ratio begins with the location of the center of the activated area. For this purpose, all nonzero pixels are taken to be part of the object. The center and the point farthest away from the center determine the major axis of the object. The minor axis is taken to be perpendicular to this. These axes provide an object-relative coordinate system in which it is possible to specify, roughly, the location of bumps and depressions in the image. The bounding rectangle of the object is taken to be the smallest rectangle, with edges parallel to the axes, that contains the image. The aspect ratio of the object is taken to be the aspect ratio of its bounding rectangle.

Moving the Finger

If the image read in is not satisfactory, as is usually the case at first, it is possible to move the finger and read another. An important part of the image analysis involves moving the finger so that an optimal image is sensed. The offset pressure, for example, is adjusted in this manner. Optimally, most of the touched area activates the mid-range of the sensor, allowing bumps and depressions to be detected easily. This is accomplished by reading in an image, computing the median pressure of all points above the noise threshold, and readjusting the finger pressure appropriately. This is repeated several times until an acceptable offset pressure is achieved.

The finger is also moved to measure the stability (resistance to roll) of an object. To measure the object's stability in a given direction, the object is pressed between the finger and the supporting surface by applying a fixed force on the object normal to the plane of the surface. The finger is then moved laterally in the desired direction. (The supporting surface was a thin layer of soft rubber, to prevent sliding.) The stability of the object is indicated by the amount of force necessary to move the finger.

The Matcher

During the hypothesize step, the program must determine which of the objects it knows best matches the known data. For a small set of possible objects, such as the fasteners, it is not really important that this be done well. For a large set of possible objects, the quality of the matcher may be a determining factor in the speed of recognition. When the hypothesis is chosen by selection of the possibility that best matches the information given, usually the choice that has the largest number of features in common with the known facts is the best choice. In a system with a large number of parameters, other factors may also be taken into consideration. For one thing, some features may be more important than others, either in general or for that particular possibility. Also, the features themselves may not exactly match—a bump may be too large, a shape distorted. In cases such as this, we wish to give the possibility only partial credit for a feature match.

The most obvious way to implement such a matcher would be to use a numerical scoring system, with the weighting of factors for feature importance and partial matches. This approach was avoided for the following two reasons. First, there would have to be a degree of arbitrariness in assigning the numbers: Is a circle a 50% match to a hexagon? Is shape 2.5 times as important as texture, or only twice as important? It is unwise to trust the sums and products of numbers if the numbers themselves are chosen arbitrarily. The second objection is more of a philosophical one—converting a complex set of symbolic structures into a single number causes us to throw away too much information too quickly. Of course, this information must eventually be lost—the matcher must terminate by selecting a single item. But the pruning can be, and is, controlled in a more reasoned manner.

The implemented matcher takes two possibilities at a time and compares them on a feature-by-feature basis. If, for a particular feature, both items match the image to about the same degree, the information is ignored. If one of the items is clearly a better match, the feature is counted in favor of the appropriate item. This procedure is repeated for each feature and then the features themselves are compared in a similar manner. A feature counted toward one item will be cancelled by a feature counted toward another, if they are of approximate importance.

In a large artificial intelligence system, the best match could be computed in parallel. Parallel

marker-propagation schemes, such as the one proposed by Fahlman (1979), would do such a task well. One important assumption, even for the parallel case, is that the binary comparison operator is transitive. Without this constraint, it would be necessary to compare each possible pair of items, a task that grows as the square of the number of items.

The transitivity of the predicate described above can be easily demonstrated, given the transitivity of individual feature comparisons. Assume that there exist three items, A, B, and C, such that A > B and B > C. Let $f(x, y)$ be the set of features counted in favor of x when compared with y. Since the individual feature comparisons are transitive,

$$f(\mathrm{A}, \mathrm{C}) = (f(\mathrm{A}, \mathrm{B}) \cup f(\mathrm{B}, \mathrm{C}))$$

and

$$f(\mathrm{C}, \mathrm{A}) = (f(\mathrm{C}, \mathrm{B}) \cup f(\mathrm{B}, \mathrm{A})).$$

If $\gg$ is the feature set comparison predicate (the second stage of the algorithm above), then A > B implies $f(\mathrm{A}, \mathrm{B}) \gg f(\mathrm{B}, \mathrm{A})$. Also, for any sets a, b, c and d such that $a \gg b$ and $c \gg d$, it must be that $(a \cup c) \gg (b \cup d)$, because features that cancel in the individual sets will also cancel in the union. The assumptions A > B and B > C imply $f(\mathrm{A}, \mathrm{B}) \gg f(\mathrm{B}, \mathrm{A})$ and $f(\mathrm{B}, \mathrm{C}) \gg f(\mathrm{C}, \mathrm{B})$ and, by the union rule, $(f(\mathrm{A}, \mathrm{B}) \cup f(\mathrm{B}, \mathrm{C})) > (f(\mathrm{C}, \mathrm{B}) \cup f(\mathrm{B}, \mathrm{A}))$. This may be rewritten as $f(\mathrm{A}, \mathrm{C}) \gg f(\mathrm{C}, \mathrm{A})$, which is the criterion for A > C. Therefore, the matching predicate is transitive.

This matcher is really overkill for a possibility set of six objects with three parameters each, but it may be necessary if the program is to be extended to a large range of objects.

Proposed Efforts

A program that distinguishes among six objects on the basis of three parameters is not too impressive. Even if it only got one bit from each parameter, it should have correctly recognized eight objects. In the future, tactile recognition programs will have much more complex and more precise representations of tactile images. Three improvements can help bring this about.

The first is texture recognition. The resolution of the tactile array sensor, while high, is not not nearly sufficient for measuring textural differences between, say, paper and glass. Texture sensing requires measuring bulk effects of many tiny surface features. It is most easily accomplished if something is slid over the surface and a pattern of vibrations is detected. This can be likened to sliding a phonograph needle over a record. Sensors of the future may use embedded piezoelectric devices, or it may be possible to use the ACS directly as sort of a carbon microphone. However the information is derived, it must be processed into a useful characterization of the texture of the surface. Of interest is the intensity and periodicity of the signal. These features may be seen directly in the frequency domain. Texture processing may bear more similarity to the analysis of sounds than to the analysis of visual images.

Another improvement might involve thermal recognition: the difference between paper and glass is that glass feels cold. This is not actually because glass is lower in temperature, but because it is a better conductor of heat and so it is more quickly able to carry away the heat generated by the body. We have constructed a small thermal conductivity sensor that works on this principle. In the sensor, a resistive heating element is sandwiched between two temperature-sensitive current sources. Any difference in the temperature of the two sensors is indicated by an easy-to-measure difference in the currents. The sensor is designed to be mounted on the finger in such a way that one temperature sensor may contact the device being tested. As the heat is drawn from the object into the sensor, a difference in temperatures will develop. The primary disadvantage of this first prototype is that it is large ($0.1 \times 0.3 \times 0.2$ in.), resulting in a relatively high thermal mass. This limits both the response time and the minimum size of the object that may be usefully tested.

The third area that shows immediate potential for further research is the coordination of multiple tactile images into a global picture. This is probably the most useful next step in tactile processing. This problem was deliberately avoided in the program

described through the choice of small objects that could be read in a single impression. Such size limitations are probably unrealistic outside the laboratory environment. The first real-world applications of tactile sensing will not be in recognizing objects that fit on the tip of the finger, but rather in orienting known objects grasped with an entire hand. This will require coordinating images from multiple sensors.

We are enthusiastic about the future prospects of automated tactile sensing. What has been described here—the sensor, the finger, and the program—is only an initial approach.

Acknowledgments

I would like to thank the following people for their help, ideas, and enthusiasm: Mike Brady, Tom Callahan, Fred Drenckhahn, Richard Greenblatt, John Purbrick, Gerald Sussman, John Hollerbach, Michael Dertouzos, Margaret Minsky, Laurel Simmons, Patrick Winston, and, most of all, Marvin Minsky.

References

Broit, M. 1979 (March). The utilization of an "artificial skin" sensor for the identification of solid objects. *Proc. 9th Int. Symp. Industrial Robotics.*

Harmon, L. D. 1982. Automated tactile sensing. *Int. J. Robotics Res.* 1(2):00–00.

Hillis, W. D. 1981 (April). Active touch sensing. A. I. Memo 629, Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Fahlman, S. 1979. NETL: *A system for representing and using real-world knowledge.* Cambridge, Mass.: MIT Press.

Larcombe, M. H. E. 1976 (March). Tactile sensors sonar and parallax sensors for robot applications. Paper delivered at 3rd Conf. Industrial Robot Tech.

Okada, T. 1979. Object handling system for manual industry. *IEEE Trans. Syst., Man, Cybern.* 9(2).

Okada, T., and Tsuchiya, S. 1977 (Oct.). On a versatile finger system. *Proc. 7th Int. Symp. Industrial Robots.*

Purbrick, J. A. (1981). A force transducer employing conductive silicone rubber. Paper delivered at 1st Robot Vision and Sensors Conf.

Stojiljkovic, Z., and Clot, J., 1977. Integrated behavior of artificial skin. *IEEE Trans. Biomed. Engineering* 24(4):396–399.

Storace, A., and Wolf, B. 1979. Functional analysis of the role of finger tendons. *J. Biomechanics* 12.

Weinreb, D., and Moon, D. 1979. *Lisp Machine manual.* Cambridge, Mass.: Massachusetts Institute of Technology Artificial Intelligence Laboratory.

A MULTI-TOUCH THREE DIMENSIONAL TOUCH-SENSITIVE TABLET

SK. Lee, W. Buxton, K.C. Smith
Computer Systems Research Institute
University of Toronto
Toronto, Ontario
Canada, M5S 1A4

(416)-978-6320

ABSTRACT

A prototype touch-sensitive tablet is presented. The tablet's main innovation is that it is capable of sensing more than one point of contact at a time. In addition to being able to provide position coordinates, the tablet also gives a measure of degree of contact, independently for each point of contact. In order to enable multi-touch sensing, the tablet surface is divided into a grid of discrete points. The points are scanned using a recursive area subdivision algorithm. In order to minimize the resolution lost due to the discrete nature of the grid, a novel interpolation scheme has been developed. Finally, the paper briefly discusses how multi-touch sensing, interpolation, and degree of contact sensing can be combined to expand our vocabulary in human-computer interaction.

1. INTRODUCTION

Rapid advancement of computer technology has opened a variety of new applications. New applications and users mean demands for new modes of interaction. One consequence of this is a growing appreciation of the importance of using appropriate input technologies (Buxton, 1982). Positioning devices are seen to be essential to graphics applications, image transducers are required for pattern recognition in medical diagnosis, touch screens are useful for the education of young children, and the QWERTY keyboard remains the usual standard for text processing. However, the range of input devices available is still quite limited, as is our understanding of how to use them in the most effective manner.

The intent of the research presented in this paper is to increase the vocabulary that can be utilized in human-computer interaction. Our approach has been to develop a new input technology that enlarges the domain of human physical gestures that can be captured for control purposes. In what follows, we will describe the technology, what it evolved from, and some aspects of how it can be used.

2. OVERVIEW

The transducer that we have developed is a touch-sensitive tablet; that is, a flat surface that can sense where it is being touched by the operator's finger. This in itself is not new. Several such devices are commercially available from a number of manufacturers (see Appendix A). What is unique about our tablet is that it combines two additional features. First, it can sense the degree of contact in a continuous manner. Second, it can sense the amount and location of a number of simultaneous points of contact. These two features, when combined with touch sensing, are very important in respect to the types of interaction that we can support. Some of these are discussed below, but see Buxton, Hill, and Rowley (1985) and Brown, Buxton and Murtagh (1985) for more detail. The tablet which we present is a continuation of work done in our lab by Sasaki et al (1981) and Metha (1982).

In the presentation which follows, we focus mainly on issues relating to the transducer's implementation. Two important contributions discussed are our method of scanning the tablet surface, and our method of maintaining high resolution despite the surface being partitioned into a discrete grid. Additional technical details can be found in Lee (1984).

3. WHY MULTI-TOUCH?

Touch sensing has a number of important characteristics. There is no physical stylus or puck to get lost, broken, or vibrate out of position. Touch tablets can be molded so as to make them easy to clean (therefore making them useful in clean environments like hospitals, or dirty environments like factories). Since there is no mechanical intermediary between hand and tablet, there is nothing to prevent multi-touch sensing. Templates can be placed over the tablet to define special regions and, since the hand is being used directly, these regions can be manually sensed, thereby allowing the trained user to effectively "touch type" on the tablet.

Without pressure sensing, however, the utility of touch tablets is quite limited. One can move a tracking symbol around the screen, for example, but when the finger is over a light button, there is nothing equivalent to the button on a mouse to push in order to make a selection. Yes, we could lift the finger off the tablet, but that would be more like pulling (rather than pushing) the button. And what if we wanted to drag an item being pointed at, or to indicate that we wanted to start inking? Lifting our finger would leave our finger off the tablet, just when we want it in contact with it the most. There are ways around this problem, but they are indirect. If, however, the tablet has pressure sensing, we can push a virtual button by giving an extra bit of pressure to signal a change in state.

Pressure has other advantages. One example is to control line thickness in a paint program. But why do we want multiple point sensing? A simple example would be if we had a template placed over the tablet which delimited three regions of 9 cm by 2 cm. Where we touch each region could control the setting of a parameter associated with each region. If we wanted to simultaneously adjust all three parameters, then we would have to be able to sense all three regions. An even easier example is using the tablet to emulate a piano keyboard that can play polyphonic music.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1985 ACM 0-89791-149-0/85/004/0021 $00.75

4. HARDWARE DESCRIPTION

A brief description of the hardware of the fast multiple-touch-sensitive input device (FMTSID) is introduced here. The design of the hardware is based on the requirements of the fast scanning algorithm and on tradeoffs between software and hardware. Many sensors have been examined for our particular application, however (Hurst, 1974; Hillis, 1982; TSD, 1982; TASA, 1980; JSRC, 1981; Metha, 1982) none seemed to have the properties that satisfy the requirements of a FMTSID. The hardware basically consists of a sensor matrix board, row and column selection registers, A/D converting circuits and a controlling CPU.

The design of the sensor matrix is based on the technique of capacitance measurement between a finger tip and a metal plate. To minimize hardware, the sensors are accessed by row and column selection. Row selection registers select one or more rows by setting the corresponding bits to a high state in order to charge up the sensors while the column selection registers select one or more columns by turning on corresponding analog switches to discharge the sensors through timing resistors. The intersecting region of the selected rows and the selected columns represents the selected sensors as a group. A/D converting circuits measure the discharging time interval of the selected sensors. A University of Toronto 6809 board is used as a controlling CPU. The touch surface of the sensor board consists of number of small metal-coated rectangular-shaped areas serving as sensor plate capacitors. The design of the metal plate area of a unit sensor depends on the measurable capacitance change that results when the area is covered by a finger tip, and on the resolution that can be implemented.

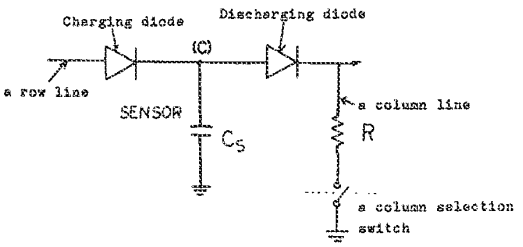


Fig. 1 A model of a selected sensor in the sensor matrix.

In order to select a sensor by row and column access, two diodes are used with each sensor. One diode, connected to the row line, is used to charge up the sensors in the row. It is referred to as the Charging Diode (CD) as shown in Figure 1. The CD also serves to block the charge flowing back to the row line when the row line voltage is dropped to zero. The other diode called the Discharging Diode(DD), connected to the column line, enables discharging of the selected row sensors to a virtual ground. Also the DD blocks charge flow from the sensors in the selected row to the sensors in the unselected rows during the discharging period. The selection of rows, by the row selection procedure, causes the sensors to be charged. The sensors in the column are then discharged through associated timing resistors connected to the column selection switches.

The charges stored in the selected row(s) flow down through the selected switches to the virtual ground of a fast operational amplifier. All the discharging currents are correspondingly added to produce a signal from which the discharging time of all the selected sensors is found by comparison with a threshold voltage.

Pressure sensitivity is incorporated by two measures: First there is the effect, here minor, of compression of the overlaying insulator. Second there is the effect of intrinsic spreading of the compressible finger tip as pressure is increased.

The software in the controlling CPU utilizes communication with the host computer to accommodate the interpolation scheme. The clock rate (10 MHz) allows about 10 counts to correspond to the sensor capacitance change due to a touch. But, of course, the capacitance of all the circuitry attached to the column line during the discharging period is much larger than the sensor capacitance. Thus before scanning the tablet for a touch, it is scanned completely in all possible resolution modes when not touched. The values so obtained are stored as references. Touches are identified by the differences between the reference values and the values measured during use.

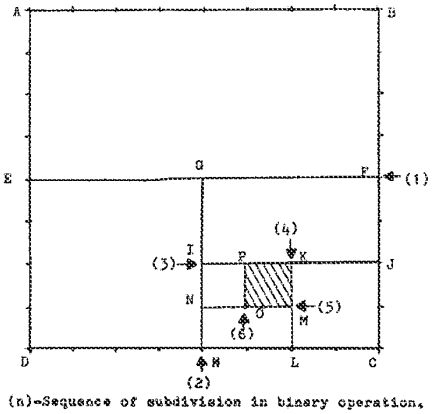
The capacitance change corresponding to the touch by more than one finger (or by the whole hand) is very large. Thus the number of bits in the counter should be enough to measure the maximum capacitance. However it is unnecessary either to have sufficient bits to measure the entire capacitance including the surrounding capacitances, or to store the corresponding "complete" counter values as references. It is necessary only to have one more bit than the number of bits required to count the value of change in the capacitance rather than the complete value in order to measure the differences of capacitance due to touch. Thus only an 8 bit counter is implemented. The counter enables the measurement of a 7 bit capacitance change regardless of the degree of overflow in the counter.

A facility is also provided for identifying templates applied to the surface of the tablet.

5. SCANNING ALGORITHM

One idea of some significance that can be introduced is to avoid scanning of all the pixels in the tablet which contain no information. For example, scanning all 2048 points of a tablet having a resolution 64 by 32 for fewer than 10 points is really quite a ridiculous idea. In fact, if the number of points to be searched is comparably small, then an improved algorithm, here called recursive area subdivision, can be used. A particular implementation example is described as follows.

Consider a tablet with resolution 8 by 8 to be searched for a touch point as shown in Figure 2. First, check the tablet for touch as a whole region as shown by the area ABCD in the figure. If touch is detected, divide the tablet into two equal regions shown by the line EF and check each of the two regions ABEF and EFCD for touchedness. Select the touched region, region EFCD in this case, and divide this into two equal regions as shown by the division line GH. Continue this process on the touched region until no further division is possible, that is, until a unit sensor, designated as the region PKMO in Figure 2, is reached. The figure also shows the sequence of subdivision in the recursive subdivision scheme.



(n)-Sequence of subdivision in binary operation.

Fig. 2 Recursive subdivision operation for 8 by 8 tablet.

Using this algorithm, a search for one point on a tablet having a resolution 64 by 32, requires 22 scanning times, that is

2 * {log sub 2} (64 * 32) = 22

If there is no overhead in the recursive subdivision process and scanning begins at the "top of the tree" (that is, with a region in which all pixels are grouped together), then using this scheme, the number of touched points that can be identified in the time that it would take to detect one touch directly (that is, if all pixels are scanned one by one sequentially) is

N = {(64 * 32) over 22} = 186.

This shows immediately that the recursive subdivision scheme is much superior to sequential scanning if the number of points to be scanned is fewer than 186.

6. INTERPOLATION

It may seem that the resolution of the hardware is too low for use in graphics applications. However touch intensity and multi-touch sensitivity can be used to enhance resolution. This is possible because the center of a touch can be most accurately estimated by an interpolation utilizing the values of the adjacent sensor intensities.

Direct interpolation schemes for a few cases has been implemented. One of interest is to interpolate an array of 3 by 3 sensors using a touched point in the center. Another is to interpolate all points on the tablet. The later one obviously provides the highest resolution but as a result it simply emulates a single touch tablet with very high resolution.

7. PERFORMANCE

7.1 Sensor

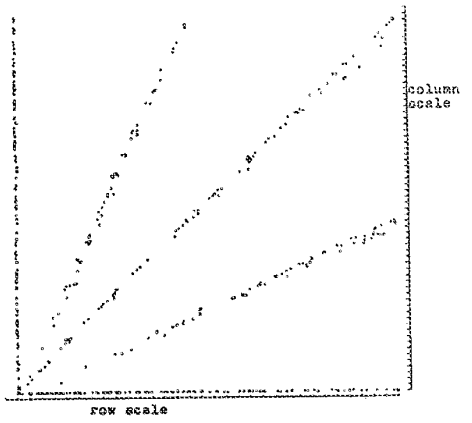
An ideal sensor matrix for a FMTSID would be one that has uniform and small reference values over a grouping level, a large variation of intensity due to a touch, and fast measurement time. The sensor matrix of the prototype, however, has a relatively wide range of reference values. However these values do not change very much over extended periods of time. The results show that doubling the number of sensors in a group in the column direction increases the reference value by a factor of about 1.5. This corresponds well to theoretical estimates. As well the results show that increasing the number of sensors in a group in the row direction, in contrast, does not increase the reference value in general, even if the number of the sensors is doubled in a group. The reference value ranges from 40 (for a single sensor in a group) to 580 (for the entire array of 64 by 32 sensors considered as a group).

In order to account for time and other variations of the reference values, a threshold is included which must be overcome in order for a touch to be detected. The threshold used ranges from 2 to 7 counts depending on group size. Using these threshold values the CPU does not report untouched points wrongly over intervals of at least 3 hours in either sequential or recursive subdivision modes. The recursive subdivision scheme uses 6 different thresholds, consequently it is very unlikely to report a wrong point whereas the linear scanning mode using only a single threshold is likely to be more sensitive.

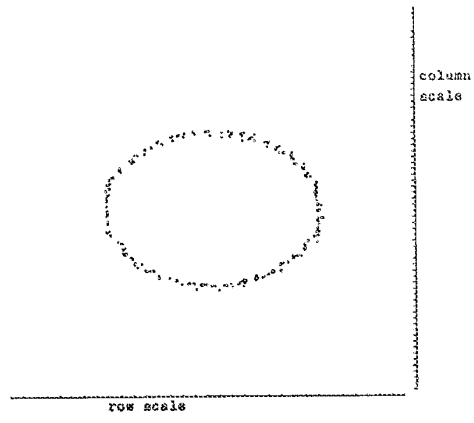
The intensity of a single touch for a single sensor group varies over the tablet but usually ranges above the threshold value by as much as 15. For a single 64 by 32 sensor group, the intensity varies from person to person but it ranges from the threshold to 124. This maximum is obtained when a palm rather than a finger touches the tablet. Another interesting feature is that the response time becomes faster as the number of sensors in a group becomes larger, and furthermore that for the 64 by 32 sensor group, it is possible to detect of a hand merely placed in the vicinity of the tablet.

7.2 Spatial Resolution

One possible and immediate interpolation scheme is to interpolate a "touched" point with all adjacent values which may not be large enough to be reported as touched. A local array of 3 by 3 points can be used for this interpolation. Some examples drawn on a laser printer (consequently having no intensity scale) are shown in Figure 3. These pictures are produced without feedback, that is, drawn without the operator looking at the output screen. This does not allow the operator to compensate, that is, to select points where data are sparse in comparison with the intended figure, but rather takes direct input from the location of the figure drawn on the input device. The first picture (a) is drawn by moving a finger in a straight line (guided by a ruler) for various angles and the second one (b) is drawn by moving a finger in a fine guide by a circle drawn on a template. These tests show that interpolation actually increases the spatial resolution as well as the locatability of a fine point on a screen.



(a) Straight lines drawn by the tablet using 3 by 3 sensor array interpolation.
The scales shown represent the boundaries of the actual sensors.



(b) A circle drawn by the tablet using 3 by 3 sensor array interpolation.
The scales shown represent the boundaries of the actual sensors.

Fig 3 Points drawn by the tablet using an interpolation method.

Since the spatial resolution in the local interpolation scheme is limited by the number of bits available from the intensities of an array of 3 by 3 sensors, other scheme was considered. In this scheme, all the points from a complete scan of a tablet are interpolated allowing the potential resolution to be almost infinite. However this process simply emulates a projective device and accordingly reports only single point, which is interpolated from all the points on the tablet. However with this scheme, there are a great many ways of pointing to a specific location on a display screen, a feature with some intriguing application possibilities.

7.3 Response Time Delay

The response time delay is the time delay from the beginning of a touch to an output received either by local terminal or by an output device attached to the host computer. For multiple touches, this delay will increase with the number of touches. The prototype used with a 9600 baud-rate terminal to measure time delays. Actual response times were measured several times and averaged for various cases and are tabulated in Table 1.

Case	best	typical	worst
(a) pts/sec	17.6	15.2	12.8
msec/pt	56.8	65.6	78.1
(b) pts/sec	19.2	17.2	16.0
msec/pt	52.1	58.1	62.5
(c) pts/sec	24.0	22.0	18.8
msec/pt	41.6	45.5	53.2

TABLE 1. Actual Response Time Delays

The cases in Table one are to be interpreted as follows:

a one sensor touched continuously

b two sensors touched at the same time continuously

c four sensors touched at the same time continuously

8. CONCLUSIONS

A prototype of a fast-scanning multiple-touch-sensitive input tablet having both the adaptability and flexibility for a broad range of applications has been designed and implemented. Capacitance measurement of individual sensor(s) which can be uniquely addressed using two diodes per sensor, makes it possible to sense both the positions and intensities of one or more simultaneous touches without ambiguity. The sensor matrix is controlled by University of Toronto 6809 board whose serial port is connected to one of the I/O ports of a host computer. Software that utilizes the recursive subdivision algorithm for fast scanning an array of 64 by 32 sensors on the tablet, and that communicates with the host computer, has been implemented and tested.

9. ACKNOWLEDGEMENTS

The research described in this paper has been funded by the Natural Sciences and Engineering Research Council of Canada. This support is gratefully acknowledged.

10. REFERENCES

Brown, E., Buxton, W. & Murtagh, K. (1985). Windows on Tablets as a Means of Achieving Virtual Input Devices. Computer Systems Research Institute, University of Toronto.

Buxton, W. (1982). Lexical and Pragmatic Considerations of Input Structures, *Computer Graphics*, 17 (1), 31 - 37.

Buxton, W., Hill, R. & Rowley, P. (1985). Issues and Techniques in Touch-Sensitive Tablet Input, Computer Systems Research Institute, University of Toronto.

Hillis, W.D. (1982), A High Resolution Imaging Touch Sensor, *International Journal of Robotics Research*, 1 (2), 33 - 44.

Hurst, G. (1974). Electrographic Sensor for Determining Planar Coordinates, United State Patent 3,798,370, March 19, 1974, Elographics, Incorporated.

JSR (1981), Pressure-Sensitive Conductive Rubber Data Sheet, Japan Synthetic Rubber Co., New Product Development Department, JSR Building, 2-11-24 Tsukiji, Chuo-Ku, Tokyo 104, Japan.

Lee, S. (1984), *A Fast Multiple-Touch-Sensitive Input Device*, M.A.Sc. Thesis, Department of Electrical Engineering, University of Toronto.

Metha, N. (1982), *A Flexible Machine Interface*, M.A.Sc. Thesis, Department of Electrical Engineering, University of Toronto.

Sasaki, L., Fedorkow, G., Buxton, W., Retterath, C., & Smith, K.C. (1981). A Touch-Sensitive Input Device. *Proceedings of the Fifth International Conference on Computer Music*, North Texas State University, Denton, Texas, November, 1981.

TASA (1980), Model: x-y 3600 and x-y controller, Model: FR-105 Data Sheet, Touch Activated Switch Arrays Inc., 1270 Lawrence Station Road., Suite G., Sunnyvale, CA 94089.

TSD (1982), Touch Screen Digitizer Data Sheet, TSD Display Products Inc., 35 Orville Drive, Bohemia, NY 11716.

11. APPENDIX A: TOUCH TABLET SOURCES

Big Briar: 3 by 3 inch continuous pressure sensing touch tablet

Big Briar, Inc.
Leicester, NC
28748

Chalk Board Inc.: "Power Pad", large touch table for microcomputers

Chalk Board Inc.
3772 Pleasantdale Rd.,
Atlanta, GA 30340

Elographics: various sizes of touch tablets, including pressure sensing

Elographics, Inc.
1976 Oak Ridge Turnpike
Oak Ridge, Tennessee
37830

KoalaPad Technologies: Approx. 5 by 7 inch touch tablet for microcomputers

Koala Technologies
3100 Patrick Henry Drive
Santa Clara, California
95050

Spiral Systems: Trazor Touch Panel, 3 by 3 inch touch tablet

Spiral System Instruments, Inc.
4853 Cordell Avenue, Suite A-10
Bethesda, Maryland
20814

TASA: 4 by 4 inch touch tablet (relative sensing only)

Touch Activated Switch Arrays Inc.
1270 Lawrence Stn. Road, Suite G
Sunnyvale, California
94089

A touch sensitive X-Y position encoder for computer input

by A. M. HLADY

National Research Council
Ottawa, Canada

INTRODUCTION

Any input device used in conjunction with a computer controlled display for interactive information exchange between man and computer must function as a position encoder. Input devices for handling two dimensional positional information can be grouped into two general types, one type encoding absolute positions and the other encoding changes in position.

Devices accepting absolute positions rely on a direct mapping of positions from an input surface to a display surface. The input surface is usually a flat plate or tablet on which positions are indicated with a movable hand held stylus. One consideration in developing a device of this type is the location of the input surface with respect to the display surface. The mapping relationship between surfaces is simplified for the user to the extent of being instinctive if the two surfaces are coincident. If the input surface is superimposed on the display surface with a finite separation, the user has to cope with the problem of parallax. A transparent input surface and a one to one mapping scale are implicit in these two arrangements. A third possibility is that the two surfaces are in different physical locations. This makes it necessary for the user to rely on a visual feedback process by observing the mapping of his selected position in relation to the desired position and then modifying his selection to decrease the difference.

The stylus used for indicating positions on the surface is typically an active one which contains a signal sensor, as for example, in the RAND Tablet,[1] or a signal radiator, as in a magnetically coupled device described by Lewin.[2] The stylus must be large enough to accommodate the necessary components, and, in addition, present devices require a cable connecting the stylus to the console for signal transmission. his makes some active styli difficult to use with dexterity.

Input devices for encoding position increments do not have separate input surfaces, and their operation depends entirely on visual feedback from the display surface. This type of device consists of a mechanical assembly having at least two degrees of freedom, such as a joy-stick or track-ball, which can be manipulated to indicate changes in the position of a cursor displayed on the screen.

Touch sensitive overlay

Work on the device described in this paper began with several primary objectives which are related to the considerations outlined above. These objectives are:

1. The device must encode absolute positions indicated by the user.
2. The input surface must be as close as possible to the display surface.
3. Positions are to be indicated with a passive stylus, including a human finger.

The first two objectives ensure that the relationships between the positional information that the user must provide and the information he observes on the screen are fundamental ones. This reduces the time and mental effort expended, especially when the device is used for item selection, that is the selec-

tion of a sub-set from a set of items shown on the display surface.

Assuming that the first two objectives are met, the third allows one to select items or positions on the screen merely by pointing at them with a finger. Because pointing with a finger is man's most natural method of indicating selection, a touch activated device creates a minimum of distraction for the user. In fact, an ideal implementation of the three objectives listed above would result in an input device that was apparent to the user in function rather than in substance.

Admittedly, the human finger is a rather coarse stylus but the resolution attainable is sufficient for many types of manual information entry. The words or phrases displayed for selection in an information retrieval system could be in a format suitable for this type of input technique. If a conventional keyboard is used in conjunction with the display terminal, a touch activated display overlay reduces the time spent in going from keyboard to display by eliminating the intermediate step of picking up a stylus. In addition, a portion of the display screen could be used as a touch sensitive keyboard with dynamic computer control of the associated key functions. The apparent simplicity, both physically and functionally, of this type of input device is a significant advantage if the user is a young child communicating with a computer-assisted instruction system.

For information entry requiring more resolution than one can obtain with a finger, a suitable passive stylus could resemble an ordinary pencil with its convenient size, light weight, and freedom of movement.

One touch sensitive device[3] that has been developed for use with a CRT consists of a number of wires terminating at the front surface of the display tube. Each wire forms the arm of an AC bridge which is unbalanced by body capacitance. A second device, developed by Control Data Corporation, has a series of translucent, touch-activated strips in front of a CRT display.

The approach taken in our case was to use an echo ranging technique with elastic surface waves. Echo ranging with pulsed ultrasonic surface waves has been applied successfully for a number of years in the field of flaw detection for structural materials. The propagation delay of ultrasonic elastic waves has been used as the basis for graphic input devices for a computer. However, these devices do not employ echo ranging and consist basically of fixed sources or radiators with the sensor in a movable stylus. One of these, developed by Woo at IBM,[4] also uses surface waves on a glass plate. The Lincoln Wand[5] provides a three dimensional input capability by using ultrasonic waves propagating in air.

In the device developed at NRC, the radiator and sensor are physically the same piezoelectric transducer which is electrically switched between the driving circuitry and the echo receiving circuitry. Pulse modulated surface waves are produced on a transparent glass plate, and any object contacting the surface reflects some of the wave energy back to the source. The distance from the radiator/sensor to the target is proportional to the time between the radiator pulse and the reception of the echo pulse.

Surface wave characteristics

An elastic surface wave can be represented mathematically as a combination of inhomogeneous longitudinal and transverse waves. This is exemplified by the particle displacements for a surface wave. The particles describe elliptical orbits with the major axis perpendicular to the surface and the minor axis parallel to the direction of propagation, corresponding to the transverse and longitudinal components respectively.

The particle displacements decrease exponentially with depth into the material, the depth decay factor being a function of the wavelength and the material. For glass, the wave energy at a depth of one wavelength is only about three percent of its value at the surface. A practical implication of this result is that, to a close approximation, a plate several wavelengths thick appears as the solid half-space necessary for true surface wave propagation.

Waves on the free surface of a solid half-space, which are also known as Rayleigh waves, are not dispersive and their phase velocity depends only on the properties of the material on which they are propagating. For plate glass the velocity is 10,400 ft/sec.

The amplitude of all elastic waves decreases with distance from the source through three mechanisms—beam divergence, scattering, and absorption. Because a surface wave is essentially a two-dimensional phenomenon, the decrease in amplitude due to beam divergence is proportional to $1/\sqrt{r}$, compared to $1/r$ for spatial waves, where r is the distance from the source. The attenuation due to scattering and absorption is related to that of spatial waves, with the attenuation factor being approximately proportional to frequency in the ultrasonic range. The attenuation coefficient of plate glass measured at 8 MHz is 0.40 nepers/inch.

An interesting property of surface waves is their ability to propagate along curved surfaces. If the radius of curvature is large with respect to the wave-

length, there is only a slight change in attenuation and velocity. This property makes it possible to employ the echo ranging principle described to produce a device which uses the curved front face of a CRT as the input surface, reducing parallax to a practical minimum.

Echo ranging parameters

All systems using echo ranging for target location have similar design parameters. Although considerable effort has gone into the refinement of echo ranging techniques for radar and sonar, the additional complexity and cost of such developments as signal correlation makes them impractical for this application.

For two dimensional space, the stylus location can be determined by measuring its distance from two fixed points or its normal distance from two fixed lines. The latter method was chosen and implemented by alternately scanning the surface in orthogonal directions using linear transducer arrays fixed at the edges of a square plate. This method can provide the stylus location directly in terms of x-y coordinates without additional computation. The line reference method also avoids the problem of edge reflections obscuring valid echoes. Furthermore, with the large beamwidths needed in the first method, it is difficult to achieve an adequate surface wave power density at frequencies in the megahertz range.

