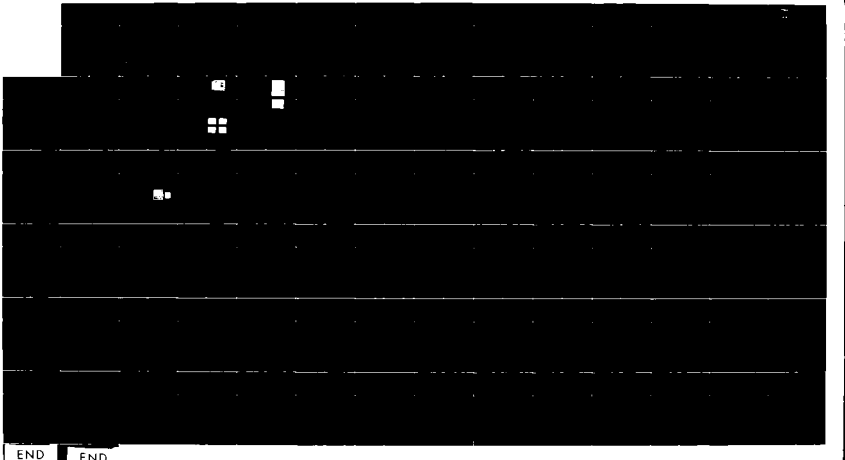
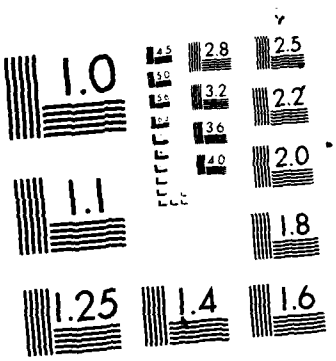
AD-A172 895 OPTICAL COMPUTING RESEARCH(U) STANFORD UNIV CA J W GOODMAN 30 JUN 86 AFOSR-TR-86-0911 AFOSR-83-0166 1/1

UNCLASSIFIED F/G 20/6 NL



END DATE FILMED 12-86 DTIC

END DATE FILMED 12-86 DTIC



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A172 895

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS:
2a. SECURITY CLASSIFICATION AUTHORITY:	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribtution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:	
4. PERFORMING ORGANIZATION REPORT NUMBER(S):	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-0911
6a. NAME OF PERFORMING ORGANIZATION: Stanford University; 6b. OFFICE SYMBOL (If applicable):	7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code): Stanford, CA 94305	7b. ADDRESS (City, State and ZIP Code): Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR; 8b. OFFICE SYMBOL (If applicable): NE	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-83-0166
8c. ADDRESS (City, State and ZIP Code): Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2305	B1	

11. TITLE (Include Security Classification): "OPTICAL COMPUTING RESEARCH"

12. PERSONAL AUTHOR(S): Dr. Joseph W. Goodman

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 18 May 85 TO 30 Jun 86	1986, June, 30	70

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report summarizes the work accomplished since May 18, 1985, on Air Force Office of Scientific Research Grant No. 83-0166. Work has been in progress in four different areas: (1) Optical interconnections; (2) Real-time defect enhancement in periodic structures using four-wave mixing; (3) Computation using neural networks; and (4) Optimal imaging concentrators. Various administrative matters pertinent to the grant are also discussed.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr. C. Lee Giles	(202) 767-4931	NE

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

OPTICAL COMPUTING RESEARCH

AFOSR-TR- 86-0911

FINAL REPORT

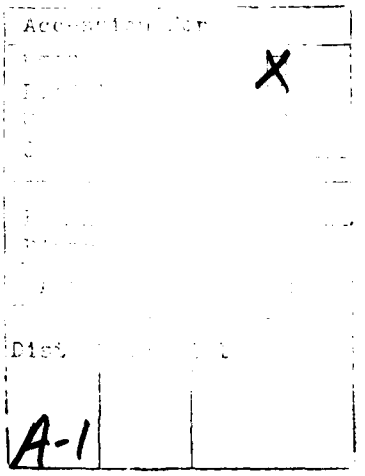
on AFOSR Contract 83-0166A-E

Submitted to the Air Force Office of Scientific Research

Joseph W. Goodman, P.I.

Stanford University

June 30, 1986



ABSTRACT

This report summarizes the work accomplished since May 18, 1985, on Air Force Office of Scientific Research Grant No. AFOSR 83-0166. Work has been in progress in four different areas: (1) Optical interconnections; (2) Real-time defect enhancement in periodic structures using four-wave mixing; (3) Computation using neural networks; and (4) Optimal imaging concentrators. Various administrative matters pertinent to the grant are also discussed.