The choice of plate material was limited by the transparency requirement. Ordinary plate glass was found to be satisfactory although its attenuation coefficient is higher than that of fused quartz and some optical glass. All the glass tested had several surface flaws per square foot but most of these were shallow enough to be eliminated by localized hand grinding and polishing. The plate size was chosen to provide a usable surface of 10 × 10 inches.

Factors involved in the choice of carrier frequency include the positional resolution, the surface wave attenuation, the radiator beamwidth, the gain in reflected energy for a given target size, and the availability of piezoelectric transducers. A carrier frequency of 8 MHz was chosen for the initial device with the corresponding wavelength on glass being about 0.015 inch.

Radiator/sensor development

One of the most efficient and convenient ways of generating surface waves at frequencies in the low megahertz range is by the mode conversion of a longitudinal spatial wave. This occurs when a longitudinal wave is incident on an interface between two solid

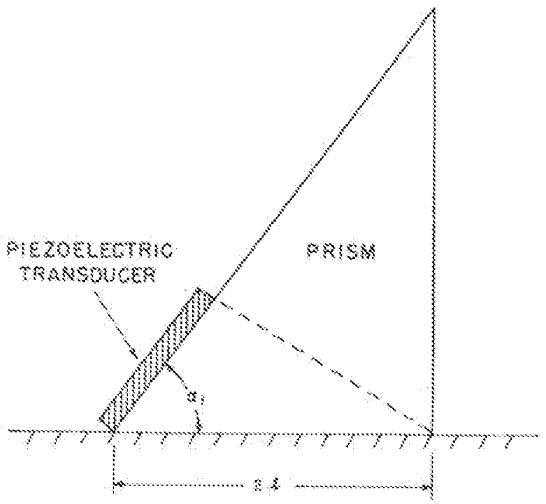


Figure 1—Surface wave radiator/sensor

materials with an angle of incidence large enough that total internal reflection occurs, and no energy is refracted into the second material. In order that the boundary conditions remain satisfied at the interface for this case, inhomogeneous longitudinal and transverse waves are produced in the second material. In ther words, a surface wave is generated.

A practical implementation of this, shown in Figure 1, consists of a thickness mode piezoelectric transducer mechanically coupled to a solid prism. Maximum surface wave output occurs for a prism angle, α_1, such that the spatial period of the surface perturbations corresponds to the wavelength of the resultant surface waves at the frequency of the incident wave. That is, when

$$c_L = c_s \sin\alpha_1$$

where c_L is the longitudinal wave velocity in the prism,

and c_s is the surface wave velocity.

For this optimum angle to be real, the prism material must be chosen so that $c_L < c_s$. One of the commonly available materials that meets this velocity requirement for generating surface waves on glass is an acrylic resin such as Plexiglass or Lucites.

The same configuration also makes an efficient surface wave sensor. In this case, incident surface waves

excite spatial waves in the prism with an angle of propagation determined by the velocity ratio. When the same transducer is used for both sending and receiving, the energy that was internally reflected within the prism during the send interval appears as clutter or noise during the receive interval. Although this excess energy is gradually absorbed by the prism material, its effect can be reduced by modifying the prism shape and coating it with an absorbent material. For the transducers actually constructed, the first two inches of range could not be used because of the clutter.

The piezoelectric transducers are made of a lead zirconate-lead titanate ceramic having a thickness mode electro-mechanical coupling coefficient of 0.66. This material is relatively good for energy transformation in both directions. The bandwidth and mechanical output power of a piezoelectric transducer are related to the mechanical impedance of the materials to which it is coupled. After some experimentation with quarter wave impedance matching transformers and various backing materials, it was decided to sacrifice bandwidth for sensitivity by using air-backed transducers bonded directly to the prism. The result was a radiator fractional bandwidth of 20 percent. The parallel components of the electrical input impedance for a small test array constructed in this way are shown in Figure 2.

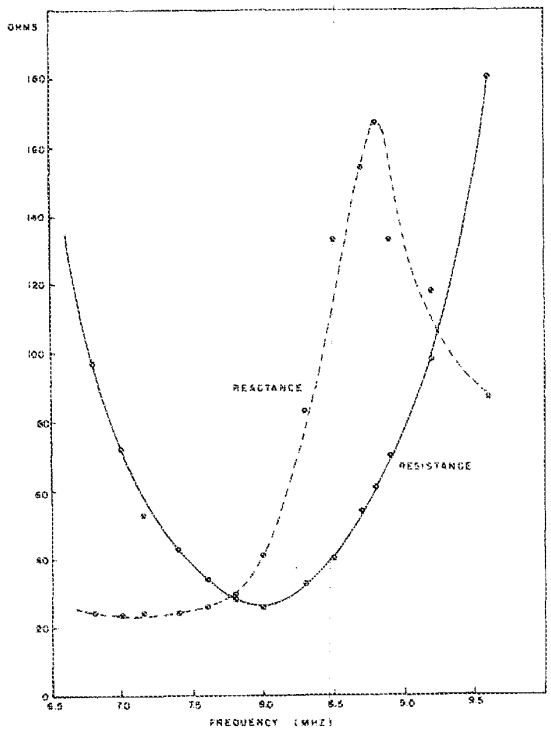


Figure 2—Parallel impedance components for a series connected array of four 1/2 × 1/4 inch transducers

For an 8 MHz pulse modulated signal with a 1.6 MHz bandwidth, the minimum resolvable stylus movement should be about 0.04 inch. As will be explained later, this resolution was attained but unusable in the first device constructed.

Array design

The method of target location being used requires a line source of waves having uniform amplitude and phase across a ten inch width. To combine separate radiator elements into a linear array with the desired characteristics, the radiation pattern of individual elements must be known. An expression for the directivity characteristics of a prism type of radiator has been derived,[5] and it yields results similar to the sin x/x function for spatial radiators. Figure 3 compares values computed for an 8 MHz radiator using this expression with experimentally measured values.

For practical plate dimensions and transducer sizes, the usable surface area lies in the far-field region of the individual elements but in the near-field region of the overall array. By computing the response for various linear array configurations, a radiator width of 0.465 inch, and a spacing of 0.565 inch, were selected.

After the arrays were assembled and tested, the measured radiation pattern was more irregular than the computations indicated. This discrepancy was attributed to the variation is spacing, orientation, and bond characteristics due to assembly tolerances and the variations in transducer sensitivity. The gaps in the pattern were sufficiently large and numerous that it was necessary to add a second set of arrays on the opposite sides of the plate. These are offset with respect to the first so that the beams from opposite arrays are effectively interleaved. The arrays are energized sequentially to avoid mutual interference.

The maximum two-way propagation time for a ten inch usable surface and a two inch buffer zone is about 200 μsec. Therefore, even with four separate arrays, the sampling rate can be greater than 1 KHz, which is more than adequate to follow normal stylus motion.

Electronic circuitry

The signal processing circuitry consists of a radiator driver, an electronic switch, and an echo receiver. The

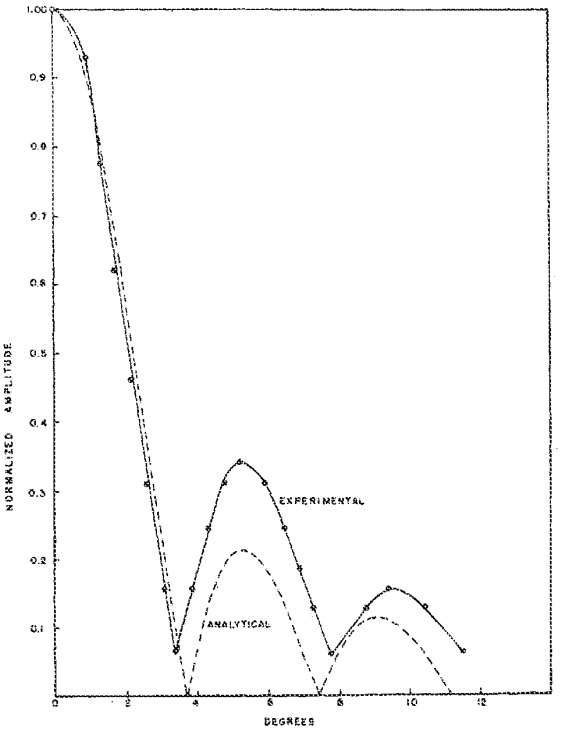


Figure 3—Directivity pattern for a surface wave radiator at 8 MHz with 0.23 inch width

timing circuitry digitizes the signal propagation time, and the control logic maintains the correct operating sequence. Figure 4 shows how these components are interconnected.

The radiator driver and the arrays are matched to 50 ohms allowing them to be connected with standard coaxial cable. The diode switch, with a four-pole double-throw action, permits the four arrays to be multiplexed into a single driver and receiver, and it also isolates the receiver during the driver pulse. The echo receiver consists of an RF amplifier followed by a demodulator and a threshold detector. The receiver gain is electronically swept during each scan to compensate for the signal attenuation with range. A range gate rejects echoes originating outside of the designated area. Figure 5 shows the demodulator and threshold detector outputs for a single scan. The signal at the centeris the echo from a finger touching the glass.

Echo timing is performed by a free running counter. Both up and down counting are required to digitize scans originating at opposite sides of the input surface. The coordinate grid is considered to have X and Y axes coincident with the edges of the usable surface, the origin being in the lower left corner. Adjustments on the range gates and counting circuitry allow the size and position of the coordinate grid to be varied slightly to permit registration with the grid of an associated display device.

The control circuitry allows two modes of operation: a continuous mode and a discrete mode. In the continuous mode, a Data Ready pulse signals the comput-

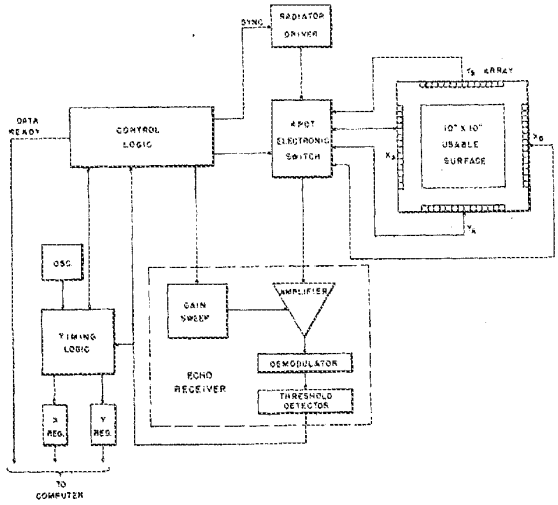


Figure 4—Position encoder block schematic

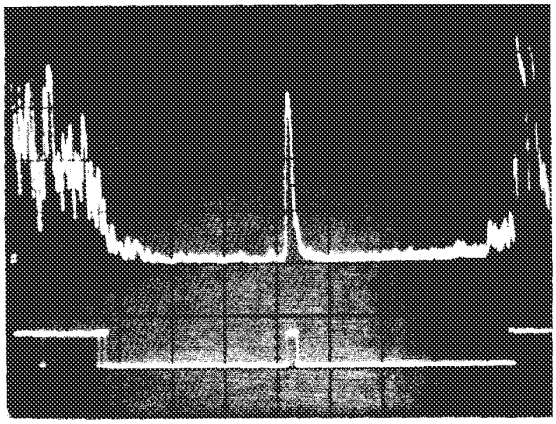


Figure 5—Echo receiver response
Vertical: Upper 0.5 v/div., Lower 5.0 v/div.
Horizontal: 25 μsec/div.

er for every set of coordinates generated while stylus contact is maintained. In the discrete mode, on the other hand, only the location of the initial contact is transferred to the computer. The stylus must be lifted and repositioned to initiate another data transfer. The discrete mode significantly reduces the amount of data that must be handled without degrading the response time when the device is being used for item selection purposes only.

In applications such as CAI which require a cluster of computer terminals in one location, it becomes feasible to time-share the electronic circuitry among several terminals, thereby decreasing cost per unit.

Device performance

The complete device is shown in Figure 6 with a static display card behind the glass for demonstration purposes. It has been interfaced with a Digital Equipment Corporation PDP-8 computer for testing and evaluation.

Tests have shown that stylus movements of 0.04 inch could be resolved, which corresponds to the calculated value mentioned earlier. However, it was found that a contact area approximately ¼ inch in diameter is necessary to ensure operation anywhere on the 10×10 inch surface. The contact area must be as large as that to bridge the regions of low sensitivity which result from the irregularities in the surface wave radiation pattern. This means that even though the device has an inherent positional resolution of 0.04 inch, the usable working resolution is considerably lower.

When using the device with a finger, a pressure of only a few ounces is adequate for operation over most of the surface. In a few places, the pressure has to be increased to enlarge the contact area sufficiently. In the former case, pointing with a finger to items displayed behind a seemingly ordinary glass plate is quite natural, and, except for the parallax, a person can make use of the device without consciously being aware of its presence.

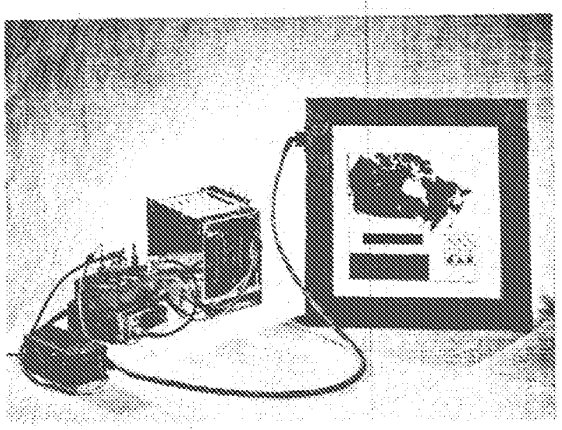


Figure 6—Touch sensitive position encoder

The position encoding is accurate and linear to about 0.5 percent. This figure takes into account the variation in wave velocity due to temperature change and material inhomogeneity, nonlinearity of the radiated wavefront, and the stability of the timing circuits.

Because scratches and marks on the glass produce small echoes which contribute to the background noise level in the receiver, some care must be used to keep the surface clean. The accumulation of fingerprints on the glass also contributes to the background noise. However, this is not a serious problem when the device is used with reasonably clean hands.

The initial device as described has served to demonstrate the feasibility of using surface wave echo ranging as the basis for a touch-sensitive position encoder. The experience gained in constructing and testing the device has been useful in determining where improvements are needed and how they should be implemented. Further computations indicate that a more sophisticated approach to the array design and assembly should improve the radiation pattern uniformity and thereby reduce the present disparity between the minimum contact size and the inherent resolution. Tests have been shown that lowering the carrier frequency to about 4 MHz should increase the signal-to-noise ratio of usable stylus echoes by decreasing the signal attenuation and lowering the sensitivity to surface contamination. The overall consequences of these changes will be to improve the performance with medium and low resolution styli and also to simplify the circuitry, and hence reduce the cost, by using two arrays instead of four. Work is progressing on the construction of a device which incorporates the improvements described.

REFERENCES

1 M R DAVIS T O ELLIS
The RAND tablet: A man-machine communication device
AFIPS FJCC Proc Vol 26 325 1964

2 M H LEWIN
A magnetic device for computer graphical input
AFIPS FJCC Proc Vol 27 881 1965

3 E A JOHNSON
Touch display: A novel input/output device for computers

Electronics Letters Vol 12 1964 Vol 13 1965
4 P W WOO
A proposal for input of hand drawn information to a digital system
IEEE Trans on Electronic Computers EC-13 609 1964
5 L G ROBERTS
The Lincoln wand
AFIPS FJCC Proc Vol 38 223 1966
6 I A VIKTOROV O M ZUROVA
Directivity diagrams of radiators of Lamb and Rayleigh waves
Soviet Physics-Acoustics Vol 9 1962 Vol 138 1963
7 I A VIKTOROV
Rayleigh waves in the ultrasonic range
Soviet Pysics-Acoustics Vol 8 1962 Vol 118 1962

A Touch-Sensitive Input Device

L. Sasaki, G. Fedorkow, W. Buxton,
C. Retterath and K. C. Smith1

Structured Sound Synthesis Project (SSSP)
Computer Systems Research Group
University of Toronto
Toronto, Ontario
Canada
M5S 1A1

INTRODUCTION

In computer music systems there is a continuing problem of finding techniques which allow suitable physical gestures to be used to express musical ideas. This is especially true in performance. This situation exists due to a lack of appropriate input transducers. Conventional computer input devices (such as sliders, joysticks, tablets, and keyboards) are being used to increased advantage (for example, Buxton, Reeves, Fedorkow, Smith, and Baecker, 1980). However, additional research is required to design new devices which lend themselves to the articulate expression of musical gestures. The "sequential drum" of Mathews (Mathews and Abbott, 1981) is one example of work in this area. The proximity sensors used in performance by Chadabe (1980) and the motion sensors used by Pinzarrone (1977) are two other examples. In the remainder of this paper we discuss yet another input device which has been developed as part of the research of the SSSP. The device is a touch-sensitive tablet which is intended to be able to be used as a pointing device, for adjusting performance parameters, and as a percussion - like input device. While the device was designed with music applications in mind, it is far more general in application.

FUNCTIONAL OVERVIEW

The basis of the tablet is a flat surface measuring 30 by 42 c.m. The surface is capable of sensing the point of contact of a finger with a resolution of 64 (horizontal) by 32 (vertical) evenly spaced units. Only one point of contact at a time can be dealt with. The device measures the capacitance at the point of contact and calculates a six-bit digit of proportional magnitude. Since capacitance is determined by the surface area covered at

1. Sasaki is currently with Bell Northern Research, Ottawa, Canada. Fedorkow is currently with Acme Widget, somewhere in New England.

the point of contact, this six-bit digit can be thought of as analogous to pressure (based on the observation that the harder you push, the more surface area your finger covers). This Z value is then transmitted to the host computer, along with the X and Y values identifying the position of the point of contact.

IMPLEMENTATION OVERVIEW

The overall architecture of the device is shown in Figure 1. Here we see that the tablet is made up of four basic modules.

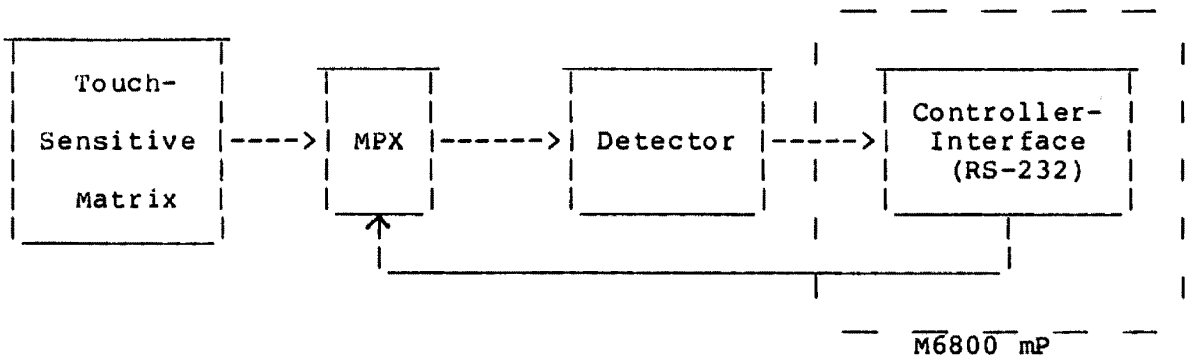


Figure 1. Tablet Block Diagram

The "Touch-Sensitive Matrix" is a printed circuit board etched with a matrix of vertical (64) and horizontal (32) conductive strips. At the edges of this array of strips are multiplexers and capacitance sensors, which are under the control of the microprocessor. Periodically, the microprocessor scans the entire tablet, reading the capacitance of each strip. The values obtained are compared to a set of reference values measured and stored upon start-up. Further processing takes note of strips which show capacitance increased beyond a threshold. Because a finger tip invariably covers several adjacent strips in both the X and Y directions, the controlling software then selects the point of highest capactance in the largest group of strips as the point of contact. A point of contact for X and Y is computed in this manner. The sum of excess capacitances for all contacted strips surrounding the contact point is scaled to a six bit number and used to indicate the pressure.

As the final step in each scan of the tablet, data is formatted and transmitted to the host, using a standard 9600 baud RS-232 serial link. Because of the amount of processing required, the tablet is scanned only about twenty times per second; this rate

is adequate for tracking hand movements, but it is too slow to be completely satisfactory as an input device for a percussive instrument.

The current version has been implemented using 49 integrated circuits. Included in this is a Motorola M6800 microprocessor which was used to implement the controller-interface module. This was realized using 1968 bytes of ROM.

EXAMPLES OF USE

To date, the tablet has been used by two programs. The first is a test program to demonstrate its sensing potential. It simply maps the three coordinates transmitted by the tablet into parameters of an FM sound being generated by the SSSP synthesizer (Buxton, Fogels, Fedorkow & Sasaki & Smith, 1978). Pressure determines volume (no contact results in silence), vertical position determines pitch, and horizontal position controls timbre (by determining the index of modulation of the FM instrument). The mapping is totally arbitrary. What is important is that the device can reliably sense pressure, and position of single points of contact, as well as track these parameters as the hand slides across the surface. In this example we have used the tablet as a position sensing device, using the absolute values of the coordinates for control purposes.

Our second software effort was to integrate the tablet into the conduct program (Buxton et al, 1980), which is the main performance system of the SSSP. Here the tablet can be used in two ways. First, it can be used as a triggering device. Thus, striking the tablet can be used to initiate events, whether they be single notes or scores. As such, the beginnings of a percussion-like interface is provided. The second use of the touch-tablet is as an alternative to sliders or the mouse for adjusting performance parameters through the control of groups. In this case the tablet can be used as a motion sensitive device, where hand motion in the horizontal and vertical domains can be independently used to increment or decrement the parameters associated with a particular group. Alternatively, the magnitude of the change of parameter values can be made proportional to the magnitude of the distance of the point of contact from the centre of the tablet. Again, the control is two dimensional, working in both the horizontal and vertical domains. Both methods of group control "delta modulate" the parameters associated with the groups in question. The two techniques have different characteristics, however. The first emulates the function of a "mouse" and a tracker-ball. The second lends itself well to combination with the triggering ability of the device. Used in combination, the tablet can be used to initiate an event, and have the properties of that event (such as duration, loudness, pitch, spectral content, etc.) controlled by where the device was hit to cause the trigger. In so doing, the full potential of the device as a

percussion instrument is greatly augmented.

CONCLUSIONS

The tablet described has clear limitations. First, the positional resolution is low, and would need to be increased for it to reach its potential as a general purpose device. It is not yet good enough, for example, to be used as a drawing device where pressure controls line thickness. Basing the design on capacitance sensing is one of the factors in this limited resolution. This also results in some variability in the pressure sensitivity. Clearly other technologies such as measuring conductance or optical techniques need to be investigated. Timing is another area where the resolution suffers. While percussion like gestures can be used effectively to trigger events, a percussionist would be frustrated by the slight lag in response and the inter-event time resolution. Such devices in the future must be designed so that the scanning can be carried out with about 5 ms of resolution. Transmission from the transducer through to the synthesis device must be traversed in about 5 ms. Finally, the most severe limitation is the device's inability to sense and track more than one point of contact at a time. A "polyphonic" version of such a tablet, one that can independently sense position and pressure for several simultaneous points of contact, would definately be welcome. However, in spite of these limitations, the tablet functions well in its present application and bodes well for the future.

ACKNOWLEDGEMENTS

The research of the SSSP has been funded by the Social Sciences and Humanities Research Council of Canada. Additional funding has also been forthcoming from the National Sciences and Engineering Research Council of Canada. This support, plus the continued physical support from the Computer Systems Research Group of the University of Toronto, is gratefully acknowledged.

REFERENCES

Buxton, W., Fogels, E., Fedorkow, G., Sasaki, L. & Smith, K. C. (1978). An Introduction to the SSSP Digital Synthesizer. Computer Music Journal 2.4: 28 - 38.

Buxton, W., Reeves, W., Fedorkow, G., Smith, K. C. & Baecker, R. (1980). A Microcomputer-based Conducting System. Computer Music Journal 4.1: 8 - 21.

Chadabe, J. (1980). "Solo": A Specific Example of Realtime Performance. In Battier, M. & Truax, B. (1980). UNESCO

Computer Music Report. Ottawa: Canadian Commission for Unesco, pp. 87 - 94.

Mathews, M. V. & Abbott, C. (1980). The Sequential Drum. Computer Music Journal 4.4: 45 - 59.

Pinzarrone, J. (1977). Interactive Woman-Machine Interaction or Live Computer Music Performed by Dance. Creative Computing 3.2: 66.

AN EMPIRICAL COMPARISON OF PIE vs. LINEAR MENUS

Jack Callahan, Don Hopkins, Mark Weiser† and Ben Shneiderman

Computer Science Department
University of Maryland
College Park, Maryland 20742

ABSTRACT

Menus are largely formatted in a linear fashion listing items from the top to bottom of the screen or window. *Pull down menus* are a common example of this format. Bitmapped computer displays, however, allow greater freedom in the placement, font, and general presentation of menus. A *pie menu* is a format where the items are placed along the circumference of a circle at equal radial distances from the center. Pie menus gain over traditional linear menus by reducing target seek time, lowering error rates by fixing the distance factor and increasing the target size in Fitts's Law, minimizing the drift distance after target selection, and are, in general, subjectively equivalent to the linear style.

KEYWORDS: menus, user interface, empirical studies, directional selection

INTRODUCTION

In presenting a list of choices to the user, most computer system designers have been limited, largely by the available hardware and software, to a linear format. The items are listed from top to bottom, sometimes with an index number for each to the item. Occasionally, the lists are multi-columned, have multiple items per line, or are even hierarchical (i.e. indented sub-choices), but for the most part lie in a strictly one dimensional structure. Many of these menus are static on the display screen or activated from mouse actions in two formats: pull-down (menu appears at a fixed label on screen when mouse directed) or pop-up (menu appears anywhere within a fixed area, occasionally the whole screen) [11]. Some systems have used the two dimensional nature of the computer display to the advantage of certain menu applications. Many flight simulation programs, for example, lay out directional headings in a typical compass format.

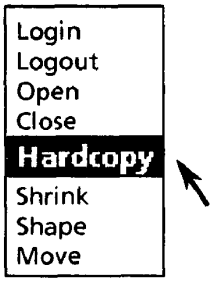


Figure 1: A typical linear menu

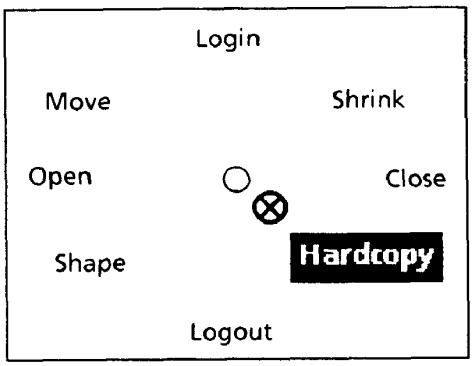


Figure 2: A crude pie menu

Item placement in menus has been an important research topic for many years. Menu organization is typically divided into three types [4]: alpha/numeric, categorical (functional), and random ordering. It is generally agreed that the performance of subjects (i.e. time to seek a target) with different placement styles converges with practice [2,10]. Further studies [9] revealed that a functional placement of items is supe-

Supported in part by the Xerox Corporation University Grants Program, NSF grant #DCR-8219507, and Office of Naval Research grant #N00014-87-K-0307.

† Computer Science Laboratory, Xerox PARC, Palo Alto, Calif. 94303.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

©1988 ACM-0-89791-265-9/88/0004/0095 $00.75

rior when the task domain is unambiguous to the user whereas an alphabetic organization can be useful in uncertain task descriptions. All of these studies have concentrated on the linear display format.

Has defaulting to a linear format (Figure 1) made some menus easier to use? Harder? By changing the menu format, can users find the item they seek faster? Is a particular menu format faster than other formats even with practice? What type of formats should be tested?

These are important questions for the designers of many systems. Software libraries of menu display routines are widely used as a default by programmers of many window systems and applications. Would it be worthwhile to present items in variable formats or perhaps in another fixed general format like the compass?

A pie menu [7] is a system facility for pop-up menus built into MIT's X windows [5] window management system, and Sun Microsystem's NeWS window system [6] and SunView window system. The pie menu interface supplies a standard library of functions that can be used by programmers to format and display menus in a circular format. The system is written in C and Forth and currently runs on a Sun Microsystems workstation. Items in the menu are placed at

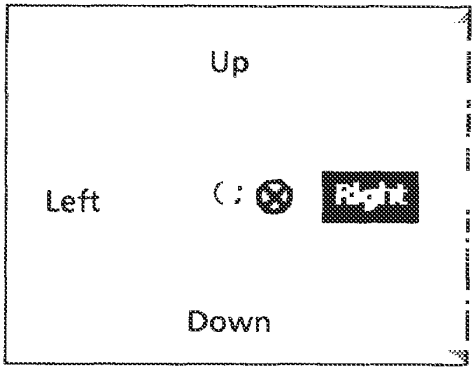


Figure 3: Pie menu activation region

equal radial distances along the circumference of a circle (Figure 2). The starting cursor position is at the center of the menu as opposed to being at the menu title or first item as in traditional pull-down menus. The cursor is under the control of a three button optical mouse on a fixed size moveable pad.

Imaginative menus formats are an inevitable future with the latest advances in window management systems. Window imaging systems using technology from laser printing protocol standards such as PostScript [1] and Interpress [12] will make it possible to display a large variety of non-rectangular shaped windows effectively on a bitmapped display. There are some obvious advantages to this organization for particular applications: compass directions, time, angular degrees, and diametrically opposed or orthogonal function names are some groupings of items that seem to fit well into the mold of the pie menu design. Alternatively, items with a sequential nature may not benefit and may in fact suffer from such a format. In addition, pie menus consume greater screen area and become polynomially larger than linear menus in both height and width with increased item size and number of items.

Distance to and size of the target are important factors that give pie menus the advantage over traditional linear menus. Even with linear menu initial cursor placement schemes where the cursor may initially be *in the middle* or *at the last item selected*, there remain target items at relatively great distances from the cursor location. Pie menus enjoy a two fold advantage because of their unique design: items are placed at equal radial distances from the center of the menu and the user need only move the cursor by a small amount in some direction for the system to recognize the intended selection. The advantages of decreased distance and increased target size can be seen as an effect on positioning time as parameters to Fitts's Law [3].

The distance to an item in any menu style can be defined as the minimum distance needed to highlight the item as selected. In both menu styles, this is defined by a region rather than a point. This region is typically of greater area than the actual target (Figure 3). Once the cursor has entered the region, the item is highlighted as feedback to the user.

Pie	*Linear*	*Unclassified*
North	First	Center
NE	Second	Bold
East	Third	Italic
SE	Fourth	Font
South	Fifth	Move
SW	Sixth	Copy
West	Seventh	Find
NW	Eighth	Undo

Table 1: Task groupings

EXPERIMENT

Introduction and hypothesis

This paper describes a controlled experiment to test two hypotheses: that pie menus decrease the seek time and error rates for menu items and that pie menus are especially useful in menu applications suited for a circular format, diametrically opposed item sets (e.g. open/close), directions (e.g. up/down) or even linear sets of items and conversely linear menus are useful for sets of linear items (e.g. one,two,three,etc.).

The experiment is a 2x3 randomized block design. Each cell is an element of the cross product of menu and task type. A typical pie task would be the compass example because it seems best suited functionally for pie menus. List of elements, like OPEN/CLOSE and UP/DOWN, whose meanings are antonyms are also classified as pie tasks. Lists, like numbers, letters and ordinals, are best suited for linear menus and are thus classified as linear tasks. Groups of menu items that have no relation to each other fall in the unclassified category. Table 1 shows an example of the groupings.

There are a total of 15 menus, a group of 5 for each task type. Subjects perform the experiment for all cells in the experiment matrix in random order in accordance with a randomized block design [8]. The subjects see each of the 15 menus four times, a total of twice in each menu format. Each cell in the experiment consists of 10 menus. Each subject therefore sees a total of 60 menus. Targets are uniformly distributed over the eight possible items.

Pilot study results

A pilot study of 16 subjects showed that users were approximately 15% faster with the pie menus and that errors were less frequent with pie menus. Statistically significant differences were found for item seek time but not task type. Subjects were split on their subjective preference of pie and linear menus. Some commented that they were able to visually isolate an item easier with linear menus and that it was hard to control the selection in pie menus because of the sensitivity of the pie menu selection mechanism. These subjects tended to be the most mouse naive of all whereas those who had heard of or seen a mouse/cursor controlled system but had not used one extensively tended to prefer pie menus. The most mouse naive users, while finding linear menus easier, tended to be better at pie menus and commented that with practice, they would probably be superior and in fact prefer the pie menus because of their speed and minimization of hand movement with the mouse. Not surprisingly, therefore, most of those preferring linear menus did not have a strong preference on the scaled subjective questionnaire.

Subjects

Subjects were volunteers from the University of Maryland Psychology Department Subject Pool. All 33 subjects were undergraduate students with little or no mouse experience. They were rewarded with 1 extra credit point for participating.

Materials

As stated, pie menus run on a Sun Microsystems Workstation as part of an enhanced version of MIT's X windows system. The screen is a 19-inch bitmapped high resolution black-and-white display. Cursor location is controlled by a three button optical mouse on a moveable mousepad made of a specially formatted reflective material.

Procedures and problems

Some changes were made from the pilot design of the experiment: a better distribution of menu targets and doubled number of menu trials, though the total number of menus remained constant.

The process of selecting items from a pop-up menu, regardless of format, can be characterized in three stages: invocation, browsing, and confirmation. To make a selection, the user invokes the menu by pressing a mouse button (*invocation*), continues to hold the mouse button down and moves to an item which is then highlighted (*browsing*) and releases the mouse button confirming the selection (*confirmation*).

The typical sequence of events for a subject is as follows:

- The target is displayed to the user in a fixed text window at the top of the screen. The cursor associated with the mouse is marked by a small hash mark "x" on the display screen.
- The user invokes the menu by pressing and holding any one of three mouse buttons. The menu appears with the cursor location unchanged (except near screen boundaries where the cursor must "jump away" to accomodate the menu). The cursor is located in the center or menu title region of pie and linear menus respectively.
- With the mouse button still depressed, the user moves the cursor with the mouse towards the textual target as indicated. Selections highlight as the cursor moves into distinct activation regions. As noted, the activations regions for pie menus are "pie" shaped sections that extend to the screen boundaries and are rectangular sections extending horizontally towards the screen boundaries for linear menus.
- Once selection is made, the user releases the mouse button to confirm the selection. The menu disappears from the display screen. The cursor remains at the screen position relative to the selection location. If the selection is correct, the process begins again with a new target and possibly a new menu style. Otherwise, if the selection is not the requested target, an audible "beep" tone is heard and the user attempts the task again.

Basically, the computer posts the target name at the

top of the screen, the user invokes the current menu, moves to the target item, and confirms the selection by releasing the mouse button. This sequence, called a task, is repeated 60 times by each subject. Each subject saw 6 sequences of 10 menus each. In each ten menu sequence, the menu type was the same, either pie or linear, and since there are only 5 menus per task type, each menu appears twice in the sequence.

	Task type			
	Pie	Linear	Unclass.	$Mean_{menu}$
Using pie menus	2.20	2.18	2.40	2.26
Using linear menus	2.68	2.30	2.94	2.64
$Mean_{task}$	2.44	2.24	2.67	

Table 2: Target seek time (sec) means per cell, menu type, and task type

	F	$PR > F$
Menu type	16.23	0.0003
Task type	6.93	0.0030
Menu type X Task type	2.82	0.0750

Table 3: repeated measures analysis of variance results for target seek time

The 10 menu sequences correspond to the cells in the experiment table design. Each subject performed a sequence for all 6 cells in random order. 60 data points are collected per subject. A total of 33 subjects performed the experiment for a total of 1980 data points.

For each task, the time from the first mouse button down to the correct target selection is the seek time for the item. If the user selected the wrong item, the time is included in this interval. The number of errors made as well as the sub-interval times when errors are made is recorded during the experiment by the system. All subjects performed the test adequately and no person failed to finish the assignment.

RESULTS AND DISCUSSION

A repeated measures analysis of variance was performed on the data. Table 2 shows the means per cell, per row, and per column. Table 3 displays the repeated measures ANOVA results. A Tukey analysis reveals that there is a statistical significant difference ($P < 0.01$) between overall menu type performance and task type performance in target seek times. Pie tasks and linear tasks did not significantly differ from each other, but both organizations are an improvement over the unclassified menu tasks. Slight statistical significant difference ($P = 0.075$) between cells in the experiment design is also observed. No other interaction was observed to be significant.

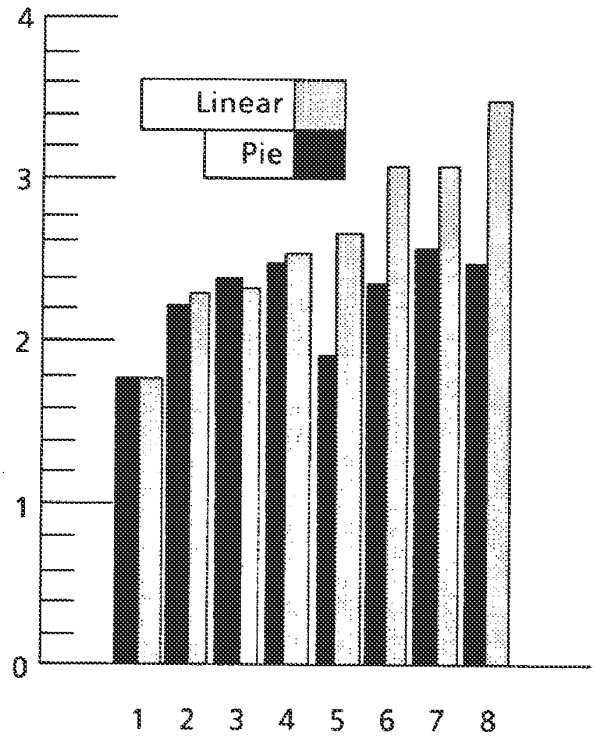


Figure 4: Target location (x) vs. seek time (y) in seconds

The statistically significant difference between menu type performance is the central result of this study. The task type difference reiterates earlier study results [2,9] that showed that some organization is helpful. Furthermore, the slight interaction between menu types and task types tends to confirm the hypothesis that certain task groupings perform well with particular menu formats. The reason for a lack of strong correlation is evident in the lower mean for pie menus even on linearly grouped tasks.

Figure 4 displays the target location by item plotted against the mean seektime. The mean seektime across target location for pie menus is fairly constant. As expected for linear menus, the mean seek time increases proportionally to the distance of the target from the initial cursor location. Analysis of seektime vs. number of menus seen shows that no strict convergence occurs between the two menu styles, though mean seektimes did decrease for both pie and linear menus with practice.

With error times removed from the results (measuring time from menu invocation to *first correct* choice), the menu styles compared relatively the same as the comparison which includes error times because of the error rates.

An analysis of seek time based on Fitts's Law $T = K_0 + Klog_2(D/S + 0.5)$ where T = time to position cursor using mouse (seek time), K_0 = constant time to adjust grasp on mouse, K = constant normalization factor (positioning device dependent), S = size of target in $pixels_2$, D = distance in screen pixels, helps explain our results because the ratio of the distance (D) to target size (S) is smaller for pie menus. The fixed target distance and increased size of targets for pie menus decreases the mean positioning time as compared with linear menus. In our experiment, the activation region for an item constitutes the target. All subjects were informed of the fact that their target was not necessarily the text, but the region containing the text target item. This was clearly understood by all participants. The font size for text items in both

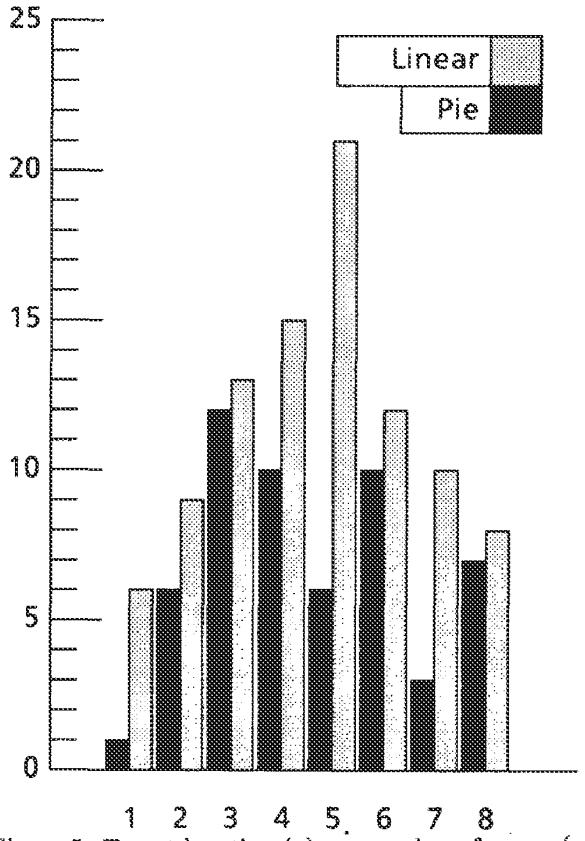


Figure 5: Target location (x) vs. number of errors (y)

menu styles was the same, yet the target region size for pie menus ($3500 - 6000 pixels^2$) was on the order of 2-3 times the size of linear menu activation region sizes ($1000 - 2000 pixels^2$). The distance from the center of a pie menu to an activation region is 10 pixels while the distance in linear menus varied from 13-200 pixels.

Figure 5 displays the target location plotted against the total number of errors across all subjects. Pie and linear menus seem to suffer from a similar phenomenon - errors are made more often on items in the central region of the menu display. These are the items with the most interaction with neighboring items [2].

	Task type			
	Pie	Linear	Unclass.	$Mean_{menu}$
Using pie menus	0.45	0.60	0.60	0.55
Using linear menus	0.88	0.73	1.24	0.95
$Mean_{task}$	0.66	0.66	0.92	

Table 4: number of errors means per cell, menu type, and task type (all observations including no errors)

Repeated measure analysis of variance results on the error rates show marginally statistically significant differences ($P = 0.087$) between pie and linear menus (Tables 4 and 5). No other statistically significant differences were observed.

Subjective results obtained in the pilot study repeated themselves in the experiment. Subjects were split on preferring one menu type over another but those who preferred linear menus had no strong conviction in this direction and most agreed that with further practice

	F	$PR > F$
Menu type	3.12	0.0869
Task type	0.93	0.4066
Menu type X Task type	1.34	0.2773

Table 5: repeated measures analysis of variance results for number of errors

they might prefer the pie menu structure. Those who preferred pie menus generally felt fairly confident in their assessment and this is reflected in the questionnaires.

One subject complained of having a problem with *menu drift* which is the phenomenon which occurs as the result of the cursor relocating to the relative screen location of the last selected target. With linear menus, this tends to "drift" the cursor towards the bottom of the screen. This may explain the higher error rate for linear menus, but the same problem occurs to a lesser degree with pie menus. This, in fact, we believe to be another positive feature of pie menus: the cursor drift distance is minimized. Most subjects had no problems coping with drift in either menu style. One area of further research is measuring the extent and effect of this problem.

CONCLUSIONS

What does this mean? Should we program pie menus

into our bitmapped window systems tomorrow and expect a 15-20% increase in productivity since users can select items slightly faster with pie menus. Pie menus seem promising, but more experiments are needed before issuing a strong recommendation.

First, this experiment only addresses fixed length menus, in particular, menus consisting of 8 items - no more, no less. Secondly, there remains the problem of increased screen real estate usage. In one trial a subject complained because the pie menu obscured his view of the target prompt message. Finally, the questionnaire showed that the subjects were almost evenly divided between pie and linear menus in subjective satisfaction. Many found it difficult to "home in on" a particular item because of the unusual activation region characteristics of the pie menu.

One assumption of this study concerns the use of a mouse/cursor control device and the use of pop-up style menus (as opposed to menus invoked from a fixed screen location or permanent menus). Certainly, pie menus can and in fact have been incorporated to use keyed input [7] and fixed "pull-down" style presentation (the pie menu becomes a *semicircle* menu). These variations are areas for further research.

One continuing issue with pie menus is the limit on the number of items that can be placed in a circular format before the size of the menu window is impractical. Perhaps, like the limiting factors in linear menus concerning their lengths, pie menus reach a similar "breaking point" beyond which other menu styles would be more useful. Hierarchical organization, arbitrarily shaped windows (Figure 6), numeric item assignment and other menu refinements as well as further analysis is contained in [7]. Pie menus offer a novel alternative worthy of further exploration.

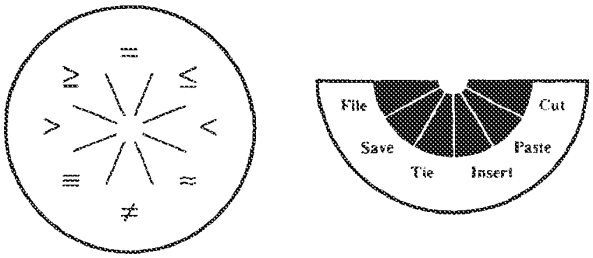


Figure 6: Advanced "pie" menus

ACKNOWLEDGEMENTS

The authors wish to thank the following people for their invaluable help in the preparation of the experiment, analysis of results and statistics, and this paper: Jim Purtilo, Nancy Anderson, Ken Norman, John Chin, Linda Weldon, Mark Feldman, Mike Gallaher, Mitch Bradley, and Glenn Pearson.

REFERENCES

[1] Adobe Systems Inc. *Postscript Reference Manual*, Palo Alto, Calif., 1985.

[2] Card, S.K. User perceptual mechanisms in the search of computer command menus, In *Proceedings - Human Factors in Computer Systems 1982* (Gaithersburg, Md., Mar. 15-17). ACM, New York, 1982, pp. 190-196.

[3] Card, S.K., Moran, T.P., and Newell, A. *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum, London, 1983.

[4] Dray, S.M., Ogden, W.G., and Vestewig, R.E. Measuring performance with a menu-selection human computer interface *Proceedings of the Human Factors Society: 25th Annual Meeting 1981* (Rochester, N.Y., Oct. 12-16). Human Factors Society, Santa Monica, Calif., 1981, pp. 746-748.

[5] Gettys, J. and Newman, R., *X Windows*. MIT, 1985.

[6] Gosling, J. *NeWS: A Definitive Approach to Window Systems* Sun Microsystems Corp., Mountain View, Calif., 1986.

[7] Hopkins, D., Callahan, J., and Weiser, M. Pies: Implementation, Evaluation and Application of Circular Menus. University of Maryland Computer Science Department Technical Report, 1988.

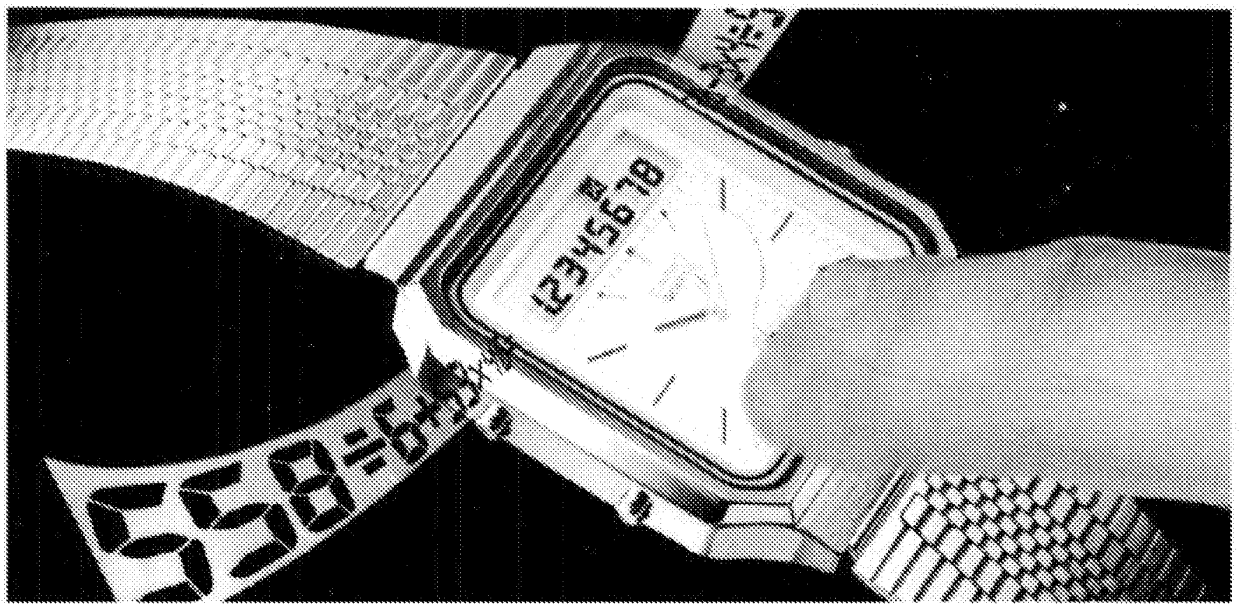
[8] Kirk, R. *Experimental Design: Procedures for the Behavioral Sciences*. Brooks-Cole, Belmont, Calif., 1968.

[9] McDonald, J.E., Stone, J.D., and Liebelt, L.S. Searching for items in menus: The effects of organization and type of target. *Proceeding of the Human Factors Society: 27th Annual Meeting 1983* (Norfolk, Virginia, Oct. 10-14). Human Factors Society, Santa Monica, Calif., 1983, pp. 834-837.

[10] Perlman, G. Making the right choices with menus. *INTERACT '84, First IFIP International Conference on Human Computer Interaction*. North-Holland, Amsterdam, 1984, pp. 291-295.

[11] Shneiderman, B. *Designing the User Interface*, Addison-Wesley, Reading, Mass., 1987.

[12] Xerox Corporation, *Interpress Electronic Printing Standard*. Stamford, Conn., 1984.



NOW... THE INVISIBLE CASIO CALCULATOR WATCH

Finger-write your figures on the watch face.

Introducing the timepiece that adds another dimension to watch technology. This new CASIO combines state-of-the-art microcomputer technology with the latest styling to give you an elegant timepiece with a multitude of functions.

And the most remarkable function of all is this... The watch face actually reads and computes math problems you trace on its face.

And there's more, much more... for less than $100.00.

ELECTRO-TOUCH TECHNOLOGY.

This handsome and superbly styled timepiece has a transparent crystal that reads finger-strokes you trace across its face. Each figure and math symbol you outline appears on the background digital display. Take your finger across twice (=) and the answer presents itself like magic.

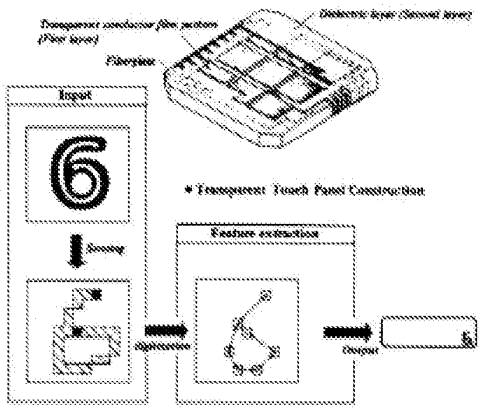
No keys, no keyboards, no need to use stylus or pen. Even the broadest fingers will work. Add, subtract, multiply, divide — perform chain and mixed calculations to eight places, plus decimal. There's even an indicator telling you which function is being performed.

DIGITAL PRECISION, ANALOG STYLE.

This handsome CASIO was created exclusively for the man who recognizes exceptional styling. And that's what you get with this sleek new timepiece. From its raised time markers to the elegant case and band of tight mesh steel or gold plate, this beauty has the special look of luxury that is never out of place... from the very casual to the most formal occasions.

You get the classic elegance and convenience of analog watch hands. Subdued in the background is the modern message of digital precision. The digital display can be set in 12-hour or 24-hour digital time. A pre-programmed calendar is set until the year 2019. It's a handsome and functional way to wear time with accuracy to 1/2 second per day.

HERE'S HOW THIS MARVEL WORKS



PRECISION ALARM AND STOPWATCH.

You can program this multi-talented wrist alarm for daily events. To wake you up. Catch your bus. Program it for any minute you choose. Or set it to beep-beep you every hour.

In addition to helping you organize your day, this gifted chronograph boasts a fiercely talented stopwatch. Record normal times and net times with accuracy to 1/10 second. A beep confirms starts and stops. Ideal for tallying cooking time, minutes on your parking meter or anything you like.

WE INTRODUCE IT AND WE GUARANTEE IT.

When we first heard the engineers at CASIO were on the brink of perfecting a finger-trace recognition calculator watch, we had hopes of being the first to offer it. And now that's a reality.

Because this innovative timepiece is now available only through On The Run, to be assured earliest delivery, please order yours now. Chrome and stainless model **AT-550** is only **$99.95** and gold-plated model **AT-550G** is **$119.95.**

See how this handsome accessory can be worn anytime, anywhere. Discover the convenience of finger-trace calculation and all the other special features of this talented timepiece. Once you see this handsome and functional timepiece, you're sure to want to keep it. If not, we guarantee your satisfaction. Simply return it in new condition within 30 days for a full and courteous refund. One-year warranty included.

CREDIT CARD HOLDERS ORDER TOLL-FREE TODAY.

To order, call toll-free number below, or send a check or money order for the total amount **plus $2.50** for the first watch, $1.00 for each additional watch for shipping and insurance. Add an additional $2.00 for UPS air delivery. ND residents add 4% tax.

800-437-4385

On The Run

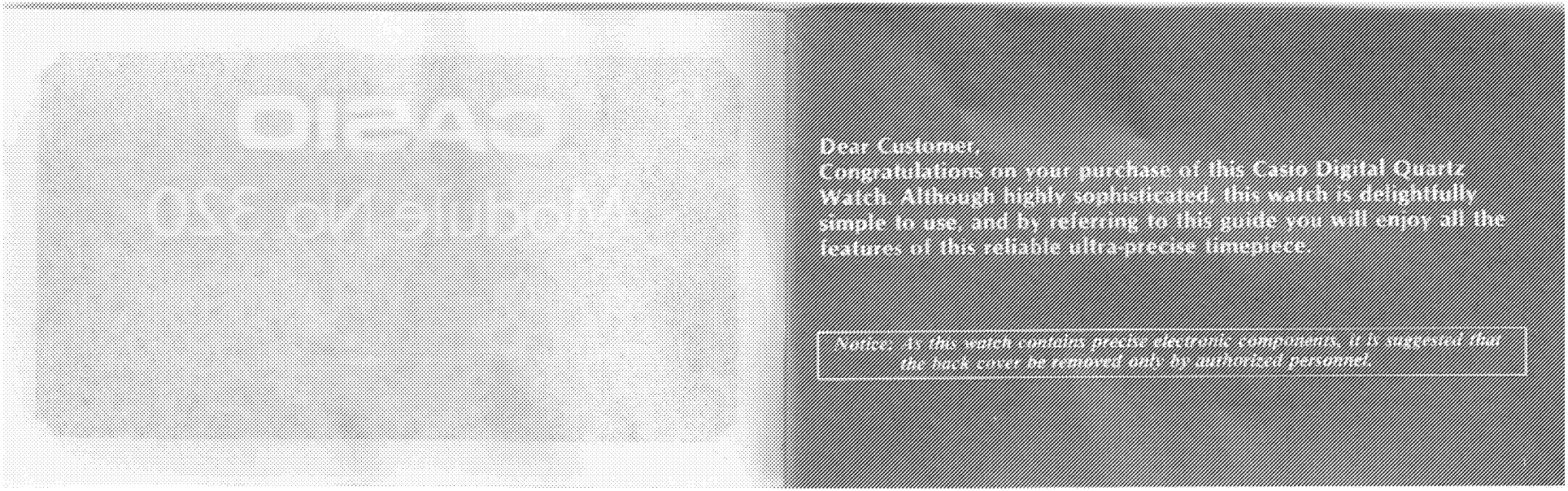
107 Roberts St., Department PS-2
P.O. Box 67, Fargo, ND 58107
Telephone (701) 232-9400

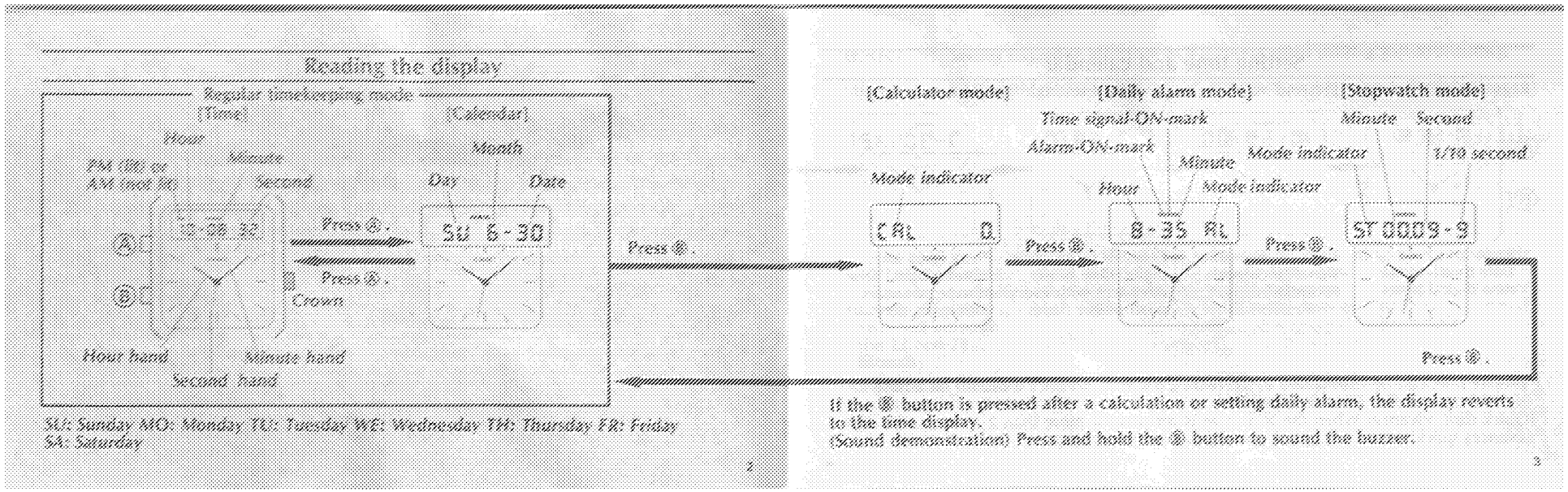
Call or write for a free one-year subscription to our catalog of timepieces and high-tech items for better living.

AT-550 FP-1

© 1986 ON THE RUN

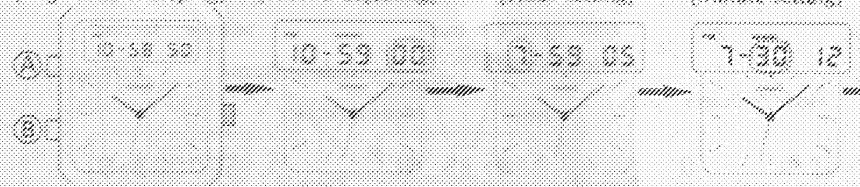
ES
CASIO
Module No.320
GUARANTEE CERTIFICATE
MODEL:
DATE OF PURCHASE:
OFFICIAL DEALER STAMP:


Dear Customer,
Congratulations on your purchase of this Casio Digital Quartz Watch. Although highly sophisticated, this watch is delightfully simple to use, and by referring to this guide you will enjoy all the features of this reliable ultra-precise timepiece.
Notice: As this watch contains precise electronic components, it is suggested that the back cover be removed only by authorized personnel.


Reading the display
Regular timekeeping mode
[Time]
Hour
Minute
PM (⏾) or AM (not ⏾)
Second
[Calendar]
Month
Day
Date
Press Ⓐ.
Press Ⓐ.
Crown
Hour hand
Minute hand
Second hand
SU 6-30
Press Ⓑ.
[Calculator mode]
Mode indicator
CAL 0
Press Ⓑ.
[Daily alarm mode]
Time signal-ON-mark
Alarm-ON-mark
Hour
Minute
Mode indicator
8-35 AL
Press Ⓑ.
[Stopwatch mode]
Minute
Second
1/10 second
Mode indicator
ST 0003-9
Press Ⓑ.
Press Ⓑ.
SU: Sunday MO: Monday TU: Tuesday WE: Wednesday TH: Thursday FR: Friday SA: Saturday
If the Ⓑ button is pressed after a calculation or setting daily alarm, the display reverts to the time display.
(Sound Demonstration) Press and hold the Ⓑ button to sound the buzzer.

Setting time and calendar

[Regular timekeeping] [Second adjusting] [Hour setting] [Minute setting]



Press and hold Ⓐ in time mode to set new digital time.

Press Ⓑ on a time signal to correct seconds.*

Press Ⓐ to set hour digits. One hour advances with every push of Ⓑ.

Press Ⓐ to set minute digits. One minute advances with every push of Ⓑ.

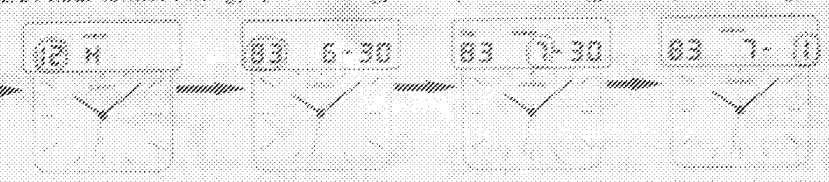
*Precise time can be maintained by correcting the seconds once a month on a time signal from a radio, TV, telephone, etc.

(Quick digit advance) When the Ⓑ button is pressed for more than 2 seconds, the digit advances quickly. When released, the digit advance will stop.

4

IMPORTANT: Setting sequence MUST BE FOLLOWED when making any new setting.

[12/24-hour format setting] [Year setting] [Month setting] [Date setting]



Press Ⓐ to set 12/24-hour format. With every push of Ⓑ, the display is switched between the 12 and 24 formats.

Press Ⓐ to set year digits. One year advances with every push of Ⓑ.

Press Ⓐ to set month digits. One month advances with every push of Ⓑ.

Press Ⓐ to set date digit. One date advances with every push of Ⓑ. Press Ⓐ to complete.

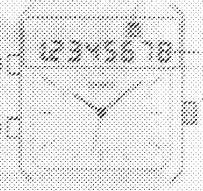
(Auto-retrieve function) When setting the watch, if you leave it alone for 2 to 4 minutes, the display will automatically return to the regular timekeeping mode.

5

Calculator operation

Deletes the displayed number by one digit for entry correction.

A function command sign (flashes when a number is set as a constant).



8-digit entry (7-digit for negatives) can be made.

Clears entry for correction.
Releases overflow or error check. Overflow is indicated by an "E" sign and stops the calculation.
Overflow occurs when the integer part of an answer, whether intermediate or final, exceeds 8 digits (7 digits for negatives).

[Before starting calculations]

Make sure the display shows 0 before starting calculations.
To input figures, write them directly on the glass with your finger-tip. (No input is possible with a fingernail or pen.)
Write figures slowly and carefully using the whole space of the glass, checking your entries on the display.
Be sure to stop your finger whenever it comes to an inflection point of a figure, and then move the finger again.
When writing a figure or symbol with multiple strokes, move your fingertip off the glass at the end of each stroke, and start the following stroke before the input recognition indicator (◄◄◄) disappears.

Making strokes not in the specified order or directions may input a wrong figure or symbol.
Please carefully read the "Notes on how to input".
Input may become difficult in low temperatures or when the air is too dry.
Avoid operation when moisture or dirt are present on the watch face. Wipe off with a dry, soft cloth.
If the Ⓒ button is pressed with your finger touching the glass, a figure or symbol may be input. Press Ⓐ or Ⓑ button to clear.
When calculating with the watch off your wrist, keep touch contact with the back of the case.
A special process has been used in manufacturing the crystal surface to allow the touch sensor function.
This is protected with a special coating to resist dirt and scratches.
Special care should be given to the crystal. This care should include avoiding scratches if at all possible as this can cause touch key sensor input problems.

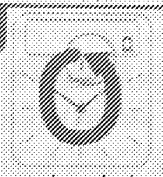
This watch incorporates a unique new mechanism which permits you to carry out calculations by writing figures directly on the glass with your fingertip.

What is this new mechanism?
Finger-Trace (handwritten figures or symbols) recognition system. The mechanism reads the order in which strokes are made and their directions.

Notes on how to input

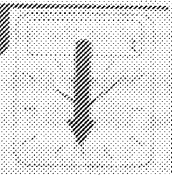
The following points should be observed when you write figures.

0



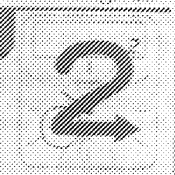
Align each end of the stroke.
(Other writing forms)

1



Draw a straight, vertical line.
(Other writing forms)

2

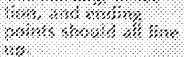


Clearly show the inflection point.
(Other writing form)

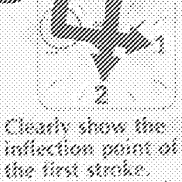
3

- Both upper and lower curves should be of the same size.
- The starting, inflection, and ending points should all line up.
- The figure should not lean over to the right.

(Other writing form)

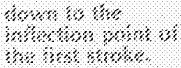


4

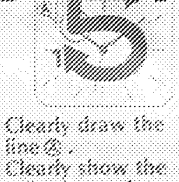


- Clearly show the inflection point of the first stroke.
- Draw the second stroke parallel to the line drawn down to the inflection point of the first stroke.

(Other writing forms)



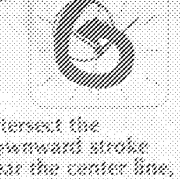
5



- Clearly draw the line ②.
- Clearly show the inflection point.

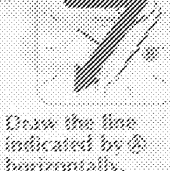
(Other writing forms)

6



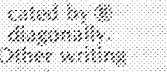
Intersect the downward stroke near the center line, dividing the figure into upper and lower halves.
(Other writing form)

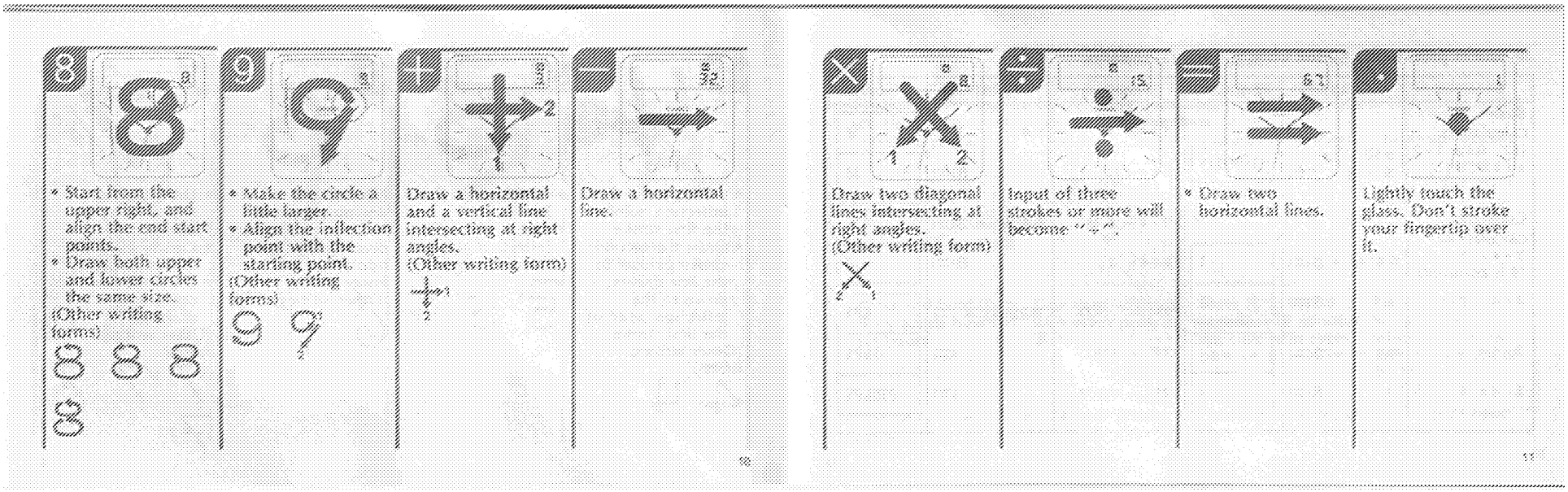
7



- Draw the line indicated by ② horizontally.
- Clearly show the inflection point.
- Draw the line indicated by ③ diagonally.

(Other writing forms)



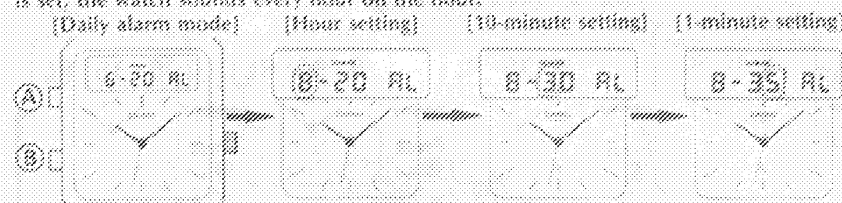

8
• Start from the upper right, and align the end start points.
• Draw both upper and lower circles the same size.
(Other writing forms)
9
• Make the circle a little larger.
• Align the inflection point with the starting point.
(Other writing forms)
+
Draw a horizontal and a vertical line intersecting at right angles.
(Other writing form)
−
Draw a horizontal line.
10
×
Draw two diagonal lines intersecting at right angles.
(Other writing form)
÷
Input of three strokes or more will become "÷".
=
• Draw two horizontal lines.
.
Lightly touch the glass. Don't stroke your fingertip over it.
11

Be sure to press the Ⓑ button when starting calculations.

EXAMPLE	OPERATION	READ-OUT	EXAMPLE	OPERATION	READ-OUT
Basic calculation: (12 − 0.5) × 3 + 7 = 4.9285714 . . .	Ⓑ 12 ⊟ (×) 5 × 3 + 7 (=)	4.9285714	3 × 4 = 12	Ⓑ 4 × × 3 (=)	12
			8 × 4 = 32	8 (=)	32
Constant calculation: 3 + 4 = 7 (4 is constant)	Ⓑ 4 ⊞ ⊞ 3 (=)	7	3 × 4 = 0.75	Ⓑ 4 × × 3 (=)	0.75
			8 × 4 = 2	8 (=)	2
8 + 4 = 12	8 (=)	12	25^2 = 625	Ⓑ 25 ⊠ ⊠ (=)	625
3 − 4 = −1	Ⓑ 4 − ⊟ 3 (=)	−1	25^3 = 15625	(=)	15625
8 − 4 = 4	8 (=)	4	25^4 = 390625	(=)	390625

Setting daily alarm

If the daily alarm is set the buzzer sounds for 20 seconds at the preset time every day until cleared. To stop the buzzer while sounding, press the Ⓐ button. If the time signal is set, the watch sounds every hour on the hour.



Press and hold Ⓐ in daily alarm mode to set hours.

One hour advances with every push of Ⓑ.*

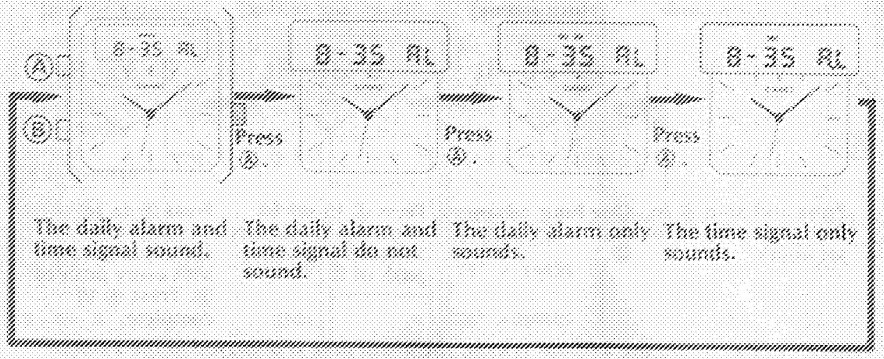
Press Ⓐ to set 10 minutes. 10 minutes advance with every push of Ⓑ.