I. INTRODUCTION

This document contains a summary of the work accomplished under Grant AFOSR 83-0166 during the time period 18 May 1985 through June 30, 1986.

Section II contains a summary of the work accomplished to date under the current year of funding. This summary is supplemented by appendices.

Section III is devoted to various administrative matters pertinent to the grant.

II. WORK ACCOMPLISHED

(a) *Optical Interconnections*

Optical interconnections has been an area of investigation under AFOSR support for several years. The powerful interconnect abilities of optical beams have led many to believe that one of the most important roles for optics in computing in the future will be as an interconnect technology.

The focus of our efforts in this area has been on the use of holographic optical elements for providing such interconnects. During the past contract year our accomplishments have been two-fold: (1) An analytic comparison of optical and electronic interconnects in the problem of chip-to-chip communication (published in *Applied Optics,* with a reprint attached as Appendix A to this report); and (2) A very detailed investigation of holographic optical elements and their capabilities in the role of interconnect elements (results presented at the 1985 Annual Meeting of the OSA, and in more detail as an invited paper at the OSA Topical Meeting on Holography, April

1986). Since the results obtained in the first of these two areas are found in the appendix, we discuss in more detail only the second area above.

Of particular interest in the interconnect problem is the diffraction efficiency that can be achieved with a holographic optical interconnect element, as well as the ability of that element to efficiently concentrate light onto a small-area photodetector. We have developed a ray-trace program that accounts not only for the density of rays in the image space (as do most conventional ray-trace programs) but also the diffraction efficiency associated with each of the rays, thus enabling us to obtain image irradiance profiles at the detector plane. The predictions of this program have been extensively verified experimentally using bleached silver halide emulsions. The approach is sufficiently general that the effects of *fan-out* on diffraction efficiency can be included, an important issue in interconnect problems. The holograms studied are generally reflection elements with focusing power. The diffraction efficiency associated with each ray is determined from coupled mode theory, using another program developed expressly for that purpose.

One Ph.D. student will be completing his degree this July in this area. A major publication on this material has been submitted recently to *Applied Optics*.

During the year we have also been devoting attention to more fundamental aspects of optical interconnections, especially the issues of fan-in and fan-out. A paper has been published by *Optica Acta* on this subject and is attached as Appendix B. In addition, an extensive survey paper on optical interconnections has been under preparation and will soon be submitted. However, much remains to be done of a fundamental nature in understanding the proper place for optical interconnects in a hierarchy of

interconnect technologies.

(b) *Defect Enhancement using Four-Wave Mixing.*

During the past contract year we have fully developed and brought to a conclusion our ideas on the use of four-wave mixing or phase conjugation as a means for enhancing defects in periodic structures. Such defect enhancement is needed in the testing of integrated circuit photomasks, as well as in other inspection problems involving periodic structures.

Our early work was devoted to the problem of intensity inversion using an inherent nonlinear property of the phase-conjugation process in photorefractives. This work was published in *Applied Optics* (Vol. 24, pp.1826-1832, 1985). Following this work, we applied the method to defect detection in periodic structures, with the results being published in *Optics Letters* (see Appendix C for a reprint).

A Ph.D. candidate finished her work on this topic in the Summer of 1985 and is now employed in industry. A patent application has been filed on the method. No further work in this area is planned, since it is ready for commercialization.

(c) *Optimal Imaging Concentrators*

During the past three years we have used a small part of our AFOSR funds to support supervision time of a U.S. Air Force Captain at Stanford in a Ph.D. program. This individual has now completed his Ph.D. thesis in the area of optimal imaging concentrators, i.e. imaging system configurations that will maximally deliver light (of an arbitrary state of partial coherence) to a prescribed detector array of arbitrary

geometrical configuration. The research is highly theoretical in nature, but has direct applications to both optical interconnections and to high-energy lasers. The early results of the work were reported at the 1985 Annual Meeting of the Optical Society of America. At about the same time a full-length technical paper was submitted to *JOSA-A* for the special issue on Coherence and Statistical Optics. We expect this paper to be published within the next month or two. The Ph.D. student working in this area will be completing his final requirements this summer.

(d) *Neural Networks and Optical Computing*

During the past contract year we have undertaken research in a new area that we feel is very exciting and promising, namely the application of neural network ideas to problems of optical computing. There is a multitude of researchers who are currently looking at such networks as a possible means for realizing associative or content-addressable memories. In view of the substantial efforts in this area elsewhere, we have chosen instead to focus on the application of such ideas to *computing*.