Press Ⓐ to set minutes. One minute advances with every push of Ⓑ. Press Ⓐ to complete.

*When the watch is in the 24-hour system, the alarm time is displayed in the 24-hour system.

[ON or OFF setting of daily alarm and time signal]

[The alarm-ON-mark and time-signal-ON-mark appear.]



[The alarm-ON-mark and time-signal-ON-mark disappear.]

The daily alarm and time signal sound.

[The alarm-ON-mark only appears.]

The daily alarm and time signal do not sound.

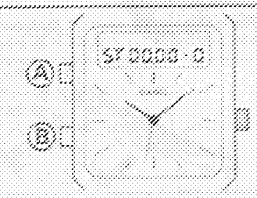
[The time-signal-ON-mark only appears.]

The daily alarm only sounds.

The time signal only sounds.

Press Ⓐ.

Stopwatch operation



Press Ⓑ to start or stop.
Press and hold Ⓑ to reset.
A signal confirms start/stop operation.
(Working range) The stopwatch display is limited to 59 minutes 59.9 seconds, for longer times reset and started again.

Setting analog timekeeping

1) Pull the crown out when the second hand is at the 12 o'clock position and the second hand stops.
2) Set the hands by turning the crown.
3) In accordance with a time signal, push the crown in.

A gain or loss of one second or less may result from properties of mechanical parts.

[How to replace the battery]

CAUTION: Battery replacement should not be attempted without use of the correct tools.

1. Check the type of back cover

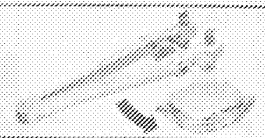
Screw-in type

Snap-on type (A)

Snap-on type (B)

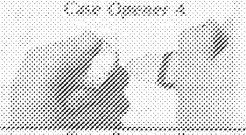
2. Removing the back cover Place the watch on soft material, like buckskin, and hold it firmly.

With the Adjustable Case Opener, turn the back cover counterclockwise.



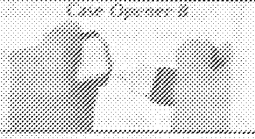
Insert Opener A in the recess and move from side to side to make a gap between the cover and the case. Then use the opener to pry off the cover.

Case Opener A



Fit Case Opener B into the notch and pry open the back cover.

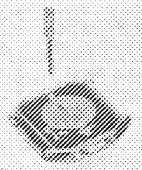
Case Opener B



BATTERY LIFE: 12-month battery life starts when battery is factory installed. At first sign of power fade (dim display), renew battery at sales store or Casio distributor.

3. Replacing the battery

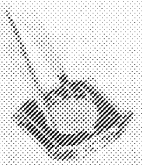
Using a screwdriver, remove screw(s) from the battery holder.
Replace dead battery(s) and attach the battery holder.



CAUTION
* Avoid touching the contact (—) of the battery.
* Never hold the contacts with metallic tweezers.

4. AC (ALL CLEAR)

As shown below, touch the AC contact and the battery (–) side with metallic tweezers. Contact should be about 3 seconds.



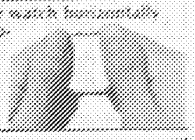
IMPORTANT
a) Contacting AC (ALL CLEAR) is necessary, when a new battery has been put in, because the memories' counters may cause erratic displays.
b) On some models, pushing the light button will turn on the display.

5. Fitting the back cover

Using the Adjustable Case Opener, tighten the back cover.

Place the watch on a hard press, and push the back cover in gently.

Hold the watch horizontally and snap-fit the back cover.



Care of your watch

- Battery Life: 12-month battery life starts when battery is factory installed. At first sign of power fade (dim display), renew battery at sales store or Casio distributor.
- As this watch contains precise electronic components, it is suggested that the back cover be removed by authorized personnel.
- Your watch is ranked A through E according to the water resistance chart below: Check the ranking of your watch to determine proper use.

Rank	Case Designation	Splashes, rain, etc.	Swimming, car-washing, etc.	Snorkeling, diving, etc.	Scuba diving
A*	—	No	No	No	No
B	WATER RESISTANT	Yes	No	No	No
C**	50M WATER RESISTANT	Yes	Yes	No	No
D***	100M WATER RESISTANT	Yes	Yes	Yes	No
E****	200M WATER RESISTANT	Yes	Yes	Yes	Yes

*The watch is sweat-resistant but not water-resistant. So be careful not to get it wet. The pendant type does not meet sweat-resistant levels.

**50M WATER RESISTANT casing model does not permit underwater button operations.

***100M WATER RESISTANT casing model permits underwater button operations (except where buttons are countersunk).
Should the watch be exposed to sea water, wash it well with fresh water and wipe dry.

****200M WATER RESISTANT casing model withstands scuba diving use (except diving using helium-oxygen gas).

- A waterproof rubber seal is used to exclude water and dust. As rubber deteriorates after long usage, the seal should be replaced periodically (every 2—3 years).
- Should water or condensation appear in the watch, immediately have it overhauled because water can corrode electronic parts inside the case.
- Avoid exposing it to extremely high and low temperatures.
- Although the watch is designed to withstand shocks under normal use, it is inadvisable to subject it to hard knocks — rough usage or drops onto hard surfaces —.
- Avoid fastening the band too tightly. You should be able to insert your finger inside the band.
- Clean the watch and bracelet with a soft cloth, dry or moistened with mild soap. To avoid surface damage, never use volatile chemicals (such as benzine, thinners, spray cleaners, etc.).
- Gold plated surfaces can be kept in good condition by regular wiping with a soft damp cloth. Discoloration can be removed with detergent.
Always keep unused watches in a dry place.

* Avoid exposing the watch to strong chemicals such as gasoline, cleaning solvent, aerosol spray, adhesive agent, paints, etc. whose chemical action will destroy the seals, case and finish.

Specifications

Accuracy at normal temperature: ± 15 seconds per month
Display capacity:
* Regular timekeeping mode
  Analog: Hour, minute and second hands
  Digital: Hour, minute, second, am/pm, month, date, day
  Time system: Changeover between 12/24-hour format
  Calendar system: Auto-calendar pre-programmed until the year 2019
* Calculator mode
  8 digits (7 digits for negatives)
  Abilities: Four basic calculations, chain & mixed operations, constants for + / − / × / ÷
  Overflow check: Indicated by the "E" sign, locking the calculator mode
* Stopwatch function
  Measuring capacity: 59 minutes 59.9 seconds
  Measuring unit: 1/10th of a second
  Measuring modes: Normal time and net time

* Daily alarm
* Hourly time signal

Battery:
  One silver oxide battery (type: No. 396)
  Approx. 12-month on No. 396 (under following condition: alarm — 20 seconds/day, calculation — 10 minutes/day)

NOTE: THERE IS NO WAY unit components can be damaged or malfunction, due to misoperation of buttons. If confusing information appears on the display it means entry sequence was incorrect. Please read the manual and try again.

Warranty Certificate

THIS WARRANTY CERTIFICATE IS VALID ONLY FOR SERVICE IN THE COUNTRY OF PURCHASE.

Should this watch malfunction under normal use, it will be repaired without charge for a period of one year from the date of purchase. If the watch requires service within the warranty period, request repair or adjustment at the store where purchased or the authorized Casio watch distributor, presenting the watch together with this warranty certificate. The customer shall not have any claim under this warranty for repair or adjustment expenses if:

(1) The trouble is caused by improper, rough or careless treatment.
(2) The trouble is caused by a fire or other natural calamity.
(3) The trouble is caused by improper repair or adjustment made by anyone other than the authorized Casio watch distributor or its retailers.
(4) The case, band, glass or battery is damaged or worn.
(5) This warranty certificate is not presented when requesting service.
(6) The name and address of the authorized distributor or the retailer are not stamped in the warranty certificate.
(7) The date of purchase, model name and manufacturing number are not entered in the warranty certificate.

* The above warranty applies to regions other than the United States of America.
* United Kingdom — This undertaking is in addition to consumers statutory rights and does not affect those rights in any way.

CASIO

CASIO ELECTRONIC WATCH LIMITED WARRANTY

This product, except the case (including buttons), band and battery is warranted by Casio Inc. to the original purchaser to be free from defects in material and workmanship under normal use for a period of one year from the date of purchase. During the warranty period, and upon proof of purchase, the product will be repaired or replaced (with the same or similar model) at our option, without charge for either parts or labor at a Casio Repair/Parts Center listed on this card. There is a $4.95 charge for handling, postage, and insurance. Please enclose your check or money order payable to Casio Inc., when returning your watch for any repair. The warranty will not apply to this product if it has been misused, abused or altered. Without limiting the foregoing, leakage of battery, bending or dropping of the unit, or visible cracking of the LCD display are presumed to be defects resulting from misuse or abuse.

NEITHER THIS WARRANTY NOR ANY OTHER WARRANTY EXPRESS OR IMPLIED, INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, SHALL EXTEND BEYOND THE WARRANTY PERIOD. NO RESPONSIBILITY IS ASSUMED FOR ANY INCIDENTAL OR CONSEQUENTIAL DAMAGES, INCLUDING BUT WITHOUT LIMITING THE SAME, TO MATHEMATICAL ACCURACY OF THE PRODUCT. SOME STATES DO NOT ALLOW LIMITATIONS ON HOW LONG AN IMPLIED WARRANTY LASTS AND SOME STATES DO NOT ALLOW THE EXCLUSION OR LIMITATION OF INCIDENTAL OR CONSEQUENTIAL DAMAGES, SO THAT THE ABOVE LIMITATIONS OR EXCLUSIONS MAY NOT APPLY TO YOU. This warranty gives you specific legal rights, and you may also have other rights which vary from state to state.

CASIO INC. 15 Gardner Road, Fairfield, NJ 07006
CASIO SERVICE CENTER
National Repair/Parts Center 175 Route 46 West, Fairfield, NJ 07006
For information on other Authorized Service Center, please call: 201-575-5695

24

Estimado Cliente:
Felicitaciones por la compra de este Reloj Digital de Cuarzo Casio. Aunque se trata de un reloj altamente sofisticado, resulta fácil de usar, y remitiéndose a esta guía Ud. podrá disfrutar todas las características de esta pieza ultra-precisa.

Nota: Como este reloj cuenta con componentes electrónicos de precisión, se recomienda que la cubierta posterior sea extraída únicamente por personal autorizado.

25

UDC 621.371.029.55: 621.376: 621.394: 681.32
Indexing Terms: H.f. radio communication systems, Modems, Microprocessor application

Bit-slice microprocessors in h.f. digital communications

S. D. SMITH, B.Sc.,*
P. G. FARRELL, B.Sc., Ph.D., C.Eng., M.I.E.E.*,
K. R. DIMOND, B.Sc., Ph.D., C.Eng., M.I.E.E.*

Based on a paper presented at the IERE Conference on Microprocessors in Automation and Communications held in London in January 1981

SUMMARY

A 2·4 kbit/s baseband modem is being designed for use at h.f., incorporating modulation/demodulation techniques that are matched to those frequencies and the problems associated with them. Fed by a continuous serial data stream, the modulator functions are implemented wholly by a bit-slice microprocessor, and controlled by another more conventional microprocessor. Analogue output waveforms are generated in a d/a converter, which is driven by the bit-slice machine. Demodulation is performed in a similar device, using an a/d input and giving a serial output.

*Electronics Laboratories, University of Kent at Canterbury, Canterbury, Kent CT2 7NZ

1 Introduction

Over the past ten years, devices for transmission and reception of data have become more digital in their realization. Not only are these devices constructed with more digital circuitry, but also signals hitherto transmitted on analogue schemes have been modulated digitally. This mode of transmission requires a modem which will convert the baseband signal into a form suitable for transmission. Design of modulators/demodulators which convert between data streams and waveforms suitable for specific types of channel has accelerated in recent years, one such channel being h.f. radio.

This paper describes a modem of this type, which has been designed at the University of Kent for use specifically on voiceband channels at h.f., and also discusses methods of realization. The modem is fed by a serial data stream at 2·4 kbits per second, which it modulates into a 3 kHz baseband channel. In the receiver, after mixing down to baseband, the second half of the modem uses the incoming signal to synchronize, and demodulates it back into a serial data stream (Fig. 1). The modulation technique employed for this system is multi-channel four-phase differential p.s.k.,[1] both with and without pilot synchronization tones inserted in the band. Although other modulation schemes are under consideration to demonstrate the versatility of the modem, this technique is the one to be used at h.f. trials.

H.f. transmission and reception have special problems associated with them. This is because h.f. channels are usually ionospheric and therefore suffer from multi-path propagation and both man-made and natural interference, properties which can cause unpredictable loss of data and synchronization. Unless modem parameters such as data rate or bandwidth are altered, little can be done to prevent loss of data. Loss of synchronization on the other hand results in an additional increase in data errors which can to some extent be controlled. Hence synchronization and the approach for its implementation have been under careful scrutiny in the design of the demodulator.

Until quite recently, nearly all modems would have consisted largely of analogue circuitry with a digital interface to the data source or sink. Utilizing microprocessors enables the construction of modems which are completely digital with just an analogue interface to the communication channel. The most obvious advantage in this case is the increased versatility of the modem. Whereas before, to change modulation type would have needed a major reconstruction of the hardware, the microprocessor realization reduces the problem to a modification in the program which it executes.

2 Operation

In the modulator, incoming data are packed into bytes which are used two or four at a time to provide sixteen or

0033–7722/81/060299+03 $1.50/0

© 1981 Institution of Electronic and Radio Engineers

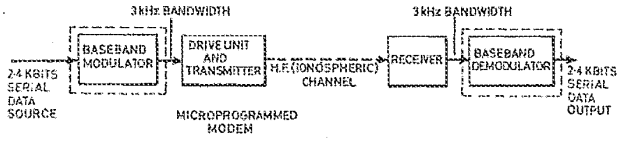


Fig. 1. Schematic diagram of the modem.

thirty-two channels of parallel information. These blocks of data are modulated by a repeating real-time programme with period τ equal to 1/16th or 1/32nd of the incoming serial data rate, into sixteen or thirty two parallel q.d.p.s.k. channels all placed side-by-side in the 3 kHz baseband. Each channel is separated from its neighbour by $2/\tau$ Hz and is also at a multiple of the frequency $2/\tau$. In the 16-channel case, eight carriers each at multiple of the frequency 300 Hz are phase modulated, carrying two bits of information on each of four 90° spaced phases (Fig. 2). In the thirty-two-channel case, sixteen carriers are modulated at a time, but the period τ is doubled too.

The individual carrier signals are generated from sine look-up tables, similar to those described in Ref. 2. These tables are sampled, scaled and summed, depending on the required frequency and phase, every 1/9600th of a second. 128 samples at each frequency of the carriers are derived from the tables at the requisite phases, and summed to obtain 128 samples for transmission. Another two or four bytes are taken from the incoming data stream and used to calculate the new phases for each carrier, so that the whole cycle may begin again. The resultant samples are clocked through a d/a converter to produce the baseband modulated waveform (Fig. 3).

The demodulator, which has to contend with synchronization and error decisions, is more complex than the modulator. (Error decisions consist of resolving the polarity of incoming data into its most likely state, and possibly implementing any error detection/correction that might have been coded into the data.) The noise-corrupted incoming signal is sampled by an a/d converter at 9·6 kbaud. Samples are used in a synchronization algorithm which is arranged to provide the start pulses to a Fast Fourier Transform (F.F.T.) routine. Output from this gives the phase and amplitude of each carrier, which may be compared with the previous phase and amplitude of the same carrier to regenerate the two bits of data.

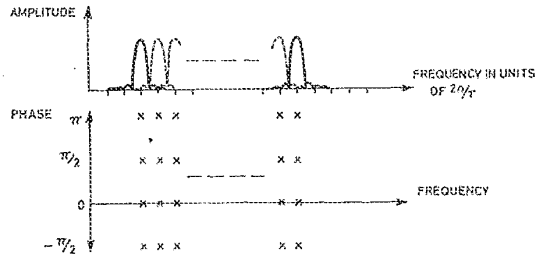


Fig. 2. Amplitude and phase spectrum for multi-channel 4-phase d.p.s.k.

Consider the sixteen carrier situation.

The incoming data from the a/d converter consist of amplitudes of an analogue waveform sampled at discrete intervals of 1/9600th of a second. Without noise, this analogue signal is a sum of sixteen sine waves of equal amplitudes at four possible discrete phases. At intervals of 1/32nd of the data rate (i.e. 75 Hz) the phase of each carrier might change by multiples of 90°, depending on the two new bits of data it carries. Assuming it is highly probable that at least one of the carriers will change phase at every discontinuity, it is possible to gain data synchronization from the phase transitions. An output from this synchronization is used to keep an F.F.T. in step with the incoming data. A double 64-point radix-2 F.F.T. routine[3,5] is applied to each block of 128 samples to produce two frequency domain samples for each carrier frequency. These are averaged and converted from complex coordinates to amplitude and phase coordinates from which not only the data may be determined, but also the rate of fading of the incoming signal and the frequency/phase shift caused by h.f. interference.

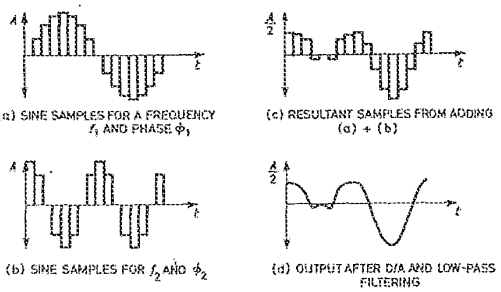


Fig. 3. Example of sine sample summation for two carriers.

Were the F.F.T. to take its 128 samples so that a phase discontinuity boundary was somewhere in the middle, the resulting data would be completely useless. In fact the errors rise fairly quickly with the number of samples at the wrong side of a phase transition, so it is essential that there is good data synchronization. This requires accurate data rate recovery from the incoming signal, which is achieved by a sliding filter algorithm in association with a local 'flywheel' clock. Whether this local clock or the generated synchronization pulses are used to synchronize the transform depends on the depth of fade or the frequency/phase error, as ascertained from previously decoded data blocks.

An additional technique is available for improving

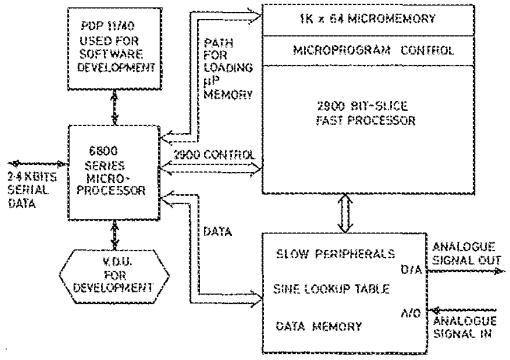


Fig. 4. Modem structure (block diagram).

synchronization, and for minimizing the possibility of performing an F.F.T. across a discontinuity. This involves spreading the carrier frequencies out so that they cover the complete bandwidth of the voice channel, rather than their frequencies being integral multiples of the data rate. The sampling frequency of the receiver is increased proportionately so that it is still an integral multiple of the carrier frequencies. However, the period between phase discontinuities in the transmitted signal remains a simple fraction of the data rate. Hence the period during which the 128 samples are taken for the F.F.T. is shorter than the time between discontinuities by approximately 16% for a data rate of 2·4 kHz in a 3 kHz bandwidth.[4] This means that there is a fairly long period of time, across the phase transition, over which no samples are taken for use in the F.F.T. This is advantageous for two reasons: (a) to allow a greater margin for synchronization error before trans-discontinuity samples cause errors in the F.F.T. algorithm, and (b) the F.F.T. does not employ samples near to the discontinuity where the 3 kHz bandlimiting causes 'rounding' of the signal on either side.

3 The Modem Structure

In both the modulator and the demodulator there are two microprocessors. A slower, one-chip microprocessor from the 6800 family is used to interface the modem to the serial data source or sink. Its responsibility is for the slower data processing, such as packing the incoming serial data into bytes and encoding it, some of the synchronization mechanism in the demodulator, and the control functions for the fast processor. (Fig. 4.)

This fast processor consists of a 2900-series bit-slice microprocessor to perform the modulation and demodulation of data, and is connected directly to the analogue port via its data bus. Its purpose is to convert data to samples of summed sine waves at the correct phases in the modulator, and to perform the F.F.T. and clock recovery in the receiver. Since as a modulator/demodulator it is repeatedly executing a dedicated routine of known duration, there is no need for macro-coding and a mapping p.r.o.m. as in the conventional bit-slice machine.[6] Hence all programming is at the microcode level. Microcode is bootstrapped into the writable microcode memory on power-up by the 6800 processor, which in turn is fed by a host mainframe computer during microprogram development. For a completed portable modem, the bootstrapping is from e.p.r.o.m.s in the 6800's memory map.

All data/address buses on the bit-slice are 12 bits wide, together with the a/d, d/a converters, while the width of the microprogram word is 64 bits. Two-level pipelining and parallel hardware stacks, together with fast data paths and devices isolated from slow data buses by registers allow minimization of processor cycle times. The bit-slice machine is connected to the slow processor by an 8-bit bidirectional data register which is directly addressable in the memory map of each machine.

4 Conclusions

This paper describes a modem which uses only digital processing to accomplish its operation. When used for h.f. trials the modem demonstrates the viability of microprocessor controlled modulation and demodulation. It also reveals its versatility to be reprogrammed with ease to a completely different modulation scheme.

5 Acknowledgments

S. D. Smith would like to acknowledge the support of the S.R.C. and of the Ministry of Defence (Procurement Executive); the authors are grateful for helpful discussions with Mr J. Pennington (A.S.W.E.)

6 References

1 Ziemer, R. E., and Tranter, W. H., 'Principles of Communication: Modulation and Noise', Sect. 7.5 (Houghton Mifflin, Boston, 1976).
2 Gorski-Popiel, J. (Ed.), 'Frequency Synthesis: Techniques and Applications' (IEEE Press, New York, 1975).
3 Rabiner, L. R., and Gold, B., 'Theory and Application of Digital Signal Processing' (Prentice Hall, Englewood Cliffs, N.J., 1975).
4 Riley, G. I., 'Error Control for Data Multiplex Systems', Ph.D. Thesis, University of Kent at Canterbury, 1975.
5 Brigham, O., 'The Fast Fourier Transform' (Prentice Hall, Englewood Cliffs, N.J., 1974)
6 'Build a Microcomputer', Advanced Micro Devices, Sunnyvale, Cal., 1979

Manuscript received by the Institution in final form on 27th March 1981 (Paper No. 1993/Comm 220)

A Novel use for Microprocessors in Designing Single and Multi-tone Generators, Ringing Generators and Inverters

Authors

Earl Rhyne
President
Permace Associates
7 Walnut Street
Millis, MA 02054

Ray Bennett
Engineering, Consultant
Permace Associates
21 Rickey Drive
Maynard, Ma 01754

ABSTRACT

This paper describes a method for producing precise single or multi-frequency tones for telephone office ringing generators, tone generators, and general purpose inverters, by using microprocessors and digital technology.

Recent technical developments have provided engineers new tools for generating the signals used for telephone equipment and for permitting remote access to the equipment for supervisory and diagnostic purposes. Figure 1 illustrates a system in which microcontrollers, counters,timers, Random Access Memory (RAM), Analog to Digital Converters (A/D), and Digital to Analog Converters (D/A) are combined to produce tone signals. These signals are then amplified to produce the required ringing or tone power.

The microcontroller is programmed with a mathematical equation to derive timing and voltage levels for the output signal. This equation is converted to digital words that are passed to the data portion of RAM and is converted to a digital word that sets a timer controlling the address portion to RAM. The RAM output is converted to an analog signal by the D/A converter. This signal is used by a power amplifier to condition the signal for use on the telephone lines.

By using a Microcontroller, the terms of the equation (i.e. frequencies and the voltage levels of each frequency independently) are input as variables, thus giving the user complete control of the output signal. The user can use a manual control (such as a keypad/readout) or a remote computer (through an RS232 port) to change the variables. The range of the signals is determined by the resolution of the D/A converter and the frequency response of the power amplifier. Because the Microcontroller is crystal controlled, frequency response and accuracy are a function of binary resolution. The output level is also a function of the binary resolution and reference voltage. Multiple frequency tones are generated by the same method. Since the quantizing frequency is much higher than the tone frequencies, simple low pass filtering is used to eliminate unwanted frequencies. Using an A/D converter, the output is monitored. This same signal may be used as a built-in diagnostic test.

The output of the power amplifier is sensed for voltage and current output. In low frequency applications, the microcontroller can monitor for inductive and capacitive loads, and make adjustments.

The versatility of the microcontroller allows a single design to cover a wide variety of uses. The power amplifier can customize the application. Other options such as zero crossing interrupting would be under control of the microcontroller.

INTRODUCTION

Currently Ringing and Tone generators consists of analog devices such as oscillators and linear amplifiers, or non-adjustable digital oscillators and bandpass filters to generate the required frequency and waveshapes. Some of the more important analog design considerations are signal linearity, symmetry, frequency stability, and temperature variations that effect all of the above. Frequency is derived from standard oscillator circuits, which contain resistors, capacitors, and/or inductors. In adjusting the frequency, fine tuning is done with variable resistors, while more coarse adjustments are done by switching capacitors and/or inductors. The frequency selective components used in ringing generators have large values and are physically large.

Some of the more important digital design elements are the master clock and count down circuits. The use of digital circuits usually solves the problem of stability and symmetry but introduces some filtering requirements because digital signals are square waves which, by definition, are rich in harmonics. Proper filter design reduces harmonics to the desired output level. Digital filters at ringing frequencies contain large capacitance and/or inductance values and, like the analog oscillators, are physically large. A resonant transformer may be used as a filter but it is physically large, and can only work with a single frequency. The filters for tones need to have high Q values in order to suppress the harmonics below the DTMF band requirements. In the current technology, interrupters are not synchronized to the tone wave shapes being interrupted.

Today's technology allows the use of Microcontrollers to generate signals. Microcontroller generate the required data and D/A converters transform the digital words to analog signals, eliminating the need for RC or LC oscillators. All the required functions are executed in firmware which calculates the sine functions for magnitude and duration of wait statements necessary to obtain the frequency. Since the microcontroller is driven by a crystal oscillator, frequency stability can be as good as a standard quartz watch (in the order of .002%). Frequency and voltage adjustments can be made by recalculating Microcontroller data. The user may input the data in many ways, such as a key pad, selector switches, or it may be down loaded via a RS232 computer/modem port.

DIGITAL SYNTHESIS

Ringing and tone analog signals are independent, continuous signals varying as a function of time. Digital signals are discrete time varying signals. A digital signal is a sequence of numbers[1].

CH2928-0/90/0000-0611$01.00 ©1990 IEEE

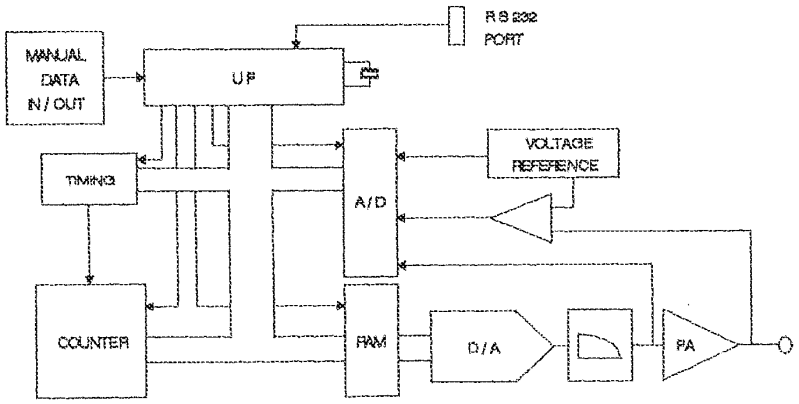


Figure 1

Each digital number represents a time variant value of an analog signal called sampled data. Sampled data is a discrete value for a sampled time. The following sampled data words are analog signals at different values in both time and magnitude. This quantized signal takes on only those values specified by the quantized levels. The quantized signal differs from the analog signal by the number of individual points taken during the duration of the analog signal. The more points taken over a given period of time, the smaller are the errors introduced by the digitizing of the signal.

Ringing and tones are repetitive, time varying signals, making their mathematical models quite simple. They lend themselves to simple calculations. Allowing the Microcontroller to generate discrete samples of data over an integral number of one or more repetitive "tone cycles".

To convert the mathematical numbers to an analog signal, a D/A converter is used. Its input is a digital word and its output is a voltage level corresponding to that word. As each new different word is applied to the D/A converter the output varies accordingly. The result is a time varying signal; but this signal is not continuous because of the quantized affect.
See figure 2.

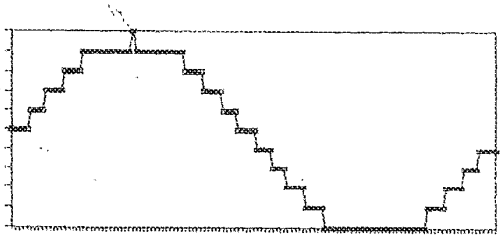


Figure 2.

The computer generated mathematical formula is turned into a digitized analog signal. This analog signal has quantized error terms that cause distortion with respect to an analog generated signal. If the correct number of points are chosen the error terms are low. These error terms contain high frequency signals that are easily filtered out with a simple bandpass filter. This process yields a signal that is amplified and applied to the telephone circuits.

Dial tone and ringback tone are produced by adding together two precise frequency tones. Ringback tones comprise 440 Hz and 480 Hz; dial tones comprise 350 hz and 440 Hz. The "Tone Cycle" is defined as follows: with all frequencies starting at zero degrees phase angle (zero volts and zero current), a "Tone Cycle" is complete when all the tones arrive at a 360° phase angle at the same time (i.e. zero volts). For a single frequency tone, quantized data is calculated for only one 360° cycle. For ringback tone, the computed data includes eleven complete 360° cycles of 440 Hz with twelve complete 360° cycles of 480 Hz to complete one "Tone Cycle". For dial tone, the computed data includes thirty five complete 360° cycles of 350 Hz with forty four complete 360° cycles of 440 Hz to complete one "Tone Cycle". After incrementing and transferring one "Tone Cycle" to the D/A converter, the Microcontroller resets to the start of the "Tone Cycle" and repeats.

Because the error terms in the digital generated dual tones are high frequency, a lowpass filter will leave only the two fundamental frequencies to be linearly amplified and distributed to the telephone circuits.
See figure 3.

Equation 1 produces a single sine wave frequency.

$$e = E*\sin(w*tq) \quad (1)$$

Where e = output signal
E = peak output voltage
w = 2*pi*f
tq = quantized time
tq = (1/f)/points

Equation 2 produces dual sine wave tones.

$$e = E_1*\sin(w_1*tq) + E_2*\sin(w_2*tq); \quad (2)$$

Where
E_1 = peak output of f_1
E_2 = peak output of f_2
$w_1 = 2*pi*f_1$
$w_2 = 2*pi*f_2$

The quantized time is

$$tq = (1/f_1)/points$$

Example:

If 512 data points are chosen and a ringing frequency of 20Hz is needed then,

$f = 20Hz$

$tq = (1/20)/512 = 98us.$

$fq = 1/98E-6 = 10.2Khz.$

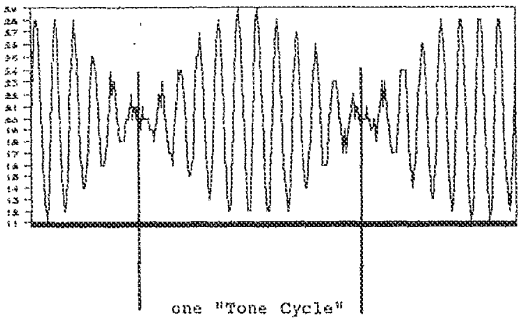


Figure 3

This quantizing frequency is twice the upper limit of the DTMF band. Filtering 10 Khz to a level of 50 dB below the fundamental can be done by a typical Chebyshev filter.

A simple listing for a single frequency is as follows:

```
 /* GET VARIABLES */
input frequency, f1;
input voltage, E1;

/* DEFINE CONSTANTS
 two_pi = 2*pi */
two_pi=6.283185307;

/* CALCULATE VARIABLES */
t1=((1/f1)/512) - overhead;
Mloc=0;
E = E1*10;

/* MAIN */
while (Mloc<=512)
   {
/* GET NEXT LOCATION */
     t=t1*Mloc;
/* OUTPUT DATA */
   P1= E*sin(two_pi*f1*t);
/* LOAD TIMER */
   TL0 = low(TICK);
   TH0 = high(TICK);
   TR0 = 1;
/* INCREMENT COUNTER */
   Mloc++;
/* START TIMER */
   while (TF0=0); TF0 = 0;
/* WAIT FOR TIMER & RESET*/
   };      /* END MAIN */
```

The value t1 is the time between quantized points. P1 is the output port attached to the D/A. Overhead is the time it takes the microcontroller to execute the instructions.

DIGITAL SYNTHESIZER

We have now identified one mathematical approach to produce a single frequency signal. One way to implement this in hardware is to use microcontrollers. The microcontroller contains timers and RAM. The instructions, also referred to as firmware, are contained in a ROM connected to the microcontroller's data bus. One data port of the controller is directly connected to the D/A converter, and the D/A converter is connected to the filter. See figure 4.

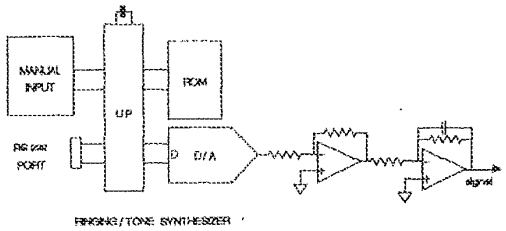


Figure 4

Installed in firmware are the calculations for a normalized sine function and the normalized timing functions. The input parameters of frequency and output level are independent functions, and are not in the main loop that generates the output signal. The input parameters are used by the calculator portion of the main loop that produces the output signal. From this, a signal containing the programmed frequency(s) and output level(s) is produced. The frequency is converted to time and is then loaded into one of the microcontroller's timers. The output value is calculated using the D/A parameters along with the loss of the filter and the gain of the output amplifier.

The firmware then takes the first normalized value from the sine calculation and algebraically adds the value of the programmed output voltage, sets the timer and passes this value to the D/A for conversion to output analog signal. The next step in the sequence, increments the count and repeats the

function. The repeated functions are put on the data bus when time t1 has elapsed. To insure t1 is strictly adhered to, the timer will interrupt the controller at the end of its programmed time. The analog value is zero and the address counter is reset to zero when the count reaches it's limit; then the cycle repeats until the controller is reset or powered down.

Multi tones may be generated by algebraically adding two computed sine functions together and passing the data to a single D/A converter. A second method is to use two independently generated sine functions and pass the individual data to two A/D converters, then sum the analog signals. This method requires the use of separate timers to keep track of the individual times of each sine function.

OUTPUT VALUES

The full scale value of a typical D/A is plus and minus five volts. In binary, the plus full scale for the D/A is 11111111 (FF in hex) and the minus full scale is 00000000 (00 in hex). The firmware must convert the value in the 128th (first peak value of the sine) location to FF (hex). The same conversion value is then used for each of the 512 sine values as they are to be loaded into the A/D.

Resolution is the function of the number of parallel data bits used by the D/A converter. For example, if an 8 bit D/A converter is used, the resolution is 1/256 times the full scale value. If the analog full scale value of the D/A converter is 5v then the resolution is equal to 5/256 or 19.5 millivolts. This is the smallest value of change allowed for this signal. The resolution is approximately 0.4%. By using a larger input D/A (i.e. 12 bits) the resolution is lowered to approximately 0.03%. The digital word for each analog value is calculated using equation 3.

$$V = N*R \qquad (3)$$

where

V = output volts
N = number of steps
R = resolution in volts

If the filter and amplifier have a combined gain of unity, the input parameters are the only multipliers.

Example:

To generate a 2.5v rms signal at the output of the 8 bit D/A converter,

Vrms/.707 = Vpk

2.5V/.707 = 3.54Vfs

2^8 = 256 steps FS

R = 5/256 = 19.53mv

N=V/R = 3.54/.0195 = 182

182 decimal is B6(hex)

Thus on the 128th count the D/A is to be loaded with B6 (hex).

Filtering is necessary because the output of the D/A converter still contains the quantizing frequency. A quantizing filter is a low pass filter with high enough value of Q to allow the desired signals to pass unattenuated, but attenuate the quantized frequency to a value at least 50 Db below the fundamental. A two or three stage Chebyshev filter is all that is required if the quantizing frequency is separated by two or three orders of magnitude.

The resulting signal is then treated as a standard analog signal and may be amplified by many means, such as a linear amplifier.

FEATURES

With a microcontroller calculating and generating the data for each step and count, it is possible to start and stop the signal on a data boundary (typically zero volts). It is also possible to modify the frequency and the output level by modifying data words, this allows for stable non-component dependent signals. If a host computer is connected to the generator it is possible for the microcontroller to collect operating data and diagnostic data (such as load peaks with respect to the time of day and active operating data). With the use of a battery backed up clock, the time of unscheduled interruption (failures etc.) may be logged, and the host computer may be used to trouble shoot the faulty equipment. Modems make it possible for remote sites to be monitored and data to be logged.

An interrupter can be included in the same package by adding appropriate hardware, and driving it by microcontrollers. Since microcontrollers are controlling both ringing generator and tone generator as well as the interrupter, it can synchronize them for zero voltage and zero current interruptions. Before the interrupter makes or breaks the signal, the microcontroller will first allow the output signal to finish its cycle to zero, then shut off the generator allowing the output voltage and current to go to zero. After waiting for transient settling, it opens or closes the interrupting relay. After the relay switching time has elapsed, the controller restarts the generator at zero phase angle and zero volts. A non-current breaking interruption has taken place.

CONCLUSIONS

Digital to analog technology is now a mature process and is supplemented with many pre-packaged circuits. The combining of functions within packages and the small physical sizes of the packages make them a viable solution to existing requirements. As the usage of these complex packages becomes more widespread, the prices reduce, and the variations increase. This gives today's designer a broad spectrum of ideas to chose from to make the telephone equipment more compatible with the present day technology.

REFERENCES

1. Lawrence R. Rabiner and Bernard Gold, "Theory and Application of Digital Signal Processing",Prentice-Hall, Inc. Englewood Cliffs, N.J.

Capacitive Impedance Readout Tactile Image Sensor

R. A. Boie

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT The transduction of mechanical forces to representative electrical signals uses a three layer sandwich structure. The top layer is columns of compliant metal strips over a central elastic dielectric sheet. The bottom layer is a flexible printed circuit board with rows of metal strips and multiplexing circuits. Electrically, the sensor is a capacitor array formed by the row and column crossings with the middle layer functioning as a dielectric spring. A readout of the capacitor values corresponds to a sampled tactile image.

The reasons for choosing this transduction method, the performance advantages of capacitive sensing and the design and integration of 64 element imagers into the fingers of a controlled compliance gripper are described.

1. INTRODUCTION

A review of touch or tactile sensor technology is given by Harmon[1]. Several sensor designs, including the one reported here, are based on measuring the thickness of an elastic layer compressed by the applied force. Resistive readout sensors of this class use conductive loading and obtain the pressure map by cross layer resistance measurement[2]. The method is inherently non linear and the materials exhibit poor elastic properties including hysteresis.

Cross layer capacitive impedance sensing is more favorable in many respects. The elastic materials need not be modified and desirable mechanical properties are generally consistent with low dielectric loss. Capacitive sensing is demonstrated to have marked advantage in terms of signal to noise ratio and measurement speed. The idea of force distribution sensing by capacitive readout and a study of suitable elastic/dielectric materials are presented in a comprehensive paper by Nicol[3]. The main contributions here are the development of a relationship for noise limited force resolution, illustrating the inherent performance of the sensing method, and the development of an appropriate robotic sensor.

2. CAPACITIVE SENSING

Figure 1 illustrates an exploded view of a sample robotic touch sensor. The topmost layer is a compliant glove that contacts objects and transmits via its elastic constant the contacting force distribution to the elastic/dielectric layer below. The lower layer is here shown rigidly supported by the printed circuit board. The glove and dielectric layer can be viewed as two springs in series under compression where the force information is obtained by measuring the displacement of the dielectric spring. The mechanical point-spread function of the glove can be narrowed, if desired, by suitably segmenting the glove material.

Orthogonal sets of conductive strips are arranged on the upper and lower surfaces of the elastic layer. A sampling of the layer thickness map is obtained by measuring the array of capacitors formed by the crossing areas, $A_{i,j}$, of row and column strips. The strip widths and spacing along with any point force spreading in the structure determine the spatial sampling and resolution. The time required to measure all capacitors determines the temporal sampling. The r.f. source, $V_0\cos(\omega_0 t)$ is connected to the lower set of strips via analog multiplexer "i". The multiplexer "j" connects pads to the amplifier input node. The pads are capacitively coupled to the upper strips via an inactive region of the elastic/dielectric layer. This contactless arrangement, due to Miller[4], is an important construction feature of this method. Cross talk signals are reduced by connecting the unselected strips and pads to ground potential. For each pair of multiplexer addresses (i,j) the r.f. source voltage is connected through the capacitance C(i,j) of strip i to strip j to the input node of the amplifier. (The strip to pad capacitance is arranged to be sufficiently large.) The output signal of the amplifier, V_A (i,j,t), is related to the strip to strip capacitance by,

$$V_A(i,j,t) = -V_D \frac{C(i,j)}{C_A} \cos \omega_0 t \qquad (1)$$

where C_A is the capacitance in feedback. C(i,j) is related to the localized layer thickness change by,

$$C(i,j) = \frac{K A}{\epsilon_0 (d_0 - x(i,j))} \qquad (2)$$

where A is the strips crossing area, K is the relative dielectric constant, ϵ_0 is the permittivity of vacuum, d_0 is the unloaded layer thickness. The local sampled force is described by the relationship,

$$F(i,j) = \lambda\, x(i,j) \qquad (3)$$

where λ is the dielectric/elastic layer spring constant.

The applied force is linearly related to measures of the reciprocal crossing capacitances with a constraint of fixed layer constants. Each crossing capacitance, independent of the shunt dielectric loss and series switch resistances, is measured in turn by phase sensitive detection during the interval T_m between sequential address advances.

Figure 2 illustrates the measurement method. The signal, V_A(i,j,t), is multiplied by the amplitude limited r.f. drive and integrated over the measurement interval, T_m. The time, T_m is synchronous with and has duration of m cycles of the r.f. drive. The integrator output is sampled and reset and the multiplexers address advanced at the end of each interval.

CH2008-1/84/0000/0370$01.00©1984 IEEE

The sampled output is related to the strip i to strip j crossing capacitance by,

$$V_s(i,j) \propto V_d\ m\ \frac{C(i,j)}{C_A} \tag{4}$$

The force information is related to reciprocals of offset corrected capacitance measurements. Two direct reading readout methods were considered and may prove practical for some sensor designs. A conceptually simple method requires only the circuit location interchange of capacitors C(i,j) and C_A. All else remaining the same, the output provides a measurement of the crossing capacitive impedance. The impedance is linearly related to the displacement and, via the elastic constant, the force. This method requires a high performance input amplifier. The central difficulty is the large loop gain required for linear measurement response over a wide dynamic range. A more robust method is described in a paper on capacitive distance measurement[5].

3. NOISE, RESOLUTION AND DYNAMIC RANGE

Capacitive sensing of mechanical displacements is in most applications the method of choice. The low noise - high bandwidth properties of the method are well known, but little practiced. The method has the virtues of a parametric measurement, that is, the output signal is proportional to the displacement times the drive signal. Capacitors are non dissipative elements and so generate no noise. Capacitive sensing has not faired well in the robotics literature to date where it is described as inappropriate because of noise[6]. This misconception most likely results from confusing man-made interference, which can be reduced to negligible levels by proper shielding and connection, with intrinsic noise related to the basic nature of the detection process.

Figures 3 illustrate the equivalent circuits used for the performance analysis. Here a simpler receiver and filter are used to better illustrate the performance relationship. Figure 3a illustrates the strip crossing capacitance measurement. The r.f. drive or pump voltage, V_D, is connected to the input node of amplifier, A, via the crossing impedance. The peak output level of the filter with bandwidth Δf and center frequency ω_0 is the measure of the crossing capacitance and thereby the displacement of the dielectric/elastic and the force.

The diagram of Fig. 3b includes the significant parasitic circuit elements and the amplifier noise sources referred to its input. The resistances, R_{si} and R_{sj}, represent the multiplexers "on" resistances that appear as uncorrelated series noise sources. The resistor R(i,j) represents the dielectric loss, a parallel source. The generators e_n and i_n are the input equivalent series and parallel noise sources of the amplifier. The capacitors C_D, C_s and C_{gm} represent the parasitic elements of the sensor, wiring strays and the amplifier input, respectively. The noise sources and parasitic elements may be combined into equivalent noise resistances R_s and R_p and total shunt capacitance C_T, without loss of generality as shown in Fig. 3c.

The signal to noise relationship is developed in terms of the thickness change δx of the dielectric at a measured crossing. The differential signal output of the filter for a small displacement is;

$$\delta V_s = V_D\ \frac{C_0}{C_T}\ \frac{\delta x}{d_0} \tag{5}$$

where the displacement is described in relationship (2). A sensor array formed of N × N strips has parasitic capacitance C_D, which is by inspection of Fig. 4a proportional to the strip length.

$$C_D \propto N\ C_0 \tag{6}$$

The stray capacitances are not intrinsic to the design and can in practice be made relatively small. The sensor represents a capacitive source to the amplifier. The signal to noise ratio is optimized if the amplifier input element is physically scaled, while preserving its gain bandwidth product, so that C_{gm} and C_D have the same value[7]. The relationship for the optimized configuration is;

$$\delta V_s = \frac{V_D}{\alpha N}\ \frac{\delta x}{d_0} \tag{7}$$

where α is excess capacitance scaling constant. The signal improves linearly with the pump magnitude and degrades by the square root of the total number of array elements.

The mean square output signal of the uncorrelated series and parallel sources may be expressed as,

$$\overline{V}_n^2 = 4kT\ \Delta f\ \left\{ \frac{1}{\omega_0^2 C_T^2 R_p} + R_s \right\} \tag{8}$$

The first term in braces is due to the parallel source. The series term is usually dominant at the measurement frequencies and values of interest. The measurement bandwidth Δf is not of direct interest, more important is the array or frame rate F. That being the case the rms noise limited displacement resolution for a fully multiplexed sensor readout is given by,

$$\frac{\sigma_x}{d_0} = \alpha N^2 \left\{ \frac{\sqrt{4kT\ R_s\ F}}{V_D} \right\} \tag{9}$$

where σ_x is the r.m.s. displacement uncertainty. The term in braces represents the ratio of the series noise to the drive voltages. A conservative value of 1K Ohm for R_s, a drive of 10 volts and a framing rate 100 Hz yields a ratio value of 4×10^{-9}. This translates into a wide available dynamic range that may in turn be advantageously traded for relaxed layer requirements. Increasing the spring constant λ and thereby restricting the total fractional excursion, may help in reducing force dependent effects in the layer constants λ and K.

4. TACTILE IMAGING FINGERS

An 8 × 8 element tactile imager and its finger are shown in Fig. 4. The U shaped flexible circuit board is shown in the lower right of the photograph. The base of the U is the active region. The eight long strips are the driven elements and the eight short strips are the signal coupling pads. The short arm of the U supports the drive circuitry. The other supports the eight amplifiers, one for each pad, and the output multiplexer. The photograph also shows the finger structure and the assembly of the U shaped touch sensor band-aid on the robot finger. A view of the instrumented gripper is shown in Fig. 5. The low loss and backdriveable robot gripper mechanism was developed to support ultrasonic eye in the hand ranging and tactile imaging fingers with independent and variable gripping impedance[8]. The ultrasonic ranging system and the gripper control system are described elsewhere in these proceedings in papers by Miller[9] and Brown[10]. Pressures up to 50,000 dynes/cm^2 are sensed using a two thickness nylon stocking mesh elastic/dielectric layer. Each capacitor of the 64 element array is measured in turn by phase sensitive detection over eight cycles of a 200 KHz r.f. drive for a 390 Hz frame rate. Figures 6b and 6c show photographs of touch sensor raw data, $V_s(i,j)$, in response to touching a 1/4 inch diameter ball. Figure 6a shows the zero force offset image. The position directions "i" and "j" are indicated. Each of the 8 × 8 square areas shown correspond to strip crossings areas of 2.5 mm × 2.5 mm. The displacement out of the picture corresponds to increasing capacitance, C(i,j), and thereby sampled force, F(i,j). Figures 7a and 7b show touch images for the lead ends of an 8 pin dual in line package.

5. DISCUSSION

Capacitive sensing provides a robust and simple method of tactile imaging. The construction is straight forward and uses well behaved materials and catalog electronics. Structurally, the sensors are thin and conformable and are easily scaled. Static as well as dynamic images are sensed with a linear response. The temporal sampling can be made short relative to the mechanical response times of the robot system. The array readout need not be fully multiplexed, all rows may be measured during the time each column strip is driven. The spatial resolution is fundamentally limited only by strip lithography. If warranted, a 32×32 elements finger mounted single chip subsystem with composite video like output could be developed using current technology.

6. ACKNOWLEDGMENTS

I would like to thank G. L. Miller and R. A. Kubli for invaluable comments and suggestions.

References

1. Harmon, L.D., SME Technical Report MSR80-03, (1980)
2. Hillis, W.D., M.I.T. A.I. Memo 629, April (1981).
3. Nicol, K., Transducer Tempcon, (1981).
4. Miller, G. L., Private Communication, (1982).
5. G.L. Miller, R.A. Boie, P.L. Cowan, J.A. Golovchenko, R.W. Kerr, D.A.H. Robinson; A Capacitance-Based Micropositioning System for X-ray Rocking Curve Measurements; RSI, Vol. 50, No. 8, August 1979.
6. Harmon, L.D., SME Technical Report MSR82-02, (1982).
7. Gillespie, A. B., Signals, Noise and Resolution in Nuclear Counter Amplifiers. McGraw Hill. (1953).
8. Hogan, N., "Programmable Impedance Control of Industrial Manipulation", Proc. of Confr. on CAD/CAM Technology in Mechanical Engineering, M.I.T., Cambridge, Ma., March (1982).
9. G. L. Miller, R. A. Boie, M. Sibilia, "Active Damping of Ultrasonic Transducers for Robotic Application".
10. M. K. Brown, "Computer Simulation of Controlled Impedance Robot Hand".

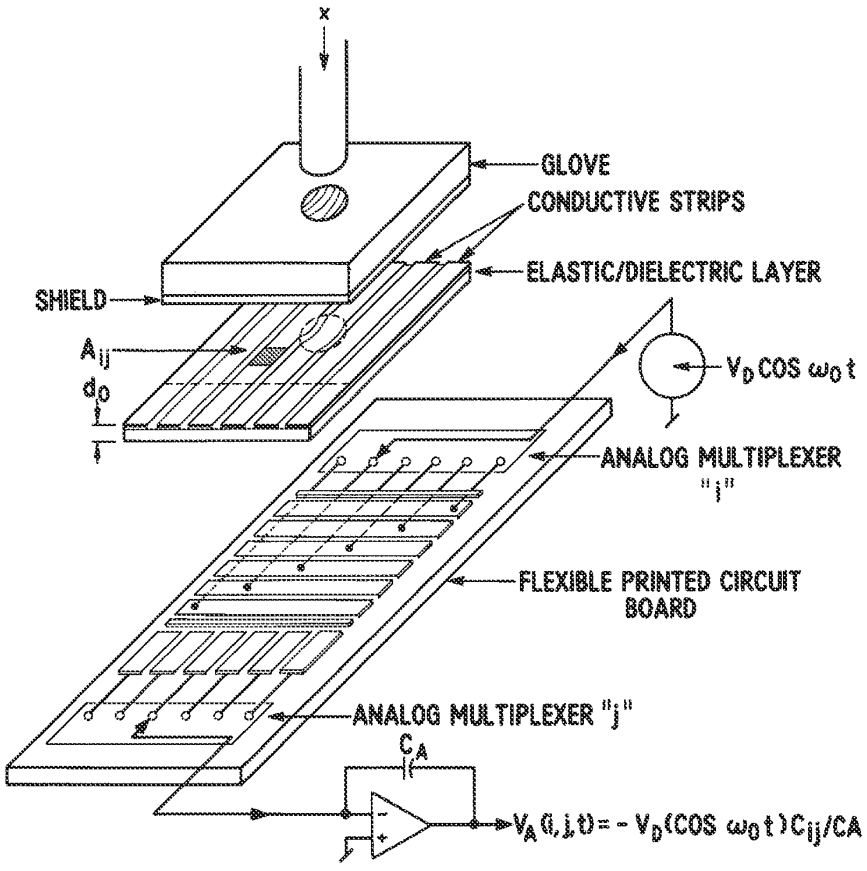


Fig. 1 Exploded view of a sample 6×6 element robotic touch sensor illustrating the layering and contactless construction. The force distribution is obtained by measurement of the cross dielectric/elastic layer capacitance.

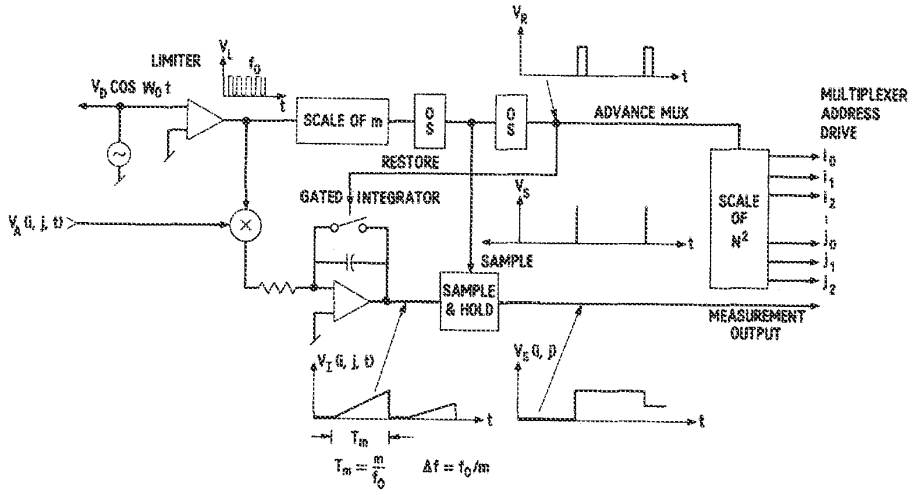
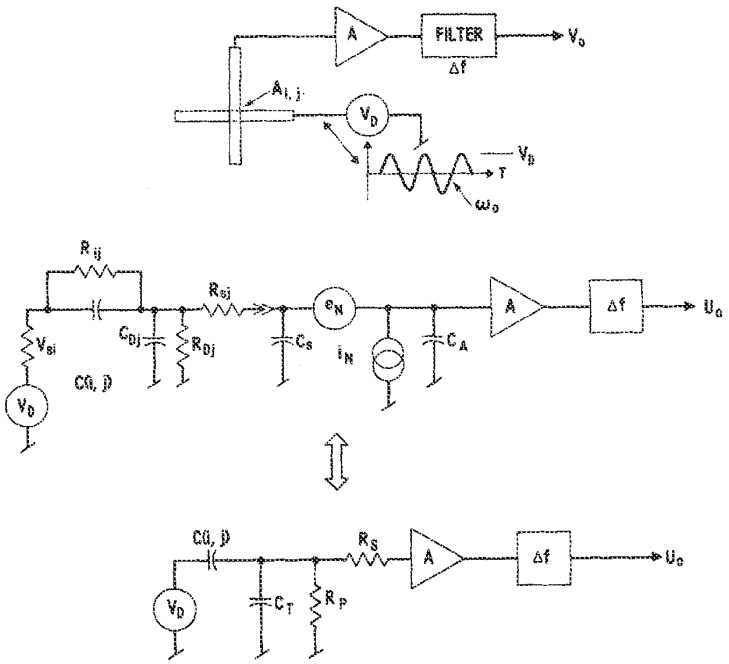


Fig. 2 Block diagram of the touch sensor control and phase sensitive receiver.



SIGNAL V_{OS}

$$C(i,j) \approx \frac{K}{\epsilon_o} \frac{A_{ij}}{d_o(i,j) - x(i,j)} \qquad C_T = \alpha N C_{ij} \qquad N^2 \text{ POSITION ELEMENTS}$$

$$\partial V_{os} \approx V_D \frac{C_o}{C_T} \frac{\partial_x}{d_o}$$

NOISE $\bar{V}_{ON}$ (RMS)

$$\bar{V}_{ON}^2 = \underbrace{4kT R_s \Delta f}_{\bar{V}_{ONS}^2} + \underbrace{\frac{4kT \Delta f}{R_P \omega_o^2 C_T^2}}_{\bar{V}_{ONP}^2} \approx \bar{V}_{ONS}^2$$

RESOLUTION

FRAME RATE $f_R \implies \Delta f > N^2 f_R$

$$\frac{\sigma_x}{d_o} \approx \alpha N^2 \frac{\sqrt{4kT R_s f_R}}{V_D}$$

Fig. 3 Illustrations of equivalent circuits used in the performance analysis.

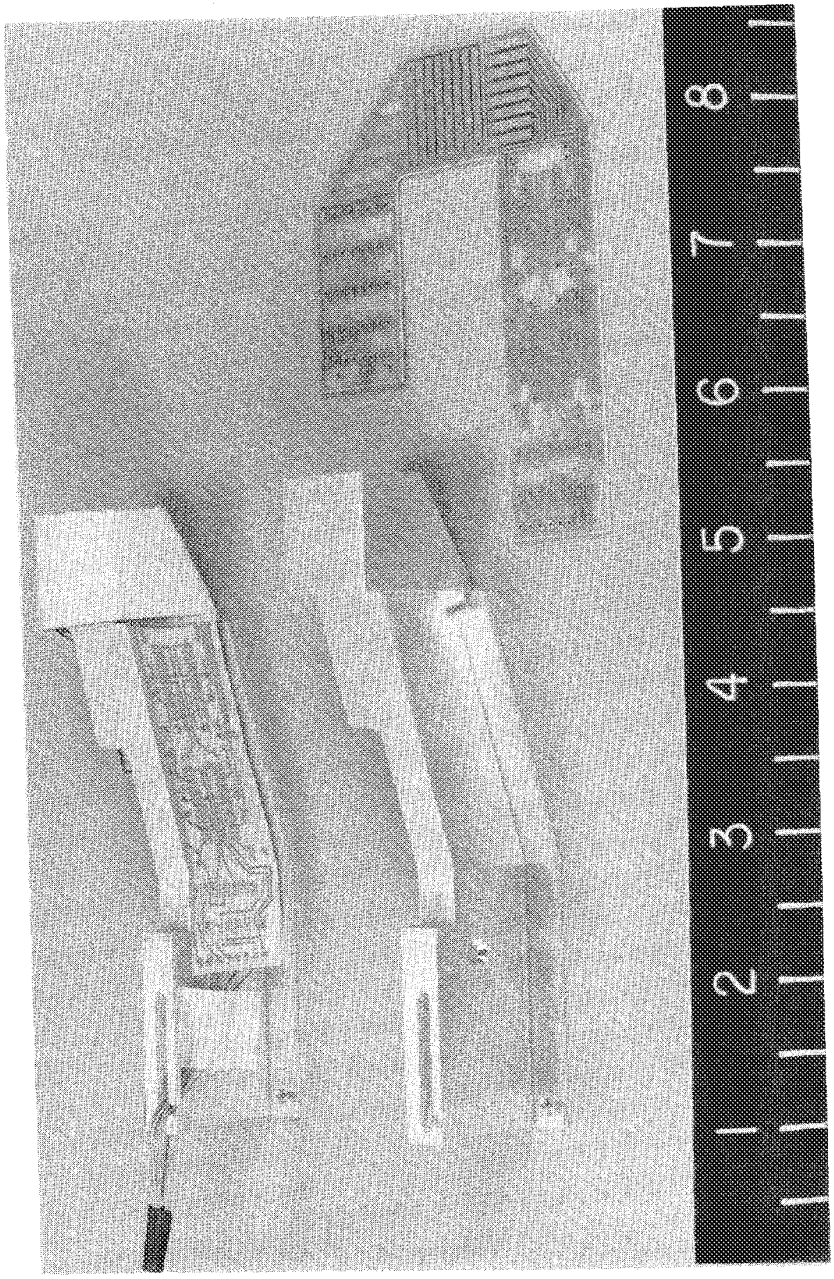


Fig. 4 Photograph showing the tactile sensor flexible printed circuit substrate, the robot finger structure and the active finger assembly.

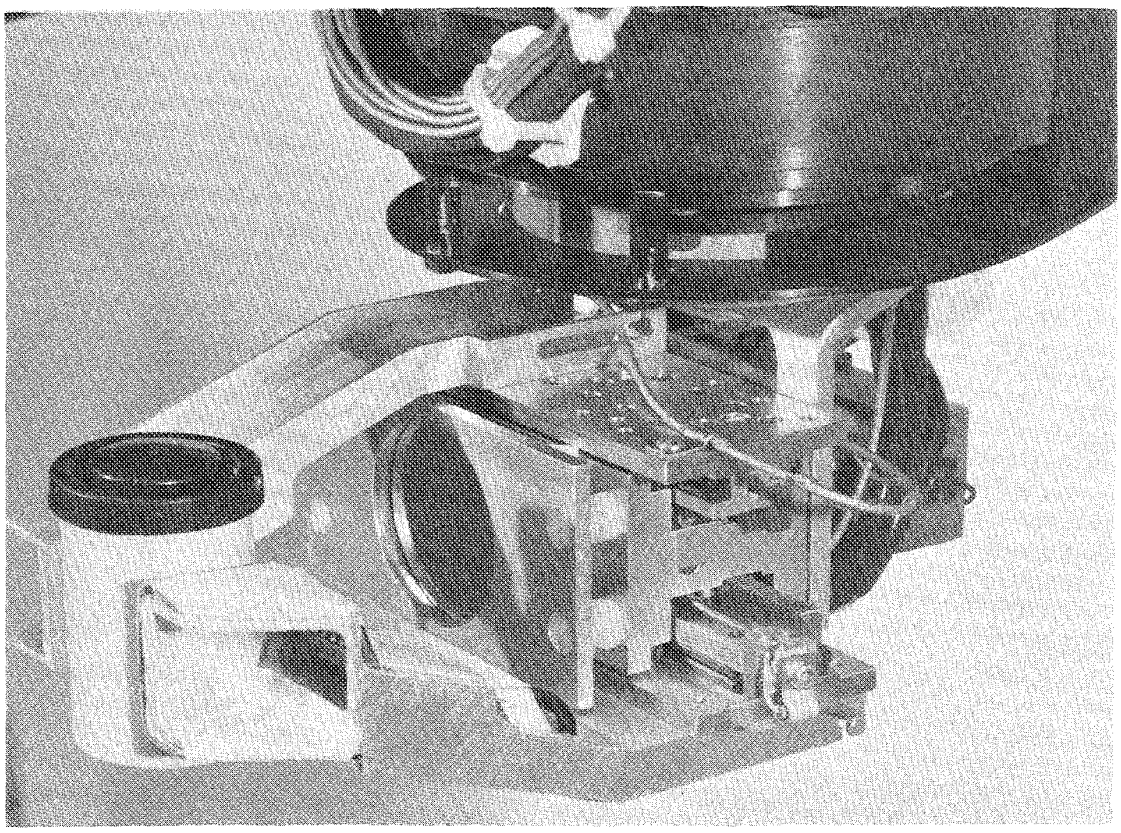


Fig. 5 A view of the instrumented gripper feeling a film container.

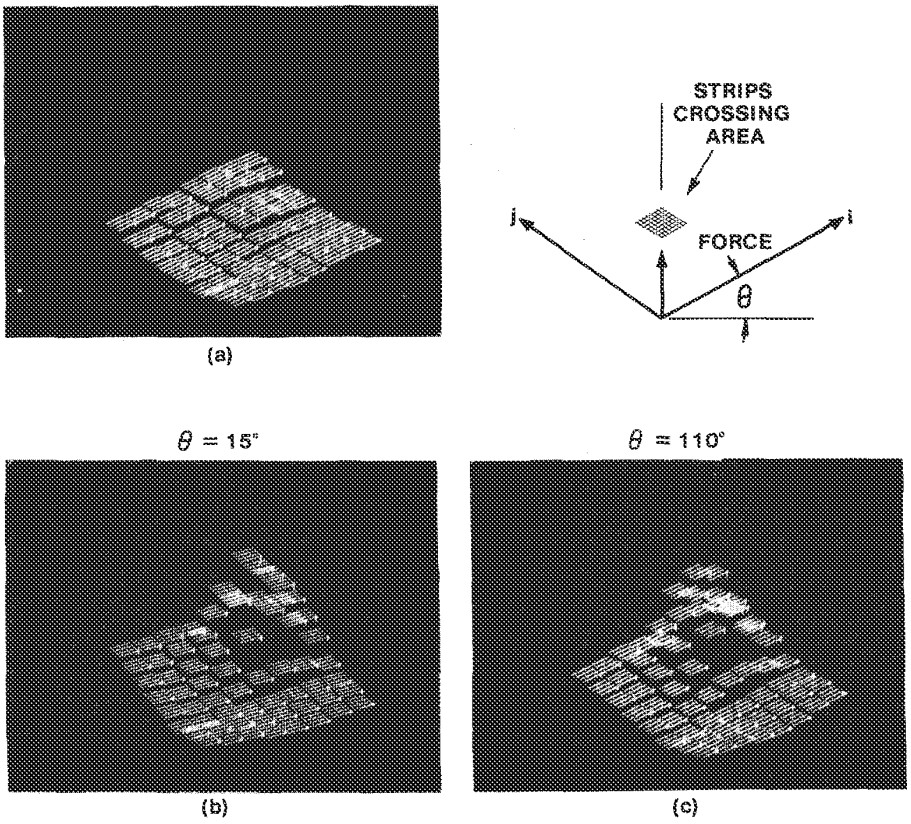


Fig. 6 Photographs of touch sensor raw data display showing zero force offset image and two views of touching a 1/4" diameter ball.

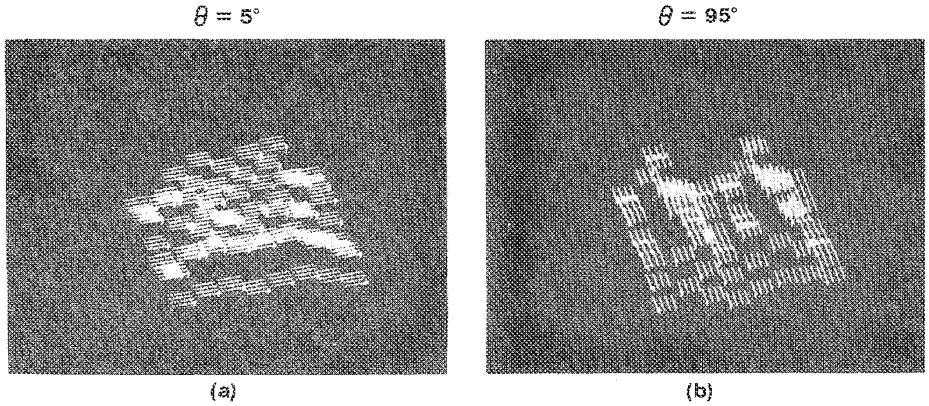


Fig. 7 Two views of touch data for lead ends of an eight pin dual in line package.

FEATURES
Rate This Article: What's Wrong with the Culture of Critique

START
Cheat with Science: Why Smart B-Ballers Bank on the Bank Shot

PLAY
Harry Potter, RIP

MAGAZINE
19.08

Clive Thompson on The Breakthrough Myth

By Clive Thompson ✉ July 26, 2011 | 12:00 pm | Wired August 2011

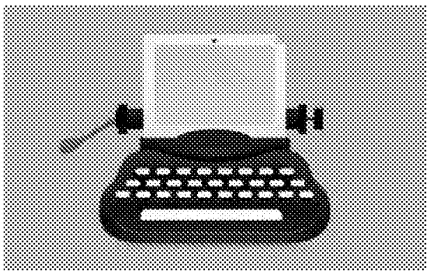


Illustration: Dev Gupta

Tech people love stories about breakthrough innovations---gadgets or technologies that emerge suddenly and take over, like the iPhone or Twitter. Indeed, there's a whole industry of pundits, investors, and websites trying feverishly to predict the Next New Big Thing. The assumption is that breakthroughs are inherently surprising, so it takes special genius to spot one coming.

But that's not how innovation really works, if you ask Bill Buxton. A pioneer in computer graphics who is now a principal researcher at Microsoft, he thinks paradigm-busting inventions are easy to see coming because they're already lying there, close at hand. "Anything that's going to have an impact over the next decade---that's going to be a billion-dollar industry---has always already been around for 10 years," he says.

Buxton calls this the "long nose" theory of innovation: Big ideas poke their noses into the world very slowly, easing gradually into view.

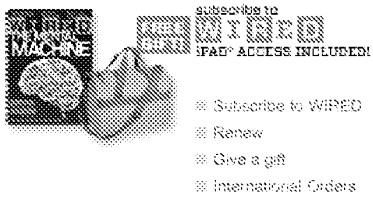
Can this actually be true? Buxton points to exhibit A, the pinch-and-zoom gesture that Apple introduced on the iPhone. It seemed like a bolt out of the blue, but as Buxton notes, computer designer Myron Krueger pioneered the pinch gesture on his experimental Video Place system in 1983. Other engineers began experimenting with it, and companies like Wacom introduced tablets that let designers use a pen and a puck simultaneously to manipulate images onscreen. By the time the iPhone rolled around, "pinch" was a robust, well-understood concept.

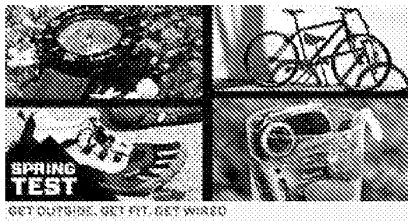
A more recent example is the Microsoft Kinect. Sure, the idea of controlling software just by waving your body seems wild and new. But as Buxton says, engineers have long been perfecting motion-sensing for alarm systems and for automatic doors in grocery stores. We've been controlling software with our bodies for years, just in a different domain.

This is why truly billion-dollar breakthrough ideas have what Buxton calls surprising obviousness. They feel at once fresh and familiar. It's this combination that lets a new gizmo take off quickly and dominate.

The iPhone was designed by Apple engineers who had learned plenty from successes and failures in the PDA market, including, of course, their own ill-fated Newton. By the time they added those pinch gestures, they'd made the obvious freshly surprising.

If you want to spot the next thing, Buxton argues, you just need to go "prospecting and mining"---looking for concepts that are already successful in one field so you can bring them to another. Buxton particularly recommends prospecting the musical world, because musicians invent gadgets and interfaces that are robust and sturdy yet creatively cool---like guitar pedals. When a team led by Buxton





developed the interface for Maya, a 3-D design tool, he heavily plundered music hardware and software. ("There's normal spec, there's military spec, and there's rock spec," he jokes.)

OK: If it's so easy to spy the future, what are Buxton's predictions? He thinks tablet computers, pen-based interfaces, and omnipresent e-ink are going to dominate the next decade. Those inventions have been slowly stress-tested for 20 years now, and they're finally ready.