For six months during 1985 we were fortunate to have as a visitor with our group Prof. Mitsuo Takeda from the University of Electro-communications, in Tokyo. Under our encouragement, Dr. Takeda began an investigation in this area in collaboration with us, and results that we feel are very significant were obtained. To summarize in a few words, we investigated the application of the Hopfield neural network model to the following computational problems:

1. The "Hitchcock" problem, which is a transportation problem or a resource allocation problem.

2. Matrix inversion and image deblurring problems.

3. Signal processing problems, including spectral analysis.

The results of these investigations revealed some interesting points that require further investigation:

1. For most (but not all) problems, the most direct solution was one that mixed the "program" and the "data" in a single interconnect pattern.

2. For many (but not all) problems, the computational load associated with determination of the required interconnect pattern is comparable with the computational load associated with direct solution of the problem.

3. For most problems, constraints must be properly weighted with respect to the energy function to be minimized, requiring rather ad hoc and empirical choices.

In view of the importance we place on this work, both with respect to work accomplished and work proposed, we are attaching a preprint of the paper to this report as Appendix D (in spite of its bulk). This work has been accepted for publication in the *Applied Optics* special issue on number representations in optical computing.

III. ADMINISTRATIVE MATTERS

This section contains miscellaneous information pertinent to the grant.

Publications on work fully or partially supported by this grant and accepted or published during the last contract year are as follows:

(1) E. Ochoa, L. Hesselink, J. W. Goodman, "Real-time intensity inversion using two-wave and four-wave mixing in photorefractive Bi12GeO20", APPLIED OPTICS, Vol. 24, pp.1826-1832 (1985).

(2) R.K. Kostuk, J.W. Goodman, L. Hesselink, "Optical imaging applied to microelectronic chip-to-chip interconnections", APPLIED OPTICS, Vol. 24, No. 17, pp 2851-2858 (1985).

(3) E. Ochoa, J.W. Goodman, L. Hesselink, "Real-time enhancement of defects in a periodic mask using photorefractive B12SiO12", OPTICS LETTERS, Vol.10, pp. 430-432 (1985).

(4) J.W. Goodman, "Fan-in and Fan-out with optical interconnections", OPTICA ACTA, Vol. 32, No. 12, 1489-1496 (1985).

(5) J.W. Goodman, R.K. Kostuk, and B. Clymer, "Optical interconnects: an overview", Proceedings of the IEEE Conference on Multilevel Interconnects for VLSI, Santa Clara, California, June 1985, pp. 219-224.

(6) J.W. Goodman, "A random walk through the field of speckle", Optical Engineering, May 1986.

Papers under submission include:

(1) P. Idell and J.W. Goodman, "Design of optimal imaging concentrators for partially coherent sources:absolute encircled energy criterion", Accepted for publication in JOSA-A.

(2) M. Takeda and J.W. Goodman, "Neural networks and computing: number representations and programming complexity", Accepted for publication in APPLIED OPTICS.

Three contributed papers were presented at the 1985 Annual meeting of the OSA. An invited paper was presented at the Workshop on Optical Interconnects sponsored by MCC in Austin Texas in November 1985. An plenary paper was presented at LASER 85 in Los Vegas, Nevada in December 1985. An invited paper entitled "Holographic optical elements for optical interconnects" was presented at the OSA Topical Meeting on Holography, Honolulu, Hawaii, April 1986. An invited paper entitled "Optical interconnects" was presented at the NSF Workshop on Lightwave Technology, Tucson, AZ, May 1986. An invited paper entitled "Optical interconnections and computing" was presented at the US-Japan Workshop on Optoelectronics, Tokyo, Japan, May 1986.

APPENDIX A

Optical imaging applied to microelectronic chip-to-chip interconnections

Raymond K. Kostuk, Joseph W. Goodman, and Lambertus Hesselink

a reprint from Applied Optics
volume 24 number 17, September 1, 1985

Optical imaging applied to microelectronic chip-to-chip interconnections

Raymond K. Kostuk, Joseph W. Goodman, and Lambertus Hesselink

An imaging system is proposed as an alternative to metallized connections between integrated circuits. Power requirements for metallized interconnects and electrooptic links are compared. A holographic optical element is considered as the imaging device. Several experimental systems have been constructed which have visible LEDs as the transmitters and PIN photodiodes as the receivers. Signals are evaluated at different source–detector separations. Multiple exposure holograms are used as a means of optical fan out allowing one source to simultaneously address several receiver locations. Limitations of this technique are also discussed.