Using a "long nose" analysis, I have a prediction of my own. I bet electric vehicles are going to become huge----specifically, electric bicycles. Battery technology has been improving for decades, and the planet is urbanizing rapidly. The nose is already poking out: Electric bikes are incredibly popular in China and becoming common in the US among takeout/delivery people, who haul them inside their shops each night to plug them in. (Pennies per charge, and no complicated rewiring of the grid necessary.) I predict a design firm will introduce the iPhone of electric bikes and whoa: It'll seem revolutionary!

But it won't be. Evolution trumps revolution, and things happen slowly. The nose knows.

Email clive@clivethompson.net.

Post Comment | Permalink

Like Confirm Tweet 63

PREVIOUS
What Bandwidth Caps Would Mean for Internet Gluttons

NEXT
Rate This Article: What's Wrong with the Culture of Critique

6 people liked this.

Add New Comment

Login

Real-time updating is **enabled**.

Showing 2 comments

Sort by popular now

RECENT ARTICLES

Storytown: Mark Norton Talks Vampires and Writing at Comic Con

Harry Potter, RIP

Five Gory Game Deaths

Surfer Geeks Build a Better Wave Pool

Soul-Crushing Realism is a Videogame Hit

Decode: Puzzles, games and harrowing mental torments

Wired Magazine RSS feed

RECENT ISSUES

19.08 - August 2011: Extreme Science
19.07 - July 2011: The Mental Machine
19.06 - June 2011: The Smartest Jobs
19.05 - May 2011: The Humor Issue
19.04 - April 2011: How To Make Stuff

ADVERTISEMENT

Overstock iPads: $30.93
Get 32GB Apple iPads for $30.93. Limit One Per Customer. Grab Yours. - www.DealFun.com/iPads

Electric Tricycles Sale
Perfect for riders not able to balance. In stock. Free Shipping. - www.cabbikes.com

Intuos4 Graphics Tablet
Perfect For Creative Professionals. New Features & Specs - Try It Now! - www.Wacom.com/Intuos4

PC Tablets
Great Selection with Free Shipping! Order Now from J&R and Save - www.JR.com/Tablets

Ads by Google

SERVICES

SharperBike

I believe the IPhone of electric bikes is already on the market and it is called the VeloMini folding electric bike. It is everything the Segway should have been as a transportation vehicle (12 miles and hour for 10 miles without pedaling) and you can purchase 7 for the price of a Segway, fold them up and put them all in a small SUV Two will fit in the trunk of a Prius. They come in IPod colors and are used by students, commuters, seniors, as well as boat, RV and private plane owners.

1 week ago Like Reply

ArizonaRider

We feel it happening. Pedego Electric Bike sales are soaring!

1 week ago Like Reply

Subscription: Subscribe | Give a Gift | Renew | International | Questions | Change Address

Quick Links: Contact Us | Sign In/Register | Newsletter | RSS Feeds | Tech Jobs | Wired Mobile | FAQ | Site Map

Search Magazine

Subscribe by email RSS

Text Size: A A A

Condé Nast Web Sites:

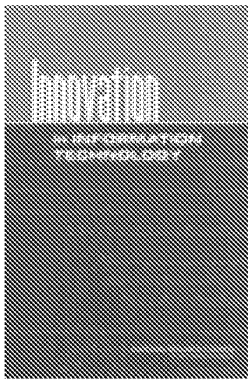
Wwwmonkey | Reddit | Arstechnica | Details | Golf Digest | GQ | New Yorker

Subscribe to a magazine: Condé Nast web sites:

Registration on or use of this site constitutes acceptance of our User Agreement (Revised 4/1/2009) and Privacy Policy (Revised 4/1/2009).

Wired.com © 2011 Condé Nast Digital. All rights reserved.

The material on this site may not be reproduced, distributed, transmitted, cached or otherwise used, except with the prior written permission of Condé Nast Digital.



Innovation in Information Technology

National Research Council
ISBN: 0-309-52622-1, 84 pages, 6x9, (2003)
This free PDF was downloaded from:
http://www.nap.edu/catalog/10795.html

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Purchase printed books and PDF files
- Explore our innovative research tools – try the Research Dashboard now
- Sign up to be notified when new books are published

Thank you for downloading this free PDF. If you have comments, questions or want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to comments@nap.edu.

This book plus thousands more are available at www.nap.edu.

Copyright © National Academy of Sciences. All rights reserved.
Unless otherwise indicated, all materials in this PDF file are copyrighted by the National Academy of Sciences. Distribution or copying is strictly prohibited without permission of the National Academies Press <http://www.nap.edu/permissions/>. Permission is granted for this material to be posted on a secure password-protected Web site. The content may not be posted on a public Web site.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Innovation in INFORMATION TECHNOLOGY

Computer Science and Telecommunications Board

Division on Engineering and Physical Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The projects that are the basis of this synthesis report were approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committees responsible for the final reports of these projects and of the board that produced this synthesis were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by the core sponsors of the Computer Science and Telecommunications Board (CSTB), which include the Air Force Office of Scientific Research, Cisco Systems, Defense Advanced Research Projects Agency, Department of Energy, Intel Corporation, Microsoft Research, National Aeronautics and Space Administration, National Institute of Standards and Technology, National Library of Medicine, National Science Foundation, and Office of Naval Research. Sponsors enable but do not influence CSTB's work. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the organizations or agencies that provide support for CSTB.

International Standard Book Number 0-309-08980-8 (book)
International Standard Book Number 0-309-52622-1 (PDF)

Copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055, (800) 624-6242 or (202) 334-3313 in the Washington metropolitan area; Internet: http://www.nap.edu.

Copyright 2003 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Copyright © National Academy of Sciences. All rights reserved.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

COMPUTER SCIENCE AND TELECOMMUNICATIONS BOARD

DAVID D. CLARK, Massachusetts Institute of Technology, *Chair*
ERIC BENHAMOU, 3Com Corporation
DAVID BORTH, Motorola Labs
JAMES CHIDDIX,** AOL Time Warner
JOHN M. CIOFFI, Stanford University
ELAINE COHEN, University of Utah
W. BRUCE CROFT, University of Massachusetts at Amherst
THOMAS E. DARCIE, University of Victoria
JOSEPH FARRELL, University of California at Berkeley
JOAN FEIGENBAUM, Yale University
HECTOR GARCIA-MOLINA, Stanford University
SUSAN L. GRAHAM,* University of California at Berkeley
JUDITH HEMPEL,* University of California at San Francisco
JEFFREY M. JAFFE,** Bell Laboratories, Lucent Technologies
ANNA KARLIN,** University of Washington
WENDY KELLOGG, IBM Thomas J. Watson Research Center
BUTLER W. LAMPSON, Microsoft Corporation
EDWARD D. LAZOWSKA,** University of Washington
DAVID LIDDLE, U.S. Venture Partners
TOM M. MITCHELL, Carnegie Mellon University
DONALD NORMAN,** Nielsen Norman Group
DAVID A. PATTERSON, University of California at Berkeley
HENRY (HANK) PERRITT, Chicago-Kent College of Law
DANIEL PIKE, GCI Cable and Entertainment
ERIC SCHMIDT, Google Inc.
FRED SCHNEIDER, Cornell University
BURTON SMITH, Cray Inc.
TERRY SMITH,** University of California at Santa Barbara
LEE SPROULL, New York University
WILLIAM STEAD, Vanderbilt University
JEANNETTE M. WING, Carnegie Mellon University

MARJORY S. BLUMENTHAL, Director
KRISTEN BATCH, Research Associate
JENNIFER BISHOP, Senior Project Assistant
JANET BRISCOE, Administrative Officer
DAVID DRAKE, Senior Project Assistant

---

*Term ended June 30, 2001.
**Term ended June 30, 2002.

Copyright © National Academy of Sciences. All rights reserved.

JON EISENBERG, Senior Program Officer
RENEE HAWKINS, Financial Associate
PHIL HILLIARD, Research Associate
MARGARET MARSH HUYNH, Senior Project Assistant
ALAN S. INOUYE, Senior Program Officer
HERBERT S. LIN, Senior Scientist
LYNETTE I. MILLETT, Program Officer
DAVID PADGHAM, Research Associate
CYNTHIA A. PATTERSON, Program Officer
JANICE SABUDA, Senior Project Assistant
BRANDYE WILLIAMS, Staff Assistant
STEVEN WOO, Dissemination Officer

NOTE: For more information on CSTB, see its Web site at <http://www.cstb.org>, write to CSTB, National Research Council, 500 Fifth Street, N.W., Washington, DC 20001, call at (202) 334-2605, or e-mail the CSTB at cstb@nas.edu.

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Preface

The health of the computer science field and related disciplines has been an enduring concern of the National Research Council's Computer Science and Telecommunications Board (CSTB). From its first reports in the late 1980s, CSTB has examined the nature, conduct, scope, and directions of the research that drives innovation in information technology.

Ironically, the success of the industries that produce information technology (IT) has caused confusion about the roles of government and academia in IT research. And it does not help that research in computer science—especially research relating to software—is hard for many people outside the field to understand. This synthesis report draws on several CSTB reports, published over the course of the past decade, to explain the what and why of IT research. It was developed by members of the board, drawing on CSTB's body of work and on insights and experience from their own careers.

This synthesis is kept brief in order to highlight key points. It is paired with a set of excerpts from previous reports, chosen either for their explanation of relevant history or for their compelling development of core arguments and principles.

David D. Clark, *Chair*
Computer Science and
Telecommunications Board

Copyright © National Academy of Sciences. All rights reserved.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Acknowledgment of Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

Frederick P. Brooks, Jr., University of North Carolina at Chapel Hill,
Linda Cohen, University of California at Irvine,
Samuel H. Fuller, Analog Devices Inc.,
Juris Hartmanis, Cornell University,
Timothy Lenoir, Stanford University,
David G. Messerschmitt, University of California at Berkeley,
Ivan E. Sutherland, Sun Microsystems Laboratories, and
Joseph F. Traub, Columbia University.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by John

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Hopcroft, Cornell University. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring board and the institution.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Contents

SUMMARY AND RECOMMENDATIONS 1

1 INNOVATION IN INFORMATION TECHNOLOGY 5
Universities, Industry, and Government: A Complex Partnership Yielding Innovation and Leadership, 5
The Essential Role of the Federal Government, 9
The Distinctive Character of Federally Supported Research, 15
University Research and Industrial R&D, 20
Hallmarks of Federally Sponsored IT Research, 22
Looking Forward, 26

2 EXCERPTS FROM EARLIER CSTB REPORTS 30
Making IT Better: Expanding Information Technology Research to Meet Society's Needs (2000), 31
The Many Faces of Information Technology Research, 31
Implications for the Research Enterprise, 33
Funding a Revolution: Government Support for Computing Research (1999), 37
Lessons from History, 37
Sources of U.S. Success, 44
Research and Technological Innovation, 46
The Benefits of Public Support of Research, 47
Maintaining University Research Capabilities, 48

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Creating Human Resources, 49
The Organization of Federal Support: A Historical Review, 50
1945-1960: Era of Government Computers, 51
The Government's Early Role, 52
Establishment of Organizations, 53
Observations, 57
1960-1970: Supporting Continuing Revolution, 58
Maturing of a Commercial Industry, 58
The Changing Federal Role, 60
1970-1990: Retrenching and International Competition, 67
Accomplishing Federal Missions, 67
Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure (1995), 68
Continued Federal Investment Is Necessary to Sustain Our Lead, 68

WHAT IS CSTB? 71

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Summary and Recommendations

Progress in information technology (IT) has been remarkable, but the best truly is yet to come: the power of IT as a *human enabler* is just beginning to be realized. Whether the nation builds on this momentum or plateaus prematurely depends on today's decisions about fundamental research in computer science (CS) and the related fields behind IT.

The Computer Science and Telecommunications Board (CSTB) has often been asked to examine how innovation occurs in IT, what the most promising research directions are, and what impacts such innovation might have on society. Consistent themes emerge from CSTB studies, notwithstanding changes in information technology itself, in the IT-producing sector, and in the U.S. university system, a key player in IT research.

In this synthesis report, based largely on the eight CSTB reports enumerated below, CSTB highlights these themes and updates some of the data that support them. Much of the material is drawn from (1) the 1999 CSTB report *Funding a Revolution: Government Support for Computing Research*,[1] written by both professional historians and computer scientists to ensure its objectivity, and (2) *Making IT Better: Expanding Information Tech-*

[1]Computer Science and Telecommunications Board, National Research Council. 1999. *Funding a Revolution: Government Support for Computing Research.* National Academy Press, Washington, D.C.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

nology Research to Meet Society's Needs,[2] the 2000 CSTB report that focuses on long-term goals for maintaining the vitality of IT research. Many of the themes achieved prominence in (3) the 1995 CSTB report *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure*,[3] known informally as the Brooks-Sutherland report. Other reports contributing to this synthesis include (4) *Computing the Future: A Broader Agenda for Computer Science and Engineering* (1992),[4] (5) *Building a Workforce for the Information Economy* (2001),[5] (6) *Academic Careers in Experimental Computer Science and Engineering* (1994),[6] (7) *Embedded, Everywhere: A Research Agenda for Networked Systems of Embedded Computers* (2001),[7] and (8) *More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure* (1997).[8] In the text that follows, these reports are cited by number as listed, for easy reference, in Box 1.

Here are the most important themes from CSTB's studies of innovation in IT:

- *The results of research*
  - America's international leadership in IT—leadership that is vital to the nation—springs from a deep tradition of research (1,3,4).
  - The unanticipated results of research are often as important as the anticipated results—for example, electronic mail and instant messaging were by-products of research in the 1960s that was aimed at making it

[2]Computer Science and Telecommunications Board, National Research Council. 2000. *Making IT Better: Expanding Information Technology Research to Meet Society's Needs*. National Academy Press, Washington, D.C.

[3]Computer Science and Telecommunications Board, National Research Council. 1995. *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure*. National Academy Press, Washington, D.C.

[4]Computer Science and Telecommunications Board, National Research Council. 1992. *Computing the Future: A Broader Agenda for Computer Science and Engineering*. National Academy Press, Washington, D.C.

[5]Computer Science and Telecommunications Board, National Research Council. 2001. *Building a Workforce for the Information Economy*. National Academy Press, Washington, D.C.

[6]Computer Science and Telecommunications Board., National Research Council. 1994. *Academic Careers in Experimental Computer Science and Engineering*. National Academy Press, Washington, D.C.

[7]Computer Science and Telecommunications Board, National Research Council. 2001. *Embedded, Everywhere: A Research Agenda for Networked Systems of Embedded Computers*. National Academy Press, Washington, D.C.

[8]Computer Science and Telecommunications Board, National Research Council. 1997. *More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure*. National Academy Press, Washington, D.C.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 1
Reference Numbers for Key CSTB Titles Cited in This Report

Reference Number	*Title*
(1)	*Funding a Revolution: Government Support for Computing Research* (1999)
(2)	*Making IT Better: Expanding Information Technology Research to Meet Society's Needs* (2000)
(3)	*Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure* (1995)
(4)	*Computing the Future: A Broader Agenda for Computer Science and Engineering* (1992)
(5)	*Building a Workforce for the Information Economy* (2001)
(6)	*Academic Careers in Experimental Computer Science and Engineering* (1994)
(7)	*Embedded, Everywhere: A Research Agenda for Networked Systems of Embedded Computers* (2001)
(8)	*More Than Screen Deep: Toward Every-Citizen Interfaces to the Nation's Information Infrastructure* (1997)

NOTE: Complete citations for these reports appear in footnotes 1 through 8 in this "Summary and Recommendations" section.

possible to share expensive computing resources among multiple simultaneous interactive users (1,3).

- The interaction of research ideas multiplies their impact—for example, concurrent research programs targeted at integrated circuit design, computer graphics, networking, and workstation-based computing strongly reinforced and amplified one another (1-4).

- *Research as a partnership*

- The success of the IT research enterprise reflects a complex partnership among government, industry, and universities (1-8).

- The federal government has had and will continue to have an essential role in sponsoring fundamental research in IT—largely university-based—because it does what industry does not and cannot do (1-8). Industrial and governmental investments in research reflect different

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

motivations, resulting in differences in style, focus, and time horizon (1-3,7,8).

- Companies have little incentive to invest significantly in activities whose benefits will spread quickly to their rivals (1,3,7). Fundamental research often falls into this category. By contrast, the vast majority of corporate research and development (R&D) addresses product and process development (1,2,4).
- Government funding for research has leveraged the effective decision making of visionary program managers and program office directors from the research community, empowering them to take risks in designing programs and selecting grantees (1,3). Government sponsorship of research especially in universities also helps to develop the IT talent used by industry, universities, and other parts of the economy (1-5).

• *The economic payoff of research*

- Past returns on federal investments in IT research have been extraordinary for both U.S. society and the U.S. economy (1,3). The transformative effects of IT grow as innovations build on one another and as user know-how compounds. Priming that pump for tomorrow is today's challenge.
- When companies create products using the ideas and workforce that result from federally sponsored research, they repay the nation in jobs, tax revenues, productivity increases, and world leadership (1,3,5).

The themes highlighted above underlie two recurring and overarching recommendations evident in the eight CSTB reports cited:

Recommendation 1 The federal government should continue to boost funding levels for fundamental information technology research, commensurate with the growing scope of research challenges (2-4,6-8). It should ensure that the major funding agencies, especially the National Science Foundation and the Defense Advanced Research Projects Agency, have strong and sustained programs for computing and communications research that are broad in scope and independent of any special initiatives that might divert resources from broadly based basic research (2,3).

Recommendation 2 The government should continue to maintain the special qualities of federal IT research support, ensuring that it complements industrial research and development in emphasis, duration, and scale (1-4,6).

This report addresses the ways that past successes can guide federal funding policy to sustain the IT revolution and its contributions to other fields.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

1

Innovation in Information Technology

UNIVERSITIES, INDUSTRY, AND GOVERNMENT: A COMPLEX PARTNERSHIP YIELDING INNOVATION AND LEADERSHIP

Figure 1 illustrates some of the many cases in which fundamental research in IT, conducted in industry and universities, led 10 to 15 years later to the introduction of entirely new product categories that became billion-dollar industries. It also illustrates the complex interplay between industry, universities, and government. The flow of ideas and people—the interaction between university research, industry research, and product development—is amply evident.

Figure 1 updates Figure 4.1 from the 2002 CSTB report *Information Technology Research, Innovation, and E-Government*.[1] The originally published figure[2] produced an extraordinary response: it was used in presentations to Congress and to administration decision makers, and it was

[1]Computer Science and Telecommunications Board, National Research Council. 2002. *Information Technology Research, Innovation, and E-Government*. National Academy Press, Washington, D.C.

[2]Known informally as the "tire-tracks chart" because of its appearance, the figure was first published in *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure* (3; p. 2).

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

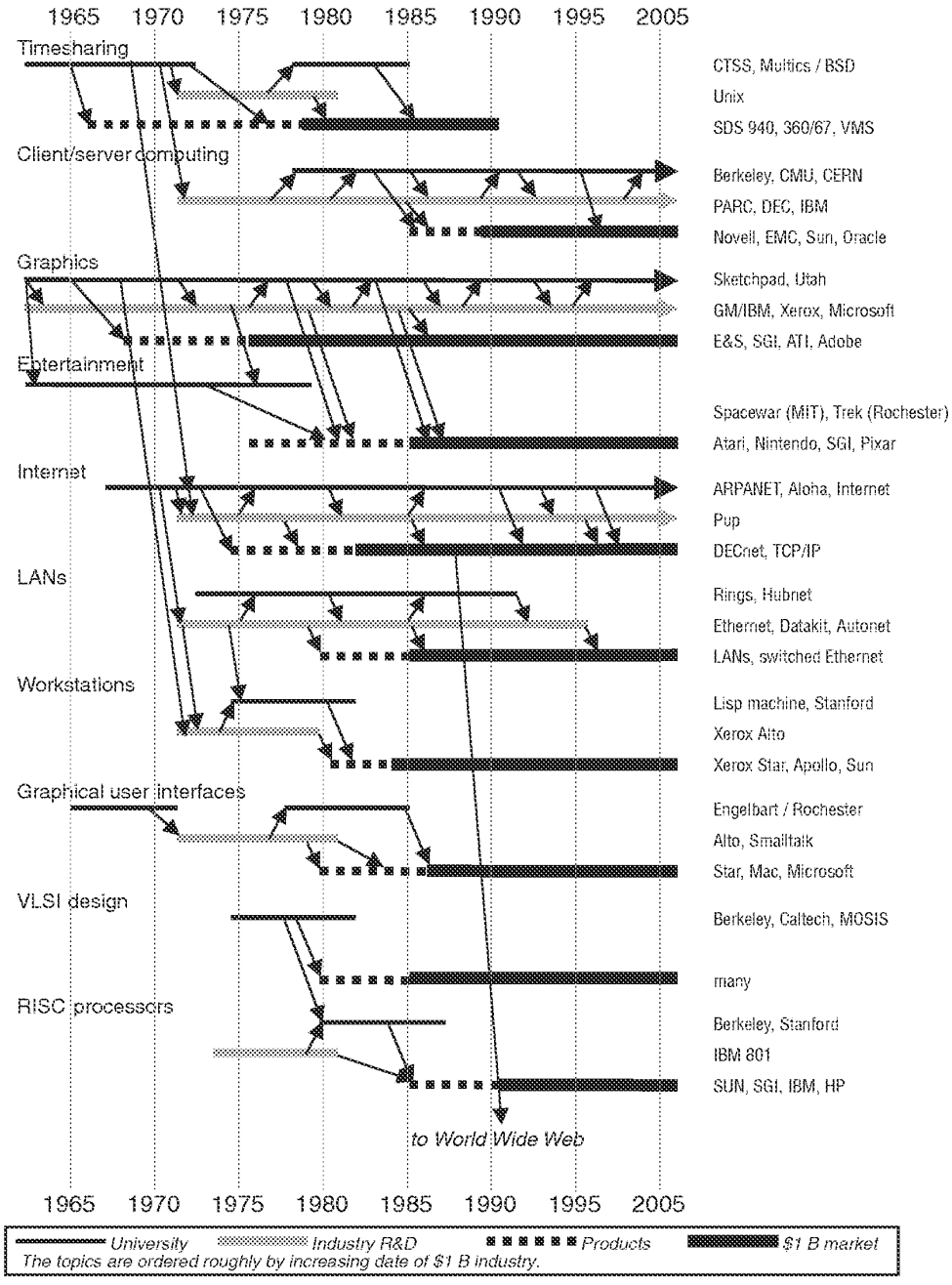
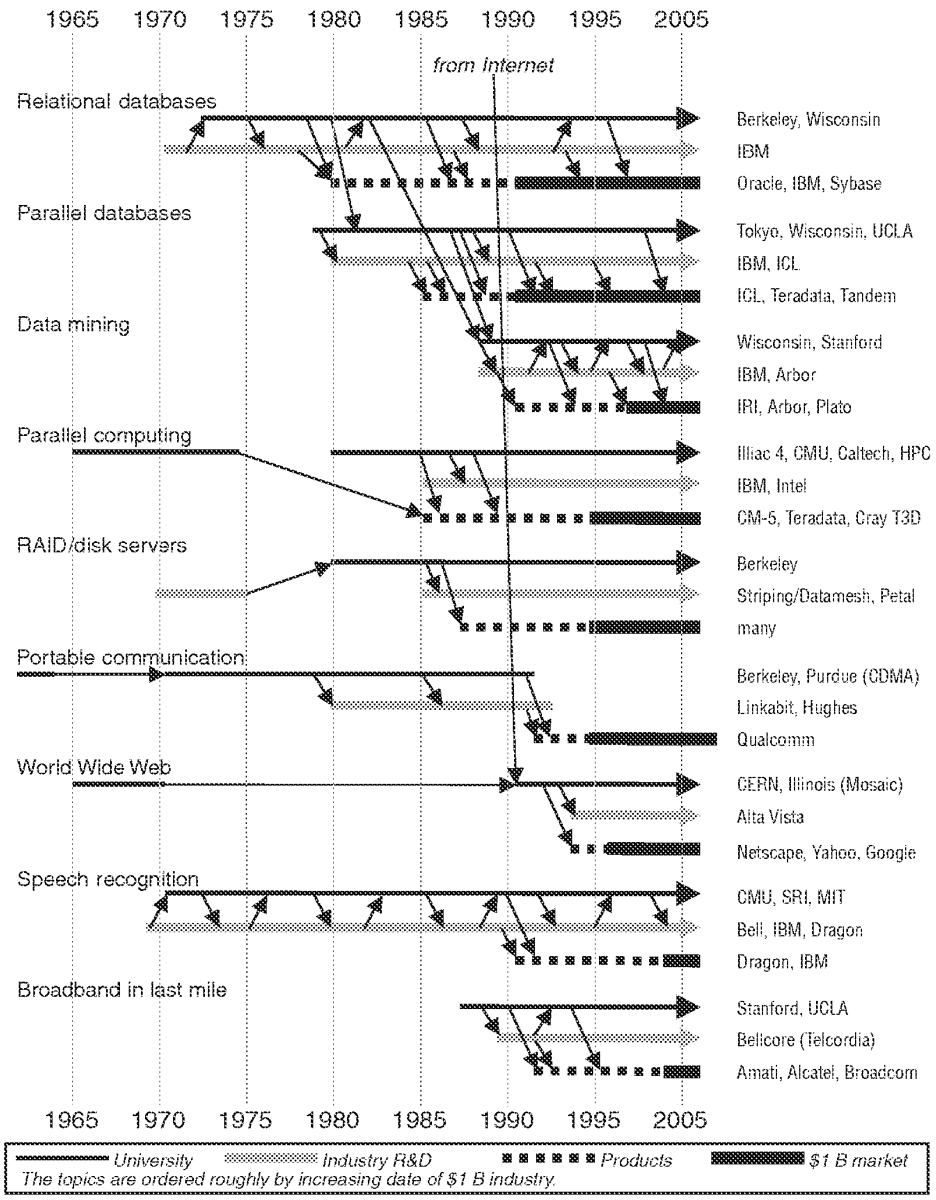


FIGURE 1 Examples of government-sponsored IT research and development in the creation of commercial products and industries. Federally sponsored research lies at the heart of many of today's multibillion-dollar information technology industries—industries that are transforming our lives and driving our economy. Ideas and people flow in complex patterns. The interaction of research ideas

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html



multiplies their effect. The result is that the United States is the world leader in this critical arena. Although the figure reflects input from many individuals at multiple points in time, ensuring readability required making judgments about the examples to present, which should be seen as illustrative rather than exhaustive. SOURCE: 2002 update by the Computer Science and Telecommunications Board of a figure (Figure ES.1) originally published in Computer Science and Telecommunications Board, National Research Council, 1995, *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure,* National Academy Press, Washington, D.C.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

discussed broadly in the research community. Although IT commercial success leads some policy makers to assume that industry is self-sufficient, the tire-tracks chart underscores how much industry builds on government-funded university research, sometimes through long incubation periods (1,3).

Figure 1 also illustrates—although sketchily—the interdependencies of research advances in various subfields. There is a complex research ecology at work, in which concurrent advances in multiple subfields—in particular within computer science but extending into other fields, too, from electrical engineering to psychology—are mutually reinforcing: they stimulate and enable one another.[3]

One of the most important messages of Figure 1 is the long, unpredictable incubation period—requiring steady work and funding—between initial exploration and commercial deployment (1,3). Starting a project that requires considerable time often seems risky, but the payoff from successes justifies backing researchers who have vision. It is often not clear which aspect of an early-stage research project will be the most important; fundamental research produces a range of ideas, and later developers select from among them as needs emerge. Sometimes the utility of ideas is evident well after they have been generated. For example, some early work in artificial intelligence has achieved unanticipated applicability in computer games, some of which are now being investigated for decision support and other professional uses as well as recreation.

It is important to remember that real-world requirements can change quickly. Although the end of the Cold War was interpreted by some as lessening the need for research,[4] September 11, 2001, underscored research needs in several areas: system security and robustness, automatic natural language translation, data integration, image processing, and biosensors, among others—areas in which technical problems are difficult to begin with, and may become harder when technology must be designed to both meet homeland security needs and protect civil liber-

[3]The idea that research in IT not only builds in part on research in physics, mathematics, electrical engineering, psychology, and other fields but also strongly influences them is consistent with what Donald Stokes has characterized in his four-part taxonomy as "Pasteur's Quadrant" research: use- or application-inspired basic research that pursues fundamental understanding (such as Louis Pasteur's research on the biological bases of fermentation and disease). See the discussion on pp. 26-29 in the 2000 CSTB report *Making IT Better* (2), and see Donald E. Stokes, 1997, *Pasteur's Quadrant: Basic Science and Technological Innovation*, Brookings Institution Press, Washington, D.C.

[4]Linda R. Cohen and Roger G. Noll. 1994. "Privatizing Public Research," *Scientific American* 271(3): 72-77.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

ties.[5] Without fundamental research, the cupboard is bare when there is a sudden need for ideas to reduce to practice.

THE ESSENTIAL ROLE OF THE FEDERAL GOVERNMENT

Federally sponsored research played a critical role in creating the enabling technologies for each of the billion-dollar market segments illustrated in Figure 1—and for many others as well. The government role coevolved with IT industries: its organization and emphases changed to focus on capabilities not ready for commercialization and on new needs that emerged as commercial capabilities grew, both moving targets (1). As this coevolution shows, successful technology development relies on flexibility in the conduct of research and in the structure of industry.

Most often, this federal investment took the form of grants or contracts awarded to university researchers by the Defense Advanced Research Projects Agency (DARPA) and/or the National Science Foundation (NSF)—although a shifting mix of other funding agencies has been involved, reflecting changes in the missions of these agencies and their needs for IT (1,3). For example, the Department of Energy (DOE), the National Aeronautics and Space Administration (NASA), and the military services have supported high-performance computing, networking, human-computer interaction, and other kinds of research.[6]

Why has federal support been so effective in stimulating innovation in computing? As discussed below, many factors have been important.

1. *Federally funded programs have supported long-term research into fundamental aspects of computing, whose widespread practical benefits typically take years to realize* (1).

"Long-term" research refers to a long time horizon for the research effort and for its impact to be realized. Examples of innovations that required long-term research include speech recognition, packet radio, computer graphics, and internetworking. In every case illustrated in Figure 1, the time from first concept to successful market is measured in

[5]See Computer Science and Telecommunications Board, National Research Council. 2003. *Information Technology for Counterterrorism: Immediate Actions and Future Possibilities*. National Academies Press, Washington, D.C.

[6]In addition to research funding, complementary activities have been undertaken by other agencies, such as the National Institute of Standards and Technology, which often brings together people from universities and industry on issues relating to standards setting and measurement.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

decades (see Box 2)—a contrast to the more incremental innovations that are publicized as evidence of the rapid pace of IT innovation.

Work on speech recognition, for example, which began in earnest in the early 1970s, took until 1997 to generate a successful product for enabling personal computers to recognize continuous speech (8). Work on packet radio also dates from the 1970s, and its realization in commercial ad hoc mobile networking also began in the late 1990s.[7] Fundamental algorithms for shading three-dimensional graphics images, which were developed with federal funding in the 1960s, saw limited use on high-performance machines until they entered consumer products in the 1990s; today these algorithms are used in a range of products in the health care, entertainment, and defense industries. The research programs behind these innovations not only were long-term but also were broad enough to accommodate within a single program the development of those unanticipated results that have in many cases provided the most significant outcomes of a project.

The benefits of a long time horizon, combined with program breadth, extend to today's challenges. This point was emphasized in CSTB's 1997 report on usability, *More Than Screen Deep* (8), which explained (at p. 192):

> Federal initiatives that emphasize long-term goals beyond the horizon of most commercial efforts and that may thus entail added risk have the potential to move the whole information technology enterprise into new modes of thinking and to stimulate discovery of new technologies for the coming century.

Because of unanticipated results and synergies, the exact course of fundamental research cannot be planned in advance, and its progress cannot be measured precisely in the short term. Even projects that appear to have failed or whose results do not seem to have immediate utility often make significant contributions to later technology development or achieve other objectives not originally envisioned. A striking example is the field of number theory (1): for hundreds of years a branch of pure mathematics without applications, it is now the basis for the public-key cryptography that underlies the security of electronic commerce.

[7]Similarly, commercial developments in broadband cellular radio (which has become essentially wireless Internet access in third-generation wireless) are built in part on many decades of federally supported research into Code Division Multiple Access technology, signal processing for antenna arrays, error-correction coding, and so on.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 2
The Role of Federal Support for Fundamental Research in IT

CSTB's 1995 report *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure* (3) examined the payoff from several decades of federal investment in IT research. Among the conclusions of that report are these:

- *Research has kept paying off over a long period.*
- *The payoff from research takes time.* As Figure [1] shows, at least 10 years, more often 15, elapse between initial research on a major new idea and commercial success. This is still true in spite of today's shorter product cycles.
- *Unexpected results are often the most important.* Electronic mail and the "windows" interface are only two examples. . . .
- *Research stimulates communication and interaction.* Ideas flow back and forth between research programs and development efforts and between academia and industry [and between research programs with different foci that are proceeding concurrently].
- *Research trains people,* who start companies or form a pool of trained personnel that existing companies can draw on to enter new markets quickly.
- *Doing research involves taking risks.* Not all public research programs have succeeded or led to clear outcomes even after many years. But the record of accomplishments suggests that government investment in computing and communications research has been highly productive.[1]

---

[1]Computer Science and Telecommunications Board, National Research Council. 1995. *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure.* National Academy Press, Washington, D.C., pp. 3-4.

2. *The interplay of government-funded and industry research has been an important factor in IT commercialization* (1-8).

The examples in Figure 1 show the interplay between government-funded research and industry research and development. In some cases, such as reduced-instruction-set computing (RISC) processors, the initial ideas came from industry, but the research that was essential to advancing these ideas came from government funding to universities. RISC was conceived at IBM, but it was not commercialized until DARPA funded additional research at the University of California at Berkeley and at Stanford University as part of its Very Large Scale Integrated Circuit (VLSI) program of the late 1970s and early 1980s (1,3). The VLSI program also supported university research that gave rise to such companies as Synopsys, Cadence, and Mentor, which have acquired dozens of smaller

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

companies that started as spinoffs of DARPA-funded[8] university research; such research has also pushed the proverbial envelope in algorithms and user interfaces. The more than $3 billion electronic design automation industry is an essential enabler to other parts of IT.

Similarly, IBM pioneered the concept of relational databases (its System R project) but did not commercialize the technology. NSF-sponsored research at the University of California at Berkeley brought this technology to the point at which it was commercialized by several start-up companies and then by more established database companies (including IBM) (1,3). In other cases, such as timesharing, the initial ideas came from the university community, and subsequent industry research, while significant for a time, was not sustained. In none of the examples in Figure 1 did industry alone provide the necessary research.

3. *There is a complex interleaving of fundamental research and focused development* (1-3).

In the case of integrated circuit (VLSI) design tools, research innovation led to products and then to major industrial markets. A still-unfolding example is the theoretical research that yielded the algorithms behind the Web-content management technology underlying Akamai. In the case of relational databases, the introduction of products stimulated new fundamental research questions, leading to a new generation of products with capabilities vastly greater than those of their predecessors. The purpose of publicly funded research is to advance knowledge and to solve hard problems. The exploitation of that knowledge and those solutions in products is fundamentally important, but the form it takes is often unpredictable, as is the impact on future research (see Box 3).

4. *Federal support for research has tended to complement, rather than preempt, industry investments in research.*

The IT sector invests an enormous amount each year in R&D. It is critical to understand, however, that the vast majority of corporate R&D has always been focused on product and process development (2). This is what shareholders (or other investors) demand. It is harder for corporations to justify funding long-term, fundamental research. Economists

[8]In some cases, the Semiconductor Research Corporation provided the funding. For additional information, see the Web site <http://www.src.org/member/about/history.asp>. Accessed June 2, 2003.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 3
The Technological Underpinnings of Electronic Commerce

Electronic commerce is becoming pervasive. It is changing many aspects of our lives, from the way we shop to the way we obtain government services.

The organizations and individuals that exploit electronic commerce employ commercial tools from companies such as Microsoft and Oracle. They may not think of themselves as the beneficiaries of federal investments in university-based IT research—but they are. Nearly every key technological component underlying electronic commerce has been shaped by this investment. For example:

- *The Internet*—Defense Advanced Research Projects Agency (DARPA) investments in the 1960s and 1970s were followed by National Science Foundation (NSF) investments in the 1980s and early 1990s, with research (supported by multiple agencies) continuing to this day (1).
- *Web browsers*—Mosaic, the first browser with a graphical user interface, was invented at the NSF-supported National Center for Supercomputer Applications at the University of Illinois (1).
- *Public-key cryptography for secure credit card transactions*—NSF sponsored university-based research in the 1970s that supported this innovation (1).
- *Back-end database and transaction processing systems*—NSF and DARPA supported key research on relational databases and transaction processing systems at the University of California at Berkeley, University of Wisconsin, and elsewhere, beginning in the early 1980s and continuing to this day (1).
- *Search engines*—Search engines grew out of federally supported university research programs, such as the ranking algorithm work at Stanford University that contributed to Google; the WebCrawler and MetaCrawler grew out of work at the University of Washington.

But the development is not complete: a range of technical challenges still exist, along with challenges for improving the fit between the technologies and the behavior and needs of the people who use them (2,8).

SOURCES: Pieces of this history are recounted in the previously cited CSTB reports (1-8) and in CSTB's series of reports on the Internet: *Toward a National Research Network* (1988), *Realizing the Information Future: The Internet and Beyond* (1990), *The Unpredictable Certainty: Information Infrastructure Through 2000* (1996), *The Internet's Coming of Age* (2001), and *Broadband: Bringing Home the Bits* (2002), all published by the National Academy Press, Washington, D.C.

have articulated the concept of "appropriability" to express the extent to which the results of an investment can be captured by the investor, as opposed to being available to all players in the market. The results of long-term, fundamental research are hard to appropriate for several reasons: they tend to be published openly and thus to become generally known; they tend to have broad value; the most important may be unpre-

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

dictable in advance; and they become known well ahead of the moment of realization as a product, so that many parties have the opportunity to incorporate the results into their thinking. In contrast, incremental research and product development can be performed in a way that is more appropriable: it can be done under wraps, and it can be moved into the marketplace more quickly and predictably.

Although individual industrial players may find it hard to justify research that is weakly appropriable, it is the proper role of the federal government to support this sort of endeavor (1,3). When companies create successful new products using the ideas and workforce that result from federally sponsored research, they repay the nation handsomely in jobs, tax revenues, productivity increases, and world leadership (1,3). Long-term research often has great benefits for the IT sector as a whole, although no particular company can be sure of reaping most of these benefits.

Appropriability helps to explain why the companies that have tended to provide the greatest support for fundamental research are large companies that enjoy dominant positions in their market (1). AT&T and IBM, for example, have historically made significant investments in fundamental research. Anything that advances IT as a whole benefits the dominant players—they may be capable of reaping a significant proportion of the returns on their research investments. As IT industries became more competitive, however, even these firms began to link their research more closely with corporate objectives and product development activities.[9] One of them (AT&T) has radically cut back its research effort. This process began with a government proceeding that resulted in the splitting up of functions formerly aggregated under "Ma Bell" and continued with the growth and contraction of a set of industry research and development endeavors (AT&T Research, Lucent Technologies, Agere Systems, and Bellcore [now Telcordia]) where once there was the monolithic Bell Laboratories.[10]

Several of the companies that have recently emerged as dominant in their sectors, such as Intel and Microsoft, have increased their support for fundamental research. However, many other successful companies with large market shares (e.g., Cisco, Dell, Oracle) have chosen not to invest in fundamental research to any significant extent. And even at Microsoft, just as at AT&T and IBM before it, the investment in fundamental research

[9]Elizabeth Corcoran, 1994, "The Changing Role of U.S. Corporate Research Labs," *Research-Technology Management* 37(4):14-20; Peter Coy, 1993, "R&D Scoreboard: In the Labs, the Fight to Spend Less, Get More," *Business Week*, June 28, pp. 102-124.

[10]CSTB launched a study of the future of telecommunications R&D in 2003.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

represents a relatively small proportion of overall corporate R&D. In 2002, Microsoft invested roughly $5 billion in R&D, but the company's fundamental research arm is small enough to suggest that 95 percent of Microsoft's R&D investment is product-related.

Start-ups represent the other end of the spectrum. A hallmark of U.S. entrepreneurship, start-ups and start-up financing promote flexibility in industry structure and industry management. They have facilitated the development of high-risk products as well as an iconoclastic, risk-taking attitude among more traditional companies and managers in the IT business. *But they do not engage in research* (2). Thus, the wave of Internet-related and other IT start-ups of the 1990s is notable for two reasons: first, these start-ups attracted some researchers away from universities and research, and second, notwithstanding the popular labeling of those start-ups as "high-tech," they applied the fruits of past research rather than generating more. Start-ups illustrate the critical role of government funding in building the foundations for innovative commercial investments.

THE DISTINCTIVE CHARACTER OF FEDERALLY SUPPORTED RESEARCH

The most important characteristic of successful government research activities is their breadth of scope—both in their long time dimension and in their focus on activities that are potentially difficult to appropriate privately in their entirety. Two specific topic areas that illustrate these principles are large-scale IT systems and social applications of IT. Growing capabilities and broadening use of IT in the 1990s motivated CSTB recommendations for greatly increased federal support in these two categories (2) (see Boxes 4 and 5).

Prospects for progress in social applications—however difficult—are one reason for confidence that IT will improve as a human enabler. The beginnings evident in all of these areas are but crude indicators of what research may make possible.

An example of particular currency is that of cybersecurity. Stimulated by the events of September 11, CSTB issued the report *Cybersecurity Today and Tomorrow: Pay Now or Pay Later*, in early 2002. The report summarized the findings of seven CSTB reports issued over the preceding decade that had cybersecurity as a principal theme. *Cybersecurity Today and Tomorrow* concludes with the following paragraph:

> Research and development on information systems security should be construed broadly to include R&D on defensive technology (including both underlying technologies and architectural issues), organizational and sociological dimensions of such security, forensic and recovery tools, and best policies and practices. Given the failure of the market to ad-

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 4
Defining Large-Scale Systems and Social Applications of Information Technology

Large-scale systems are IT systems that contain many (thousands, millions, billions, or trillions or more) interacting hardware and software components. They tend to be heterogeneous—in that they are composed of many different types of components—and highly complex because the interactions among the components are numerous, varied, and complicated. They also tend to span multiple organizations (or elements of organizations) and have changing configurations. Over time, the largest IT systems have become ever larger and more complex, and at any given point in time, systems of a certain scale and complexity are not feasible or economical to design with existing methodologies.

Social applications of IT serve groups of people in shared activities. The most straightforward of these applications improve the effectiveness of geographically dispersed groups of people who are collaborating on some task in a shared context. More sophisticated applications may support the operations of a business or the functioning of an entire economy; systems for e-commerce are an example. Characteristic of social applications of IT is the embedding of IT into a large organizational or social system to form a "sociotechnical" system in which people and technology interact to achieve a common purpose—even if that purpose is not obviously social, such as efficient operation of a manufacturing line (which is a conjunction of technological automation and human workers) or rapid and decisive battlefield management (which is a conjunction of command-and-control technology and the judgment and expertise of commanders). Social applications of IT—especially those supporting organizational and societal missions—tend to be large-scale and complex, mixing technical and nontechnical design and operational elements and involving often-difficult social and policy issues such as those related to privacy and access.

SOURCE: Reprinted from Computer Science and Telecommunications Board, National Research Council. 2000. *Making IT Better: Expanding Information Technology Research to Meet Society's Needs.* National Academy Press, Washington, D.C., p. 3.

dress security challenges adequately, government support for such research is especially important.[11]

CSTB's 2001 study on networked systems of embedded computers (7) sounds a similar theme (at p. 9):

> [T]he committee (composed of people from both academia and industry) believes that while some of the questions raised in this report may

[11]Computer Science and Telecommunications Board, National Research Council. 2002. *Cybersecurity Today and Tomorrow: Pay Now or Pay Later.* National Academy Press, Washington, D.C., pp. 14-15.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 5
Research on the Social Applications of Information Technology

Research on the social applications of information technology (IT) combines work in technical disciplines, such as computing and communications, with research in the social sciences to understand how people, organizations, and IT systems can be combined to most effectively perform a set of tasks. Such research can address a range of issues related to IT systems, as demonstrated by the examples below . . . :

- *Novel activities and shifts in organizational, economic, and social structures*—What will people do (at work, in school, at play, in government, and so on) when computers can see and hear better than people can? How will activities and organizations change when robotic technology is widespread and cheap? How will individual and organizational activities change when surveillance via IT becomes effectively universal? New technologies will affect all kinds of people in many ways, and they hold particular promise for those with special situations or capabilities, because they will give them broader access to social and economic activities.
- *Electronic communities*—How can IT systems be best designed to facilitate the communication and coordination of groups of people working toward a common goal? Progress requires an understanding of the sociology and dynamics of groups of users, as well as of the tasks they wish to perform. Psychologists and sociologists could offer insight for the conceptualization and refinement of these social applications, and technologists could mold their technological aspects.
- *Electronic commerce*—How can buyers and sellers be best brought together to conduct business transactions on the Internet? What kinds of security technologies will provide adequate assurances of the identities of both parties and protect the confidentiality of their transactions without imposing unnecessary burdens on either? How will electronic commerce affect the competitive advantage of firms, their business strategies, and the structure of industries (e.g., their horizontal and vertical linkages)? Such work requires the insight of economists, organizational theorists, business strategists, and psychologists who understand consumer behavior, as well as of technologists.
- *Critical infrastructures*—How can IT be better embedded into the nation's transportation, energy, financial, telecommunications, and other infrastructures to make them more efficient and effective without making them less reliable or more prone to human error? For example, how can an air traffic control system be designed to provide controllers with sufficient information to make critical decisions without overwhelming them with data? Such work requires the insight of cognitive psychologists and experts in air traffic control, as well as of technologists.
- *Complexity*—How can the benefits of IT be brought to the citizenry without the exploding complexity characteristic of professional uses of IT? Although networks, computers, and software can be assembled and configured by professionals to support the mission-critical computing needs of large organizations, the techniques that make this possible are inadequate for information appliances designed for the home, car, or individual. Research is needed to simplify and automate

continues

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 5
Continued

system configuration, change, and repair. Such research will require insight from technologists, cognitive psychologists, and those skilled in user interface design.

SOURCE: Reprinted from Computer Science and Telecommunications Board, National Research Council. 2000. *Making IT Better: Expanding Information Technology Research to Meet Society's Needs.* National Academy Press, Washington, D.C., p. 7.

> be answered without a concerted, publicly funded research agenda, leaving this work solely to the private sector raises a number of troubling possibilities. Of great concern is that individual commercial incentives will fail to bring about work on problems that have a larger scope and that are subject to externalities: interoperability, safety, upgradability, and so on. Moreover, a lack of government funding will slow down the sharing of the research, since the commercial concerns doing the research tend to keep the research private to retain their competitive advantage. The creation of an open research community within which results and progress are shared is vital to making significant progress in this arena.

Another example of the distinctive role that federal funding can play in computing research comes from two recent CSTB studies of the Internet. The 2001 report *The Internet's Coming of Age* examined the role of the government in funding research that leads to open standards, exemplified by the work that defined the Internet. One of the Internet's hallmarks has been its openness. Proprietary research can enhance a particular product, but research leading to open standards can create a new marketplace for products. Each company that is an Internet "player" will be tempted to diverge from the common standard if it looks possible to capture a large portion of it—we have seen this during the past decade in protocols for transport, electronic mail, instant messaging, and many other areas (see Box 6). However, a common, open standard maximizes overall social welfare as a result of the network externalities obtained from the larger market. When effective open standards are made available, they can be attractive in the marketplace and may win out over proprietary ones. The report notes:

> The government's role in supporting open standards for the Internet has not been, and should not be, to directly set or influence standards. Rather, its role should be to provide funding for the networking research community, which has led to both innovative networking ideas as well

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 6
The Origins of Electronic Mail and Instant Messaging

The invention of timesharing systems in the 1960s not only contributed important technical developments in hardware, software, and system security but also provided the environment that led to the development of the most useful and widespread of popular applications, namely, e-mail and instant messaging (1).

Timesharing allowed concurrent multiple users to share the power of a computer, which provided a fresh way for colleagues to interact. By 1970, programmers in federally funded research laboratories had developed both asynchronous electronic mail and facilities for real-time interaction between users, in research operating systems such as Tenex, Multics, and CalTSS.

These modalities—now widely known as e-mail and instant messaging—proved so powerful that they have spread far and wide with the availability of low-cost personal computers, public networking, and client-server computing. These popular and visible tools, as well as all of the other forms of collaborative computing, have truly transformed our work and our lives. They owe their origins to the funding of IT research by the Defense Advanced Research Projects Agency and the National Science Foundation (1,3).

as specific technologies that can be translated into new open standards.[12]

A 2002 report, *Broadband: Bringing Home the Bits,* outlines an even broader role for federally funded research to enable openness in infrastructural systems:

> Support research and development on access technologies, especially targeting the needs of nonincumbent players and other areas that are not targets of stable, private sector funding. . . . [One target area is] technologies that foster the accommodation of multiple competitive service providers over facilities. Such open access-ready systems might not be a natural research and development target of large incumbent providers but will be the preferred form for a variety of public sector or public-private deployments.[13]

Broadband: Bringing Home the Bits notes that federally funded research can complement the more proprietary-oriented industry approaches to innovation, whether in communications architecture or content. It also

[12]Computer Science and Telecommunications Board, National Research Council. 2001. *The Internet's Coming of Age.* National Academy Press, Washington, D.C., p. 18.

[13]Computer Science and Telecommunications Board, National Research Council. 2002. *Broadband: Bringing Home the Bits.* National Academy Press, Washington, D.C., p. 40.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

calls for the support of research on economic, social, and regulatory factors relating to broadband technologies—nontechnical factors that interact with the design and deployment of broadband.

UNIVERSITY RESEARCH AND INDUSTRIAL R&D

Much of the government-funded research in IT has been carried out at universities.[14] Federal support has constituted roughly 70 percent of total university research funding in computer science and electrical engineering since 1976 (2). Among the many benefits of federally funded university research, the generation of new knowledge is only one (see Box 7).

Strong research institutions are recognized as being among the most critical success factors in high-tech economic development (5). In computing, electronics, telecommunications, and biotechnology, evidence of the correlation abounds—in Boston (Harvard University and the Massachusetts Institute of Technology); Research Triangle Park (Duke University, the University of North Carolina, and North Carolina State University); New Jersey (Princeton University, Rutgers University, and New York City-based Columbia University); Austin (the University of Texas); southern California (the University of California at San Diego, the University of California at Los Angeles, the California Institute of Technology, and the University of Southern California); northern California (the University of California at Berkeley, the University of California at San Francisco, and Stanford University); and Seattle (the University of Washington).

In addition to creating ideas and companies, universities often import forefront technologies to their regions (e.g., the nationwide expansion of ARPANET in the 1970s and of NSFnet in the 1980s, and the continuation of those efforts through the private Internet2 activities in the 1990s and early 2000s). Universities also serve as powerful magnets for companies seeking to relocate. These contributions are not reflected in Figure 1.

Figure 1 also does not capture the most important product of universities: people. The American research university is unique in the degree to which it integrates research with education—both undergraduate and graduate education. Not only do graduating students serve to staff industry (5,6), but they also are *by far* the most effective vehicle for technol-

[14]The concentration of research in universities is particularly true for computer science research; industry played an important role in telecommunications research before the breakup of AT&T and the original Bell Labs.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 7
The Diverse Benefits of University Research

Universities have a number of important characteristics that contribute to their success as engines of innovation. Among them are the following:

- *Universities can focus on long-term research.* Focusing on long-term research is the special role of universities—one that IT companies cannot be expected to fill to any significant extent (1-3). America's IT companies are extraordinarily adept at improving current products, but the track record is at best mixed on the invention and adoption of "disruptive technologies," and corporate research in IT has been becoming more applied (2).
- *Universities provide a neutral ground for collaboration.* Universities encourage movement and collaboration among faculty through leave and sabbatical policies that allow professors to visit industry, government, and other university departments or laboratories. These uniquely valuable components of the R&D structure in the United States are not generally present in industry. Universities also provide sites at which researchers from competing companies can come together to explore technical issues. At the same time that industry people share their wisdom and experience with university researchers, they have the opportunity to learn from one another (2,6).
- *Universities integrate research and education.* Universities provide a forum for educating the skilled IT workers of the future (5). The presence of research activities in an educational setting creates very powerful synergy (2,4). IT is a rapidly changing field. Many of the specific facts and techniques that a student learns become obsolete early in his or her career. The educational foundation for continuous learning—"keeping up with the field"—is a crucial component of IT education (5). Students, even beginning undergraduates, get that education not only in the classroom, but also by serving as apprentices on leading-edge research projects, where knowledge is being discovered, not read from a book. Often, new ideas are a by-product of what goes on in the classroom: in an attempt to explain the solutions to emerging problems, teachers often deepen their own understanding, while discovering interesting research questions whose answers are as yet unknown. Additionally, students are the most powerful vehicle for technology transfer, not only from university to industry but also between university laboratories and departments, through the hiring of postdoctoral researchers and assistant professors (5).
- *Universities are inherently multidisciplinary.* University researchers are well situated to draw on experts from a variety of other fields (2). There are often cultural barriers to cross-disciplinary collaboration, but physical proximity and collegial values go a long way in enabling collaboration. The multidisciplinary nature of universities is of historic and growing importance to computer science, which interfaces with so many other fields.
- *Universities are "open."* This characteristic of universities, which is true both literally and figuratively, can pay enormous unanticipated dividends. Chance interactions in an open environment can change the world; for example, when Microsoft founders Paul Allen and Bill Gates were students at Seattle's Lakeside School in the early 1970s, they were exposed to computing and computer science at the University of Washington and a university spinoff company, Computer Center Corporation.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

ogy transfer (see Box 7). Federal support for university research drives this process (1-6). In top university computer science programs, over half of all graduate students receive financial support from the federal government, mostly in the form of research assistantships. In addition, most of the funding for research equipment—that is, research infrastructure—comes from federal agencies. Industry also contributes significantly to equipment but is usually attracted by existing research excellence and collaborations. Thus, by placing infrastructure in universities, the federal government directly and indirectly makes possible hands-on learning experiences for countless young engineers and scientists, as well as enabling university researchers to continue their work (1-6).

HALLMARKS OF FEDERALLY SPONSORED IT RESEARCH

As discussed below, the hallmarks of federally sponsored IT research include scale, diversity, vision, and flexibility.

1. *Federal programs have been effective in supporting the construction of large-scale systems and testbeds that have motivated research and demonstrated the feasibility of new technological approaches* (1-3).

Some research challenges are too large and require too much research infrastructure to be carried out by small, local research groups (6). In IT research, as in other areas of scientific investigation, federal programs have played an important role in stimulating and supporting large-scale efforts. DARPA's decision to construct a packet-switched network (called the ARPANET) to link computers at its many contractor sites prompted diverse, high-impact research on networking protocols, the design of packet switches and routers, software structures for managing large networks (such as the Domain Name System), and applications (such as remote log-in, file transfer, and ultimately the Web). Moreover, by constructing a successful system, DARPA demonstrated the value of large-scale packet-switched networks, motivating subsequent deployment of other networks—such as the NSF's NSFnet, which ultimately served as the foundation of the Internet—and also a series of high-speed networking testbeds (1,3).

Much of the success of major system-building efforts derives from their ability to bring together large groups of researchers from universities and industry that develop a common vocabulary, share ideas, and create a critical mass of people who subsequently extend the technology (2,6).

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

2. *Computing research has benefited from diverse modes of research sponsored by different federal agencies* (1-3).

Funding for research in computing has been provided by various federal agencies—most notably DARPA and NSF, but also including other parts of the Department of Defense (DOD) besides DARPA, and other federal agencies such as NASA, DOE, and the National Institutes of Health (NIH; in particular through the National Library of Medicine). Complementary investments have supported technology transfer to industry (e.g., activities of the National Institute of Standards and Technology, or NIST). Funding agencies have continually evolved in order to match their structures better to the needs of the research and policy-making communities (1). (See Box 8.)

In supporting research, these agencies pursue different objectives and employ different mechanisms. In contrast to NSF, for example—which has a mandate to support a very broad research agenda—"mission agencies" tend to focus on topics that appear to have the greatest relevance to their specific missions. Additionally, the early DARPA programs chose to concentrate large research awards in so-called centers of excellence (many of which over time have matured into some of the nation's leading university computer science programs), while NSF and the Office of Naval Research have supported individual researchers at a more diverse set of institutions (1). NSF has been active in supporting educational and research needs more broadly, awarding graduate student fellowships and providing funding for research equipment and infrastructure.

CSTB has recognized the effective leadership of NSF and DARPA, calling on them to step up to larger roles (2; p. 11):

> The programs run by [NSF and DARPA] should complement one another and should together [do the following]:
>
> - Support both theoretical and experimental work;
> - Offer awards in a variety of sizes (small, medium, and large) to support individual investigators, small teams of researchers, and larger collaborations;
> - Investigate a range of approaches to large-scale systems problems, such as improved software design methodologies, system architecture, reusable code, and biological and economic models . . . ;
> - Attempt to address the full scope of large-scale systems issues, including scalability, heterogeneity, trustworthiness, flexibility, and predictability; and
> - Give academic researchers some form of access to large-scale systems for studying and demonstrating new approaches.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

BOX 8
Federal Agency Evolution

In response to proposals by Vannevar Bush and others for an organization to fund basic research, especially in universities, the U.S. Congress established the National Science Foundation (NSF) in 1950 (1). A few years earlier, the U.S. Navy had founded the Office of Naval Research to draw on science and engineering resources in the universities.

In the early 1950s, during an intense phase of the Cold War, the military services became the preeminent funders of computing and communications research. The Soviet Union's launching of Sputnik in 1957 raised fears in Congress and the country that the Soviets had forged ahead of the United States in advanced technology. In response, the U.S. Department of Defense, pressured by the Eisenhower administration, established the Advanced Research Projects Agency (ARPA, now DARPA) to fund technological projects with military implications. In 1962 DARPA created the Information Processing Techniques Office (IPTO), whose initial research agenda gave priority to further development of computers for command-and-control systems.

With the passage of time, new organizations have emerged, and old ones have often been reformed or reinvented to respond to new national imperatives and counter bureaucratic trends (2). DARPA's IPTO has transformed itself several times to bring greater coherence to its research efforts and to respond to technological developments and changes in perceived national needs for IT.

In 1967 NSF established the Office of Computing Activities, and in 1986 it formed the Computer and Information Science and Engineering Directorate to advance and coordinate support for research, education, and infrastructure in computing (1). In the 1980s NSF, which customarily has focused on fundamental research in universities, also began to encourage joint university-industry research centers through its Engineering Research Centers program (these centers focus on research and education in the context of long-time-horizon, complex engineering challenges[1]) and its Science and Technology Center program (aimed at long-term research in areas that are new or that can bridge disciplines and/or institutions and sectors[2]).

With the growth in the IT sector and corresponding IT development together with the maturation of the field of computer science, more recent federal funding has been characterized by a series of multiagency, long-term, high-risk initiatives. The first was the High Performance Computing and Communications Initiative, which emerged in the late 1980s and broadened through the mid-1990s (1,3). By the late 1990s and the establishment of the multiagency Information Technology for the Twenty-First Century initiative (in NSF, the Information Technology Research initiative), social science research—relating IT innovation to the people who use IT—was an important complement to the science and technology research per se (3,8).

---

[1]See <http://www.eng.nsf.gov/eec/erc.htm>. Accessed June 2, 2003.
[2]See <http://www.nsf.gov/od/oia/programs/stc/>. Accessed June 2, 2003.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

> Given the wide circle of agencies interested in and involved with IT research and the even wider circle coming to depend on large-scale IT systems, the NSF and DARPA should attempt to involve in their research other federal agencies . . . that operate large-scale IT systems and would benefit from advances in their design. Such involvement could provide a means for researchers to gain access to operational systems for analytical and experimental purposes.

The diversity of research funding objectives and program management styles offers many benefits (1,3). It helps ensure exploration of a diverse set of research topics and consideration of a range of applications. For example, DARPA, NASA, and NIH (in addition to NSF) have all supported work in expert systems. However, because the systems have had different applications—decision aids for pilots, tools for determining the structure of molecules on other planets, and medical diagnostics—each agency has supported different groups of researchers who tried different approaches. And no one's judgment is infallible. If one agency declines to support a particular topic, researchers have other sources of funding.

3. *Visionary program managers who were willing to take risks have been a hallmark of many of the highest-impact federal research initiatives* (1,3).

The program manager is responsible for initiating, funding, and overseeing research programs. The funding and management styles of program managers at DARPA during the 1960s and 1970s, for example, reflected an ability to marry visions for technological progress with strong technical expertise and an understanding of the uncertainties of the research process (1,3). Many of these program managers and program office directors were recruited from universities and industrial research laboratories for limited tours of duty and were themselves leading researchers. With close ties to the field, they were trusted by—and trusted—the research community. They tended to lay down broad guidelines for new research areas and to draw specific project proposals from principal investigators. They were willing to place bets—to pursue high-risk/high-gain projects.

This style of funding and management allowed researchers room to pursue new venues of inquiry. The funding style resulted in advances in areas as diverse as computer graphics, artificial intelligence, networking, and computer architecture. As that experience illustrates, because unanticipated outcomes of research are so valuable, federal mechanisms for funding and managing research need to recognize the inherent uncertainties and build in enough flexibility to accommodate midcourse changes (1,3).

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

LOOKING FORWARD

Federal funding agencies will have to continue to adjust their strategies and tactics as national needs and imperatives change. Today there is an escalation in concern about homeland security, the globalization of industry, a rise of commodity IT products and an IT mass market, the growing dependence of economic and social activity on networking and distributed computing capabilities, and a variety of industry retrenchments. Coevolution with industry thus means different things for federally funded computing research today than it did in the middle to late decades of the 20th century.

Challenges as well as opportunities have grown: computer science is a larger field with more subdisciplines; telecommunications is increasingly intertwined with computing while evolving across multiple media;[15] the interdisciplinary problems that engage computer science and telecommunications are broader-ranging; and the number of hard problems—reflecting growth in scale, complexity, and interactions with people—has increased. Evolving capabilities motivate a range of stretch goals that can help realize the potential of information technology as a human enabler.[16] Examples include new forms of prosthetics (beginning with systems that can hear, speak, or see as well as a person can) and better ways to observe or participate in activities from a distance (i.e., telepresence).

These circumstances imply that the challenge to federal research program managers has also grown. For example, while IT is at the core of a number of interdisciplinary programs (such as the multiagency Digital Libraries Initiative and NSF's Digital Government and Computing and Social System programs), it takes more work to review proposals for interdisciplinary work and to assure its quality. It may thus be more important to engage IT-using organizations in research projects, which may involve more work for the researchers (2). The growth in opportunities at the intersection of computing and biology, for example, or even computing and the arts—both topics of CSTB projects[17]—suggests new horizons

---

[15]Innovations are enhancing the potential of optical fiber, various forms of wireless, and even older media, such as copper.

[16]These and other problems were outlined by Jim Gray in his 1998 A.M. Turing Award lecture. See Jim Gray. 1998. "What's Next? A Few Remaining Problems in Information Technology." Available online at <http://research.Microsoft.com/~Gray/talks>. Accessed June 9, 2003.

[17]The project on computing and the arts and design was completed in early 2003. See Computer Science and Telecommunications Board, National Research Council. 2003. *Beyond Productivity: Information Technology, Innovation, and Creativity.* National Academies Press, Washington, D.C.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

for IT innovation that depend on the nurturing that is available through university-based research programs.

The challenges confronting program managers underscore the need to attract talent from universities and industry to such public service positions. Past advances fostered by federal funding leveraged the energies and wisdom of people who went from universities and industry into the government, for at least a limited period. It is ironic that their success has increased the incentives for researchers to stay in universities or to try their hand in industry instead of cultivating the field as program managers.

Government support for IT research will also be shaped by categories of problems in which it has a special interest. The events of September 11, 2001, remind us that computer and communications security, constrained by market failure, has always depended on federal investments. But so, too, has research in human-computer interaction, another arena in which market forces have been limited (8) and where the rise of e-government reinforces long-standing government interest associated with its own applications.[18] The post-September 11 focus on homeland security and intelligence analysis also puts a spotlight on supercomputing architectures, numerical analysis, parallel programming languages and tools, and other areas in which IT advances have flowed from scientific and engineering computing needs within the research community at large—and in which purely commercial development was unlikely at best (1,3).

The downturn in the telecommunications industry presents opportunities for the government to stimulate new directions through its support for research. We may see a consolidation and a loss of viable competition, or a realignment of the sector boundaries to better reflect economic realities. Government funding, supporting the development of open standards, can help shape the structure of industry.[19] Given the "chicken-and-egg" tension shaping advances in infrastructure and applications, government support for exploration of new kinds of applications can have great impact.[20] The government can encourage competition by supporting the definition of critical interfaces and demonstrations of feasibil-

---

[18]Computer Science and Telecommunications Board, National Research Council. 2002. *Information Technology Research, Innovation, and E-Government*. National Academy Press, Washington, D.C.

[19]See Computer Science and Telecommunications Board, National Research Council, 2001, *The Internet's Coming of Age*, National Academy Press, Washington, D.C.; and Computer Science and Telecommunications Board, National Research Council, 2002, *Broadband: Bringing Home the Bits*, National Academy Press, Washington, D.C.

[20]This was demonstrated by the evolution of the early Internet and Web, involving development and refinement of both the underlying infrastructure and a suite of compelling

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

ity for open standards, and it can demonstrate new architectures through field trials and testbeds. This role was critical in the emergence of the Internet, and the relevance and importance of this sort of leadership have not waned.[21]

More generally, the 2001-2002 downturn in the economy and the crisis in the telecommunications industry caused a reduction in investment across all of IT. Spending remained down in 2003, and internal investment has dropped accordingly. Venture and equity capital has also become harder to obtain in the IT industries. In times such as these, research, especially longer-term research, is an obvious target for cost cutting. But if we as a nation do not continue to invest in the foundations of innovation, we run the risk that when an improving economy justifies an increase in investment, there may be few ideas in which to invest. For that reason this time is especially important for government-sponsored research.

Today's research investments are essential to tomorrow's world leadership in IT. From its position of leadership today—reinforced by an aggregation of universities, companies, government programs, and talent—the United States is better positioned than other nations are to make the most of nonappropriable research (and even appropriable research). Properly managed, publicly funded research in IT will continue to create important new technologies and industries, some of them unimagined today. The process will continue to take 10 to 15 years from the inception of a new idea to the creation of a billion-dollar industry. Without continued federal investment in fundamental research there would still be innovation, but the quantity and range of new ideas for U.S. industry to draw from would be greatly diminished—as would the flow of people edu-

---

applications by researchers focused not only on IT but also on other fields of science and engineering in which people used IT. The Internet probably could never have developed commercially without this phase of government-supported experimentation and refinement coordinated between infrastucture and applications. For a discussion of new opportunities in the support of applications, see Computer Science and Telecommunications Board, National Research Council, 2002, *Broadband: Bringing Home the Bits*, National Academy Press, Washington, D.C.

[21]For a discussion of the role of government in setting a vision, see Computer Science and Telecommunications Board, National Research Council, 1994, *Realizing the Information Future: The Internet and Beyond*, National Academy Press, Washington, D.C. For a discussion of government leadership and the importance of government funding of research as a policy tool, see Computer Science and Telecommunications Board, National Research Council, 1996, *The Unpredictable Certainty: Information Infrastructure Through 2000*, National Academy Press, Washington, D.C.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

cated at the forefront, the most important product of the nation's research universities (1-8).

The lessons of history are clear, as many CSTB studies in the past decade have shown, and many of those lessons are relevant to 21st-century realities. A complex partnership among government, industry, and universities has made the United States the world leader in IT, and information technology has become essential to our national security and economic and social well-being. Turn-of-the-century turmoil and structural changes in IT industries have diminished their inherently limited capacity to support fundamental IT research. The role of the federal government in sponsoring fundamental research in IT—largely university-based—has been and will continue to be essential.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

2

Excerpts from Earlier CSTB Reports

This section contains excerpts from three CSTB reports:

- *Making IT Better: Expanding Information Technology Research to Meet Society's Needs* (2000),
- *Funding a Revolution: Government Support for Computing Research* (1999), and
- *Evolving the High Performance Computing and Communications Initiative to Support the Nation's Information Infrastructure* (1995).

While this synthesis report is based on all the CSTB reports listed in Box 1 in the "Summary and Recommendations," the excerpts from these three reports are the most general and broad. To keep this report to a reasonable length, nothing was excerpted from the other five reports. Readers are encouraged to read all eight reports, which can be found online at <http://www.nap.edu>.

For the sake of simplicity and organizational clarity, footnotes and reference citations appearing in the original texts have been omitted from the reprinted material that follows. A bar in the margins beside the excerpted material is used to indicate that it is extracted text. Section heads show the topics addressed.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

MAKING IT BETTER: EXPANDING INFORMATION TECHNOLOGY RESEARCH TO MEET SOCIETY'S NEEDS (2000)

CITATION: Computer Science and Telecommunications Board (CSTB), National Research Council. 2000. *Making IT Better: Expanding Information Technology Research to Meet Society's Needs.* National Academy Press, Washington, D.C.

The Many Faces of Information Technology Research

(From pp. 23-26): IT research takes many forms. It consists of both theoretical and experimental work, and it combines elements of science and engineering. Some IT research lays out principles or constraints that apply to all computing and communications systems; examples include theorems that show the limitations of computation (what can and cannot be computed by a digital computer within a reasonable time) or the fundamental limits on capacities of communications channels. Other research investigates different classes of IT systems, such as user interfaces, the Web, or electronic mail (e-mail). Still other research deals with issues of broad applicability driven by specific needs. For example, today's high-level programming languages (such as Java and C) were made possible by research that uncovered techniques for converting the high-level statements into machine code for execution on a computer. The design of the languages themselves is a research topic: how best to capture a programmer's intentions in a way that can be converted to efficient machine code. Efforts to solve this problem, as is often the case in IT research, will require invention and design as well as the classical scientific techniques of analysis and measurement. The same is true of efforts to develop specific and practical modulation and coding algorithms that approach the fundamental limits of communication on some channels. The rise of digital communication, associated with computer technology, has led to the irreversible melding of what were once the separate fields of communications and computers, with data forming an increasing share of what is being transmitted over the digitally modulated fiber-optic cables spanning the nation and the world.

Experimental work plays an important role in IT research. One modality of research is the design experiment, in which a new technique is proposed, a provisional design is posited, and a research prototype is built in order to evaluate the strengths and weaknesses of the design. Although much of the effect of a design can be anticipated using analytic techniques, many of its subtle aspects are uncovered only when the prototype is studied. Some of the most important strides in IT have been made through such experimental research. Time-sharing, for example, evolved

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

in a series of experimental systems that explored different parts of the technology. How are a computer's resources to be shared among several customers? How do we ensure equitable sharing of resources? How do we insulate each user's program from the programs of others? What resources should be shared as a convenience to the customers (e.g., computer files)? How can the system be designed so it's easy to write computer programs that can be time-shared? What kinds of commands does a user need to learn to operate the system? Although some of these tradeoffs may succumb to analysis, others—notably those involving the user's evaluation and preferences—can be evaluated only through experiment.