I. Introduction

A limitation of increasing importance in VLSI electronic integrated circuit design is the interconnections between devices and systems. Restrictions of conventional interconnects arise from (a) increased space allocated to wiring, (b) propagation delays with increased line lengths and RC time constants, (c) inductive noise between lines, (d) dominance of line capacitance over other sources of capacitance as line lengths increase, and (e) degrading electromigration effects on wiring materials.[1-3] Since different optical signals can propagate through the same spatial volume without interference, the possibility of using optical methods to alleviate this space restriction is attractive.[4-6] In this paper we discuss a number of aspects of optical imaging which are applicable to the electronic interconnect problem and evaluate an experimental system.

II. Comparison of Optical and Electronic Interconnections

Figure 1 shows a typical VLSI microelectronic circuit mounted and bonded to a package which can be connected to other electronic systems. There are several thousand gates on this circuit and several hundred output pins which allow communication to other systems. Two levels of interconnection can be identified: One connects two or more devices on a common chip, and another connects an integrated system or chip to another chip.

There are a number of ways to compare the performance and capability of different types of interconnection.[1,3] Consider one such criterion, the reactive power required of one electronic inverter to trigger another inverter. Reactive power is given by

$$P = \frac{CV^2}{2\tau}, \tag{1}$$

where C is the capacitance of the line and attached devices, V is the device threshold level (assumed 1 V), and τ is the clocking period (assumed 1 nsec).

Figure 2 illustrates gate-to-gate connection.[7] The gate capacitance of two devices and the metal line connecting them must be charged to the threshold potential for the gate. The gate capacitance is given by

$$C_g = \frac{\epsilon_r \epsilon_0 A}{d}, \tag{2}$$

where $\epsilon_r = 3.9$ for SiO_2, $\epsilon_0 = 8.854 \times 10^{-4}$ F/cm, A is the device area, and d is the oxide thickness layer. Projected VLSI device lengths and oxide layer thickness are 0.5 and 0.02 μm, respectively. This gives a gate capacitance of $C_g = 50$ fF/device. The capacitance of the line joining two devices is

$$C_l = \epsilon_r \epsilon_0 \frac{w}{h} l,$$

where l is the line length and w the linewidth. The width/height ratio is restricted by fringing field effects to a minimum value of ~2. For a typical VLSI circuit the average length is ~1 mm long. This gives a line capacitance of $C_l = 70$ fF. The total capacitance of this link is then $C_t = 2C_g + C_l = 170$ fF, and the corresponding reactive power $P_g = 85$ μW.

The authors are with Stanford University, Electrical Engineering Department, Stanford, California 94305.

Received 26 February 1985.
0003-6935/85/172851-08$02.00/0.

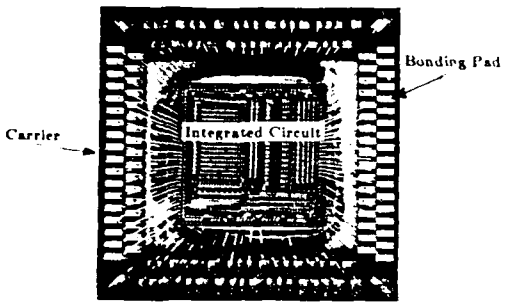


Fig. 1. VLSI circuit (manufactured by Honeywell) with ~3000 gates and 150 bonding pads. Interconnections exist between gates on a common substrate and from bonding pads to other circuits and outside systems.

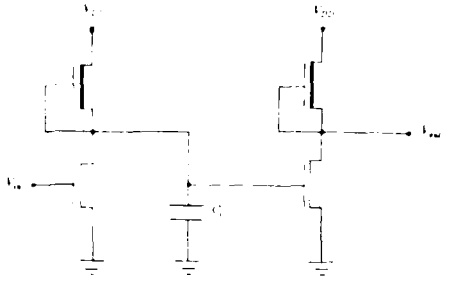


Fig. 2. Schematic of gate-to-gate connection for two inverters. The line between gates is modeled as a single capacitor.

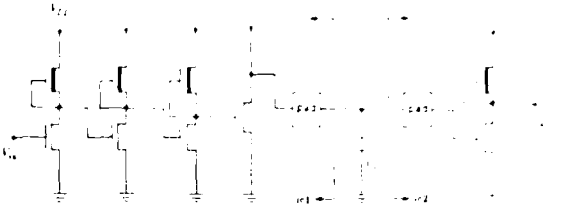


Fig. 3. Schematic of chip-to-chip connection. Inverters have [illegible] with increasing capacitance to minimize signal delay. Lines are assumed short enough so as not to be influenced by transmission line effects.