Ideas for IT research can be gleaned both from the research community itself and from applications of IT systems. The Web, initiated by physicists to support collaboration among researchers, illustrates how people who use IT can be the source of important innovations. The Web was not invented from scratch; rather, it integrated developments in information retrieval, networking, and software that had been accumulating over decades in many segments of the IT research community. It also reflects a fundamental body of technology that is conducive to innovation and change. Thus, it advanced the integration of computing, communications, and information. The Web also embodies the need for additional science and technology to accommodate the burgeoning scale and diversity of IT users and uses: it became a catalyst for the Internet by enhancing the ease of use and usefulness of the Internet, it has grown and evolved far beyond the expectations of its inventors, and it has stimulated new lines of research aimed at improving and better using the Internet in numerous arenas, from education to crisis management.

Progress in IT can come from research in many different disciplines. For example, work on the physics of silicon can be considered IT research if it is driven by problems related to computer chips; the work of electrical engineers is considered IT research if it focuses on communications or semiconductor devices; anthropologists and other social scientists studying the uses of new technology can be doing IT research if their work informs the development and deployment of new IT applications; and computer scientists and computer engineers address a widening range of issues, from generating fundamental principles for the behavior of information in systems to developing new concepts for systems. Thus, IT research combines science and engineering, even though the popular—and even professional—association of IT with systems leads many people to concentrate on the engineering aspects. Fine distinctions between the science and engineering aspects may be unproductive: computer science is special because of how it combines the two, and the evolution of both is key to the well-being of IT research.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Implications for the Research Enterprise

(From pp. 42-43): The trends in IT suggest that the nation needs to reinvent IT research and develop new structures to support, conduct, and manage it. . . .

As IT permeates many more real-world applications, additional constituencies need to be brought into the research process as both funders and performers of IT research. This is necessary not only to broaden the funding base to include those who directly benefit from the fruits of the research, but also to obtain input and guidance. An understanding of business practices and processes is needed to support the evolution of e-commerce; insight from the social sciences is needed to build IT systems that are truly user-friendly and that help people work better together. No one truly understands where new applications such as e-commerce, electronic publishing, or electronic collaboration are headed, but business development and research together can promote their arrival at desirable destinations.

Many challenges will require the participation and insight of the end user and the service provider communities. They have a large stake in seeing these problems addressed, and they stand to benefit most directly from the solutions. Similarly, systems integrators would benefit from an improved understanding of systems and applications because they would become more competitive in the marketplace and be better able to meet their estimates of project cost and time. Unlike vendors of component technologies, systems integrators and end users deal with entire information systems and therefore have unique perspectives on the problems encountered in developing systems and the feasibility of proposed solutions. Many of the end-user organizations, however, have no tradition of conducting IT research—or technological research of any kind, in fact—and they are not necessarily capable of doing so effectively; they depend on vendors for their technology. Even so, their involvement in the research process is critical. Vendors of equipment and software have neither the requisite experience and expertise nor the financial incentives to invest heavily in research on the challenges facing end-user organizations, especially the challenges associated with the social applications of IT. Of course, they listen to their customers as they refine their products and strategies, but those interactions are superficial compared with the demands of the new systems and applications. Finding suitable mechanisms for the participation of end users and service providers, and engaging them productively, will be a big challenge for the future of IT research.

Past attempts at public-private partnerships, as in the emerging arena of critical infrastructure protection, show it is not so easy to get the public

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

and private sectors to interact for the purpose of improving the research base and implementation of systems: the federal government has a responsibility to address the public interest in critical infrastructure, whereas the private sector owns and develops that infrastructure, and conflicting objectives and time horizons have confounded joint exploration. As a user of IT, the government could play an important role. Whereas historically it had limited and often separate programs to support research and acquire systems for its own use, the government is now becoming a consumer of IT on a very large scale. Just as IT and the widespread access to it provided by the Web have enabled businesses to reinvent themselves, IT could dramatically improve operations and reduce the costs of applications in public health, air traffic control, and social security; government agencies, like private-sector organizations, are turning increasingly to commercial, off-the-shelf technology.

Universities will play a critical role in expanding the IT research agenda. The university setting continues to be the most hospitable for higher-risk research projects in which the outcomes are very uncertain. Universities can play an important role in establishing new research programs for large-scale systems and social applications, assuming that they can overcome long-standing institutional and cultural barriers to the needed cross-disciplinary research. Preserving the university as a base for research and the education that goes with it would ensure a workforce capable of designing, developing, and operating increasingly sophisticated IT systems. A booming IT marketplace and the lure of large salaries in industry heighten the impact of federal funding decisions on the individual decisions that shape the university environment: as the key funders of university research, federal programs send important signals to faculty and students.

The current concerns in IT differ from the competitiveness concerns of the 1980s: the all-pervasiveness of IT in everyday life raises new questions of how to get from here to there—how to realize the exciting possibilities, not merely how to get there first. A vital and relevant IT research program is more important than ever, given the complexity of the issues at hand and the need to provide solid underpinnings for the rapidly changing IT marketplace.

(From p. 93): Several underlying trends could ultimately limit the nation's innovative capacity and hinder its ability to deploy the kinds of IT systems that could best meet personal, business, and government needs. First, expenditures on research by companies that develop IT goods and services and by the federal government have not kept pace with the expanding array of IT. The disincentives to long-term, fundamental research have become more numerous, especially in the private sector,

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

which seems more able to lure talent from universities than the other way around. Second, and perhaps most significantly, IT research investments continue to be directed at improving the performance of IT components, with limited attention to systems issues and application-driven needs. Neither industry nor academia has kept pace with the problems posed by the large-scale IT systems used in a range of social and business contexts—problems that require fundamental research. . . . New mechanisms may be needed to direct resources to these growing problem areas.

(From pp. 6-9): Neither large-scale systems nor social applications of IT are adequately addressed by the IT research community today. Most IT research is directed toward the *components* of IT systems: the microprocessors, computers, and networking technologies that are assembled into large systems, as well as the software that enables the components to work together. This research nurtures the essence of IT, and continued work is needed in all these areas. But component research needs to be viewed as part of a much larger portfolio, in which it is complemented by research aimed directly at improving large-scale systems and the social applications of IT. The last of these includes some work (such as computer-supported cooperative work and human-computer interaction) traditionally viewed as within the purview of computer science. Research in all three areas—components, systems, and social applications—will make IT systems better able to meet society's needs, just as in the medical domain work is needed in biology, physiology, clinical medicine, and epidemiology to make the nation's population healthier.

Research on large-scale systems and the social applications of IT will require new modes of funding and performing research that can bring together a broad set of IT researchers, end users, system integrators, and social scientists to enhance the understanding of operational systems. Research in these areas demands that researchers have access to operational large-scale systems or to testbeds that can mimic the performance of much larger systems. It requires additional funding to support sizable projects that allow multiple investigators to experiment with large IT systems and develop suitable testbeds and simulations for evaluating new approaches and that engage an unusually diverse range of parties. Research by individual investigators will not, by itself, suffice to make progress on these difficult problems.

Today, most IT research fails to incorporate the diversity of perspectives needed to ensure advances on large-scale systems and social applications. Within industry, it is conducted largely by vendors of IT components: companies like IBM, Microsoft, and Lucent Technologies. Few of the companies that are engaged in providing IT services, in integrating large-scale systems (e.g., Andersen Consulting [now Accenture], EDS, or

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Lockheed Martin), or in developing enterprise software (e.g., Oracle, SAP, PeopleSoft) have significant research programs. Nor do end-user organizations (e.g., users in banking, commerce, education, health care, and manufacturing) tend to support research on IT, despite their increasing reliance on IT and their stake in the way IT systems are molded. Likewise, there is little academic research on large-scale systems or social applications. Within the IT sector, systems research has tended to focus on improving the performance and lowering the costs of IT systems rather than on improving their reliability, flexibility, or scalability (although systems research is slated to receive more attention in new funding programs). Social applications present an even greater opportunity and have the potential to leverage research in human-computer interaction, using it to better understand how IT can support the work of individuals, groups, and organizations. Success in this area hinges on interdisciplinary research, which is already being carried out on a small scale.

One reason more work has not been undertaken in these areas is lack of sufficient funding. More fundamentally, the problems evident today did not reach critical proportions until recently. . . . From a practical perspective, conducting the types of research advocated here is difficult. Significant cultural gaps exist between researchers in different disciplines and between IT researchers and the end users of IT systems.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

FUNDING A REVOLUTION: GOVERNMENT SUPPORT FOR COMPUTING RESEARCH (1999)

CITATION: Computer Science and Telecommunications Board (CSTB), National Research Council. 1999. *Funding a Revolution: Government Support for Computing Research.* National Academy Press, Washington, D.C.

(From p. 1): The computer revolution is not simply a technical change; it is a sociotechnical revolution comparable to an industrial revolution. The British Industrial Revolution of the late 18th century not only brought with it steam and factories, but also ushered in a modern era characterized by the rise of industrial cities, a politically powerful urban middle class, and a new working class. So, too, the sociotechnical aspects of the computer revolution are now becoming clear. Millions of workers are flocking to computing-related industries. Firms producing microprocessors and software are challenging the economic power of firms manufacturing automobiles and producing oil. Detroit is no longer the symbolic center of the U.S. industrial empire; Silicon Valley now conjures up visions of enormous entrepreneurial vigor. Men in boardrooms and gray flannel suits are giving way to the casually dressed young founders of start-up computer and Internet companies. Many of these entrepreneurs had their early hands-on computer experience as graduate students conducting federally funded university research.

As the computer revolution continues and private companies increasingly fund innovative activities, the federal government continues to play a major role, especially by funding research. Given the successful history of federal involvement, several questions arise: Are there lessons to be drawn from past successes that can inform future policy making in this area? What future roles might the government play in sustaining the information revolution and helping to initiate other technological developments?

Lessons from History

(From pp. 5-13): Why has federal support been so effective in stimulating innovation in computing? Although much has depended on the unique characteristics of individual research programs and their participants, several common factors have played an important part. Primary among them is that federal support for research has tended to *complement*, rather than preempt, industry investments in research. Effective federal research has concentrated on work that industry has limited incentive to pursue: long-term, fundamental research; large system-building efforts that require the talents of diverse communities of scientists and engi-

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

neers; and work that might displace existing, entrenched technologies. Furthermore, successful federal programs have tended to be organized in ways that accommodate the uncertainties in scientific and technological research. Support for computing research has come from a diversity of funding agencies; program managers have formulated projects broadly where possible, modifying them in response to preliminary results; and projects have fostered productive collaboration between universities and industry. The lessons below expand on these factors. The first three lessons address the complementary nature of government- and industry-sponsored research; the final four highlight elements of the organizational structure and management of effective federally funded research programs. . . .

1. Government supports long-range, fundamental research that industry cannot sustain.

Federally funded programs have been successful in supporting long-term research into fundamental aspects of computing, such as computer graphics and artificial intelligence, whose practical benefits often take years to demonstrate. Work on speech recognition, for example, which was begun in the early 1970s (some started even earlier), took until 1997 to generate a successful product for enabling personal computers to recognize continuous speech. Similarly, fundamental algorithms for shading three-dimensional graphics images, which were developed with defense funding in the 1960s, entered consumer products only in the 1990s, though they were available in higher-performance machines much earlier. These algorithms are now used in a range of products in the health care, entertainment, and defense industries.

Industry does fund some long-range work, but the benefits of fundamental research are generally too distant and too uncertain to receive significant industry support. Moreover, the results of such work are generally so broad that it is difficult for any one firm to capture them for its own benefit and also prevent competitors from doing so. . . . Not surprisingly, companies that have tended to support the most fundamental research have been those, like AT&T Corporation and IBM Corporation, that are large and have enjoyed a dominant position in their respective markets. As the computing industry has become more competitive, even these firms have begun to link their research more closely with corporate objectives and product development activities. Companies that have become more dominant, such as Microsoft Corporation and Intel Corporation, have increased their support for fundamental research.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

2. Government supports large system-building efforts that have advanced technology and created large communities of researchers.

In addition to funding long-term fundamental research, federal programs have been effective in supporting the construction of large systems that have both motivated research and demonstrated the feasibility of new technological approaches. The Defense Advanced Research Projects Agency's (DARPA's) decision to construct a packet-switched network (called the ARPANET) to link computers at its many contractor sites prompted considerable research on networking protocols and the design of packet switches and routers. It also led to the development of structures for managing large networks, such as the domain name system, and development of useful applications, such as e-mail. Moreover, by constructing a successful system, DARPA demonstrated the value of large-scale packet-switched networks, motivating subsequent deployment of other networks, like the National Science Foundation's NSFnet, which formed the basis of the Internet.

Efforts to build large systems demonstrate that, especially in computing, innovation does not flow simply and directly from research, through development, to deployment. Development often precedes research, and research rationalizes, or explains, technology developed earlier through experimentation. Hence attempts to build large systems can identify new problems that need to be solved. Electronic telecommunications systems were in use long before Claude Shannon developed modern communications theory in the late 1940s, and the engineers who developed the first packet switches for routing messages through the ARPANET advanced empirically beyond theory. Building large systems generated questions for research, and the answers, in turn, facilitated more development.

Much of the success of major system-building efforts derives from their ability to bring together large groups of researchers from academia and industry who develop a common vocabulary, share ideas, and create a critical mass of people who subsequently extend the technology. Examples include the ARPANET and the development of the Air Force's Semi-Automatic Ground Environment (SAGE) project in the 1950s. Involving researchers from MIT, IBM, and other research laboratories, the SAGE project sparked innovations ranging from real-time computing to core memories that found widespread acceptance throughout the computer industry. Many of the pioneers in computing learned through hands-on experimentation with SAGE in the 1950s and early 1960s. They subsequently staffed the companies and laboratories of the nascent computing and communications revolution. The impact of SAGE was felt over the course of several decades.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

3. Federal research funding has expanded on earlier industrial research.

In several cases, federal research funding has been important in advancing a technology to the point of commercialization after it was first explored in an industrial research laboratory. For example, IBM pioneered the concept of relational databases but did not commercialize the technology because of its perceived potential to compete with more-established IBM products. National Science Foundation (NSF)-sponsored research at UC-Berkeley allowed continued exploration of this concept and brought the technology to the point that it could be commercialized by several start-up companies—and more-established database companies (including IBM). This pattern was also evident in the development of reduced instruction set computing (RISC). Though developed at IBM, RISC was not commercialized until DARPA funded additional research at UC-Berkeley and Stanford University as part of its Very Large Scale Integrated Circuit (VLSI) program of the late 1970s and early 1980s. A variety of companies subsequently brought RISC-based products to the marketplace, including IBM, the Hewlett-Packard Company, the newly formed Sun Microsystems, Inc., and another start-up, MIPS Computer Systems. For both relational databases and VLSI, federal funding helped create a community of researchers who validated and improved on the initial work. They rapidly diffused the technology throughout the community, leading to greater competition and more rapid commercialization.

4. Computing research has benefited from diverse sources of government support.

Research in computing has been supported by multiple federal agencies, including the Department of Defense (DOD)—most notably the Defense Advanced Research Projects Agency and the military services—the National Science Foundation, National Aeronautics and Space Administration (NASA), Department of Energy (DOE), and National Institutes of Health (NIH). Each has its own mission and means of supporting research. DARPA has tended to concentrate large research grants in so-called centers of excellence, many of which over time have matured into some of the country's leading academic computer departments. The Office of Naval Research (ONR) and NSF, in contrast, have supported individual researchers at a more diverse set of institutions. They have awarded numerous peer-review grants to individual researchers, especially in universities. NSF has also been active in supporting educational and research needs more broadly, awarding graduate student fellowships and providing funding for research equipment and infrastructure. Each of these organizations employs a different set of mechanisms to support research,

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

from fundamental research to mission-oriented research and development projects, to procurement of hardware and software.

Such diversity offers many benefits. It not only provides researchers with many potential sources of support, but also helps ensure exploration of a diverse set of research topics and consideration of a range of applications. DARPA, NASA, and NIH have all supported work in expert systems, for example, but because the systems have had different applications—decision aids for pilots, tools for determining the structure of molecules on other planets, and medical diagnostics—each agency has supported different groups of researchers who tried different approaches.

Perhaps more importantly, no single approach to investing in research is by itself a sufficient means of stimulating innovation; each plays a role in the larger system of innovation. Different approaches work in concert, ensuring continued support for research areas as they pass through subsequent stages of development. Organizations such as NSF and ONR often funded seed work in areas that DARPA, with its larger contract awards, later magnified and expanded. DARPA's Project MAC, which gave momentum to time-shared computing in the 1960s, for example, built on earlier NSF-sponsored work on MIT's Compatible Time-Sharing System. Conversely, NSF has provided continued support for projects that DARPA pioneered but was unwilling to sustain after the major research challenges were resolved. For example, NSF funds the Metal Oxide Semiconductor Implementation Service (MOSIS)—a system developed at Xerox PARC and institutionalized by DARPA that provides university researchers with access to fast-turnaround semiconductor manufacturing services. Once established, this program no longer matched DARPA's mission to develop leading-edge technologies, but it did match NSF's mission to support university education and research infrastructure. Similarly, NSF built on DARPA's pioneering research on packet-switched networks to construct the NSFnet, a precursor to today's Internet.

5. Strong program managers and flexible management structures have enhanced the effectiveness of computing research.

Research in computing, as in other fields, is a highly unpredictable endeavor. The results of research are not evident at the start, and their most important contributions often differ from those originally envisioned. Few expected that the Navy's attempt to build a programmable aircraft simulator in the late 1940s would result in the development of the first real-time digital computer (the Whirlwind); nor could DARPA program managers have anticipated that their early experiments on packet switching would evolve into the Internet and later the World Wide Web.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

The potential for unanticipated outcomes of research has two implications for federal policy. First, it suggests that measuring the results of federally funded research programs is extremely difficult. Projects that appear to have failed often make significant contributions to later technology development or achieve other objectives not originally envisioned. Furthermore, research creates many intangible products, such as knowledge and educated researchers whose value is hard to quantify. Second, it implies that federal mechanisms for funding and managing research need to recognize the uncertainties inherent in computing research and to build in sufficient flexibility to accommodate mid-course changes and respond to unanticipated results.

A key element in agencies' ability to maintain flexibility in the past has been their program managers, who have responsibility for initiating, funding, and overseeing research programs. The funding and management styles of program managers at DARPA during the 1960s and 1970s, for example, reflected an ability to marry visions for technological progress with strong technical expertise and an understanding of the uncertainties of the research process. Many of these program managers and office directors were recruited from academic and industry research laboratories for limited tours of duty. They tended to lay down broad guidelines for new research areas and to draw specific project proposals from principal investigators, or researchers, in academic computer centers. This style of funding and management resulted in the government stimulating innovation with a light touch, allowing researchers room to pursue new avenues of inquiry. In turn, it helped attract top-notch program managers to federal agencies. With close ties to the field and its leading researchers, they were trusted by—and trusted in—the research community.

This funding style resulted in great advances in areas as diverse as computer graphics, artificial intelligence, networking, and computer architectures. Although mechanisms are clearly needed to ensure accountability and oversight in government-sponsored research, history demonstrates the benefits of instilling these values in program managers and providing them adequate support to pursue promising research directions.

6. Collaboration between industry and university researchers has facilitated the commercialization of computing research and maintained its relevance.

Innovation in computing requires the combined talents of university and industry researchers. Bringing them together has helped ensure that industry taps into new academic research and that university researchers

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

understand the challenges facing industry. Such collaboration also helps facilitate the commercialization of technology developed in a university setting. All of the areas described in this report's case studies—relational databases, the Internet, theoretical computer science, artificial intelligence, and virtual reality—involved university and industry participants. Other projects examined, such as SAGE, Project MAC, and very large scale integrated circuits, demonstrate the same phenomenon.

Collaboration between industry and universities can take many forms. Some projects combine researchers from both sectors on the same project team. Other projects involve a transition from academic research laboratories to industry (via either the licensing of key patents or the creation of new start-up companies) once the technology matures sufficiently. As the case studies demonstrate, effective linkages between industry and universities tended to emerge from projects, rather than being thrust upon them. Project teams assembled to build large systems included the range of skills needed for a particular project. University researchers often sought out productive avenues for transferring research results to industry, whether linking with existing companies or starting new ones. Such techniques have often been more effective than explicit attempts to encourage collaboration, many of which have foundered due to the often conflicting time horizons of university and industry researchers.

7. Organizational innovation and adaptation are necessary elements of federal research support.

Over time, new government organizations have formed to support computing research, and organizations have continually evolved in order to better match their structure to the needs of the research and policy-making communities. In response to proposals by Vannevar Bush and others that the country needed an organization to fund basic research, especially in the universities, for example, Congress established the National Science Foundation in 1950. A few years earlier, the Navy founded the Office of Naval Research to draw on science and engineering resources in the universities. In the early 1950s during an intense phase of the Cold War, the military services became the preeminent funders of computing and communications. The Soviet Union's launching of Sputnik in 1957 raised fears in Congress and the country that the Soviets had forged ahead of the United States in advanced technology. In response, the U.S. Department of Defense, pressured by the Eisenhower administration, established the Advanced Research Projects Agency (ARPA, now DARPA) to fund technological projects with military implications. In 1962 DARPA created the Information Processing Techniques Office (IPTO), whose initial re-

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

search agenda gave priority to further development of computers for command-and-control systems.

With the passage of time, new organizations have emerged, and old ones have often been reformed or reinvented to respond to new national imperatives and counter bureaucratic trends. DARPA's IPTO has transformed itself several times to bring greater coherence to its research efforts and to respond to technological developments. NSF in 1967 established the Office of Computing Activities and in 1986 formed the Computer and Information Sciences and Engineering (CISE) Directorate to couple and coordinate support for research, education, and infrastructure in computer science. In the 1980s NSF, which customarily has focused on basic research in universities, also began to encourage joint academic-industrial research centers through its Engineering Research Centers program. With the relative increase in industrial support of research and development in recent years, federal agencies such as NSF have rationalized their funding policies to complement short-term industrial R&D. Federal funding of long-term, high-risk initiatives continues to have a high priority.

As this history suggests, federal funding agencies will need to continue to adjust their strategies and tactics as national needs and imperatives change. The Cold War imperative shaped technological history during much of the last half-century. International competitiveness served as a driver of government funding of computing and communications during the late 1980s and early 1990s. With the end of the Cold War and the globalization of industry, the U.S. computing industries need to maintain their high rates of innovation, and federal structures for managing computing research may need to change to ensure that they are appropriate for this new environment.

Sources of U.S. Success

(From pp. 27-28): That the United States should be the leading country in computing and communications was not preordained. Early in the industry's formation, the United Kingdom was a serious competitor. The United Kingdom was the home of the Difference Engine and later the Analytical Engine, both of which were programmable mechanical devices designed and partially constructed by Charles Babbage and Ada, Countess of Lovelace, in the 19th century. Basic theoretical work defining a universal computer was the contribution of Alan Turing in Cambridge just before the start of World War II. The English defense industry—with Alan Turing's participation—conceived and constructed vacuum tube computers able to break the German military code. Both machines and their accomplishments were kept secret, much like the efforts and suc-

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

cesses of the National Security Agency in this country. After the war, English universities constructed research computers and developed computer concepts that later found significant use in U.S. products. Other European countries, Germany and France in particular, also made efforts to gain a foothold in this new technology.

How then did the United States become a leader in computing? The answer is manifold, and a number of external factors clearly played a role. The state of Europe, England in particular, at the end of World War II played a decisive role, as rebuilding a country and industry is a more difficult task than shifting from a war economy to a consumer economy. The movement of people among universities, industry, and government laboratories at the end of World War II in the United Kingdom and the United States also contributed by spreading the experience gained during the war, especially regarding electronics and computing. American students and scholars who were studying in England as Fulbright Scholars in the 1950s learned of the computer developments that had occurred during the war and that were continuing to advance.

Industrial prowess also played a role. After World War II, U.S. firms moved quickly to build an industrial base for computing. IBM and Remington Rand recognized quite early that electronic computers were a threat to their conventional electromechanical punched-card business and launched early endeavors into computing. . . . Over time, fierce competition and expectations of rapid market growth brought billions in venture money to the industry's inventors and caused a flowering of small high-tech innovators. Rapid expansion of the U.S. marketplace for computing equipment created buyers for new computing equipment. The rapid post-World War II expansion of civilian-oriented industries and financial sources created new demands for data and data processing. Insurance companies and banks were at the forefront of installing early computers in their operations. New companies, such as Engineering Research Associates, Datamatic, and Eckert-Mauchly, as well as established companies in the data processing field, such as IBM and Sperry Rand, saw an opportunity for new products and new markets. The combination of new companies and established ones was a powerful force. It generated fierce competition and provided substantial capital funds.

These factors helped the nation gain an early lead in computing that it has maintained. While firms from other nations have made inroads into computing technology—from memory chips to supercomputers—U.S. firms have continued to dominate both domestic and international markets in most product categories. This success reflects the strength of the nation's innovation system in computing technology, which has continually developed, marketed, and supported new products, processes, and services.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Research and Technological Innovation

(From pp. 28-31): Innovation is generally defined as the process of developing and putting into practice new products, processes, or services. It draws upon a range of activities, including research, product development, manufacturing, and marketing. Although often viewed as a linear, sequential process, innovation is usually more complicated, with many interactions among the different activities and considerable feedback. It can be motivated by new research advances or by recognition of a new market need. Government, universities, and industry all play a role in the innovation process.

Research is a vital part of innovation in computing. In dollar terms, research is just a small part of the innovation process, representing less than one-fifth of the cost of developing and introducing new products in the United States, with preparation of product specifications, prototype development, tooling and equipment, manufacturing start-up, and marketing start-up comprising the remainder. Indeed, computer manufacturers allocated an average of just 20 percent of their research and development budgets to research between 1976 and 1995, with the balance supporting product development. Even in the largest computer manufacturers, such as IBM, research costs are only about 1 to 2 percent of total operating expenses. Nevertheless, research plays a critical role in the innovation process, providing a base of scientific and technological knowledge that can be used to develop new products, processes, and services. This knowledge is used at many points in the innovation process—generating ideas for new products, processes, or services; solving particular problems in product development or manufacturing; or improving existing products, for example. . . .

Traditionally, research expenditures have been characterized as either basic or applied. The term "basic research" is used to describe work that is exploratory in nature, addressing fundamental scientific questions for which ready answers are lacking; the term "applied research" describes activities aimed at exploring phenomena necessary for determining the means by which a recognized need may be met. These terms, at best, distinguish between the motivations of researchers and the manner in which inquiries are conducted, and they are limited in their ability to describe the nature of scientific and technological research. Recent work has suggested that the definition of basic research be expanded to include explicitly both basic scientific research and basic technological research. This definition recognizes the value of exploratory research into basic technological phenomena that can be used in a variety of products. Examples include research on the blue laser, exploration of biosensors, and much of the fundamental work in computer engineering.

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

(From pp. 21-23): Clearly, the future of computing will differ from the history of computing because both the technology and environmental factors have changed. Attempts by companies to align their research activities more closely with product development processes have influenced the role they may play in the innovation process. As the computing industry has grown and the technology has diffused more widely throughout society, government has continued to represent a proportionally smaller portion of the industry.

The Benefits of Public Support of Research

(From pp. 46-47): The development of scientific and technological knowledge is a cumulative process, one that depends on the prompt disclosure of new findings so that they can be tested and, if confirmed, integrated with other bodies of reliable knowledge. In this way open science promotes the rapid generation of further discoveries and inventions, as well as wider practical exploitation of additions to the stock of knowledge.

The economic case for public funding of what is commonly referred to as basic research rests mainly on that insight, and on the observation that business firms are bound to be considerably discouraged by the greater uncertainties surrounding investment in fundamental, exploratory inquiries (compared to commercially targeted R&D), as well as by the difficulties of forecasting when and how such outlays will generate a satisfactory rate of return.

The proposition at issue here is quantitative, not qualitative. One cannot adequately answer the question "Will there be enough?" merely by saying, "There will be some." Economists do not claim that without public patronage (or intellectual property protection), basic research will cease entirely. Rather, their analysis holds that there will not be enough basic research—not as much as would be carried out were individual businesses (like society as a whole) able to anticipate capturing all the benefits of this form of investment. Therefore, no conflict exists between this theoretical analysis and the observation that R&D-intensive companies do indeed fund some exploratory research into fundamental questions. Their motives for this range from developing a capability to monitor progress at the frontiers of science, to identifying ideas for potential lines of innovation that may be emerging from the research of others, to being better positioned to penetrate the secrets of their rivals' technological practices.

Nevertheless, funding research is a long-term strategy, and therefore sensitive to commercial pressures to shift research resources toward advancing existing product development and improving existing processes,

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

rather than searching for future technological options. Large organizations that are less asset constrained, and of course the public sector, are better able to take on the job of pushing the frontiers of science and technology. Considerations of these kinds are important in addressing the issue of how to find the optimal balance for the national research effort between secrecy and disclosure of scientific and engineering information, as well as in trying to adjust the mix of exploratory and applications-driven projects in the national research portfolio.

(From p. 137): Quantifying the benefits of federal research support is a difficult, if not impossible, task for several reasons. First, the output of research is often intangible. Most of the benefit takes the form of new knowledge that subsequently may be instantiated in new hardware, software, or systems, but is itself difficult to measure. At other times, the benefits take the form of educated people who bring new ideas or a fresh perspective to an organization. Second, the delays between the time a research program is conducted and the time the products incorporating the research results are sold make measurement even more difficult. Often, the delays run into decades, making it difficult to tell midcourse how effective a particular program has been. Third, the benefits of a particular research program may not become visible until other technological advances are made. For example, advances in computer graphics did not have widespread effect until suitable hardware was more broadly available for producing three-dimensional graphical images. Finally, projects that are perceived as failures often provide valuable lessons that can guide or improve future research. Even if they fail to reach their original objectives, research projects can make lasting contributions to the knowledge base.

Maintaining University Research Capabilities

(From pp. 139-140): Federal funding has . . . maintained university research capabilities in computing. Universities depend largely on federal support for research programs in computer science and electrical engineering, the two academic disciplines most closely aligned with computing and communications. Since 1973, federal agencies have provided roughly 70 percent of all funding for university research in computer science. In electrical engineering, federal funding has declined from its peak of 75 percent of total university research support in the early 1970s, but still represented 65 percent of such funding in 1995. Additional support has come in the form of research equipment. Universities need access to state-of-the-art equipment in order to conduct research and train students. Although industry contributes some equipment, funding for uni-

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

versity research equipment has come largely from federal sources since the 1960s. Between 1981 and 1995, the federal government provided between 59 and 76 percent of annual research equipment expenditures in computer science and between 64 and 83 percent of annual research equipment expenditures in electrical engineering. Such investments have helped ensure that researchers have access to modern computing facilities and have enabled them to further expand the capabilities of computing and communications systems.

Universities play an important role in the innovation process. They tend to concentrate on research with broad applicability across companies and product lines and to share new knowledge openly. Because they are not usually subject to commercial pressures, university researchers often have greater ability than their industrial counterparts to explore ideas with uncertain long-term payoffs. Although it would be difficult to determine how much university research contributes directly to industrial innovation, it is telling that each of the case studies and other major examples examined in [the source] report—relational databases, the Internet, theoretical computer science, artificial intelligence, virtual reality, SAGE, computer time-sharing, very large scale integrated circuits, and the personal computer—involved the participation of university researchers. Universities play an especially effective role in disseminating new knowledge by promoting open publication of research results. They have also served as a training ground for students who have taken new ideas with them to existing companies or started their own companies. Diffusion of knowledge about relational databases, for instance, was accelerated by researchers at the University of California at Berkeley who published the source code for their Ingres system and made it available free of charge. Several of the lead researchers in this project established companies to commercialize the technology or brought it back to existing firms where they championed its use.

Creating Human Resources

(From pp. 140-141): In addition to supporting the creation of new technology, federal funding for research has also helped create the human resources that have driven the computer revolution. Many industry researchers and research managers claim that the most valuable result of university research programs is educated students—by and large, an outcome enabled by federal support of university research. Federal support for university research in computer science grew from $65 million to $350 million between 1976 and 1995, while federal support for university research in electrical engineering grew from $74 million to $177 million (in constant 1995 dollars). Much of this funding was used to support gradu-

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

ate students. Especially at the nation's top research universities, the studies of a large percentage of graduate students have been supported by federal research contracts. Graduates of these programs, and faculty researchers who received federal funding, have gone on to form a number of companies, including Sun Microsystems, Inc. (which grew out of research conducted by Forest Baskett and Andy Bechtolsheim with sponsorship from DARPA) and Digital Equipment Corporation (founded by Ken Olsen, who participated in the SAGE project). Graduates also staff academic faculties that continue to conduct research and educate future generations of researchers.

Furthermore, the availability of federal research funding has enabled the growth and expansion of computer science and computer engineering departments at U.S. universities, which increased in number from 6 in 1965 to 56 in 1975 and to 148 in 1995. The number of graduate students in computer science also grew dramatically, expanding more than 40-fold from 257 in 1966 to 11,500 in 1995, with the number of Ph.D. degrees awarded in computer science increasing from 19 in 1966 to over 900 in 1995. Even with this growth in Ph.D. production, demand for computing researchers still outstrips the supply in both industry and academia.

Beyond supporting student education and training, federal funding has also been important in creating networks of researchers in particular fields—developing communities of researchers who could share ideas and build on each other's strengths. Despite its defense orientation, DARPA historically encouraged open dissemination of the results of sponsored research, as did other federal agencies. In addition, DARPA and other federal agencies funded large projects with multiple participants from different organizations. These projects helped create entire communities of researchers who continued to refine, adopt, and diffuse new technology throughout the broader computing research community. Development of the Internet demonstrates the benefits of this approach: by funding groups of researchers in an open environment, DARPA created an entire community of users who had a common understanding of the technology, adopted a common set of standards, and encouraged their use broadly. Early users of the ARPANET created a critical mass of people who helped to disseminate the technology, giving the Internet Protocol an important early lead over competing approaches to packet switching.

The Organization of Federal Support: A Historical Review

(From pp. 85-86): Rather than a single, overarching framework of support, federal funding for research in computing has been managed by a set of agencies and offices that carry the legacies of the historical periods in which they were created. Crises such as World War II, Korea, Sputnik,

Copyright © National Academy of Sciences. All rights reserved.

Innovation in Information Technology
http://www.nap.edu/catalog/10795.html

Vietnam, the oil shocks, and concerns over national competitiveness have all instigated new modes of government support. Los Alamos National Laboratory, for example, a leader in supercomputing, was created by the Manhattan Project and became part of the Department of Energy. The Office of Naval Research and the National Science Foundation emerged in the wake of World War II to continue the successful contributions of wartime science. The Defense Advanced Research Projects Agency (DARPA) and the National Aeronautics and Space Administration (NASA) are products of the Cold War, created in response to the launch of Sputnik to regain the nation's technological leadership. The National Bureau of Standards, an older agency, was transformed into the National Institute of Standards and Technology in response to . . . concerns about national competitiveness. Each organization's style, mission, and importance have changed over time; yet each organization profoundly reflects the process of its development, and the overall landscape is the result of numerous layers of history.

Understanding these layers is crucial for discussing the role of the federal government in computing research. [The following sections briefly set] out a history of the federal government's programmatic involvement in computing research since 1945, distinguishing the various layers in the historical eras in which they were first formed. The objective is to identify the changing role the government has played in these different historical periods, discuss the changing political and technological environment in which federal organizations have acted, and draw attention to the multiplicity, diversity, and flexibility of public-sector programs that have stimulated and underwritten the continuing stream of U.S. research in computing and communications since World War II. In fulfilling this charge, [the following text] reviews a number of prominent federal research programs that exerted profound influence on the evolving computing industry. These programs are illustrative of the effects of federal funding on the industry at different times. Other programs, too numerous to describe here, undoubtedly played key roles in the history of the computing industry but are not considered here.

1945-1960: Era of Government Computers

(From pp. 86-87): In late 1945, just a few weeks after atomic bombs ended World War II and thrust the world into the nuclear age, digital electronic computers began to whir. The ENIAC (Electronic Numerical Integrator and Computer), built at the University of Pennsylvania and funded by the Army Ballistics Research Laboratory, was America's first such machine. The following 15 years saw electronic computing grow from a laboratory technology into a routine, useful one. Computing hard-

Copyright © National Academy of Sciences. All rights reserved.