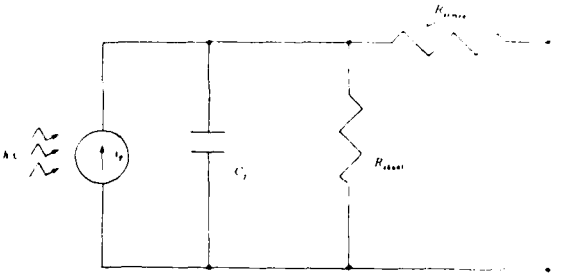


Fig. 4. Detector circuit model. The space-charge region of the junction results in a capacitance shunting a photon induced current source. The series resistance is typically a few ohms and can be neglected. The parallel resistance is of the order of $10^9\ \Omega$ and can be assumed to be an open circuit.

Figure 3 shows a chip-to-chip connection. To minimize propagation delays, gate capacitances are gradually increased in size until the device capacitance is comparable to that of a bonding pad.[7] A voltage pulse from a logic element must have sufficient power to charge these gates, two bonding pads, the line connecting them, and a receiving gate to the device threshold level. The total capacitance of this link is $C_t = 2C_b + C_l + 2C_g$, where C_b is the bonding pad capacitance. For a pad area of ~100 μm^2 and assuming a SiO_2 dielectric, this capacitance is ~0.4 pF. Lines connecting the pads are 25 μm in width and are assumed to be 500 μm above the ground plane. When a number of chips are connected on the same substrate, a typical length separating a nearby pair is of the order of 1 cm. At this distance transmission line standing wave effects are not significant (i.e., $\lambda = 30$ cm).

The line capacitance in this case is only 4.5 fF. The total capacitance becomes $C_t = 0.8$ pF + 0.0045 pF + 0.1 pF = 0.9 pF and the switching power $P_c = 430$ μW.

Next consider a simple electrooptic link consisting of a semiconductor source and detector. Initially it is assumed that all the light from the source is focused on the detector. The detector circuit model is shown in Fig. 4. The current generated is a function of the physical parameters of the junction and the illumination.[8]

$$i_p = \Phi \frac{q(1-r)}{h\nu}[1 - \exp(-\alpha_\lambda z)],$$

where i_p is the photocurrent, Φ is the optical flux, q is the electronic charge, r is the Fresnel reflection coefficient of the detector surface, $h\nu$ is photon energy, α_λ is the semiconductor absorption coefficient at λ, and z is the absorption width. Typical responsivity for a silicon device is 0.4 A/W.

The usual condition of low series and large shunt resistance simplifies the model to a capacitance shunting a current source. Current from the detector must charge the gate to its threshold level in a time less than the clocking period τ. If no preamplifier is assumed, all current must originate from electrons generated from the incident optical flux Φ. For a 2-μm thick, 25-μm square active area detector, the junction capacitance is $C_d = 32.5$ fF. Since the detector must charge the capacitance of a gate, the total capacitance is $C_t = C_d + C_g = 82.5$ fF. For a threshold voltage of ~1 V,

$$V = Q/C_t,$$

$$Q = \int_0^\tau i\,dt \simeq i\tau,$$

$$\tau = 1\ \text{nsec}.$$

With 200 μW of incident optical power, 80 μA of current can be generated in the detector and can produce 80 fC of charge. This is sufficient to produce the 1-V threshold value. Assuming a laser diode electrical to optical conversion efficiency of 30%, the electrooptic link will require ~670 μW of electrical power. (A large fraction of the power needed to drive a diode is not reactive. The important consideration here is the amount

of power required to transmit comparable optical and electrical signals. In the electrical case with FET type devices, the power is primarily reactive in nature, while the optical system also requires a real power component. The consequence of this difference is not significant at this level of analysis. It would need consideration if heat dissipation effects were investigated.)

These first-order considerations indicate that, with currently available electrooptic technology, the power required for an electrooptic link is of the same order of magnitude as that necessary for the electrical chip-to-chip interconnection and would not suffer from the problems previously outlined for conventional interconnections. The electrooptic link compares less favorably with gate-to-gate connections on the same chip.

III. Optical Chip-to-Chip Layout

The chip-to-chip interconnect problem can be formulated in more specific terms as shown in Fig. 5. One or more integrated systems are mounted on a common substrate separated by distances of ~1 cm. As mentioned previously, at these lengths and frequencies of 1 GHz, transmission line effects are not significant. Bonding pads are assumed to be 100-μm square and separated by 100 μm. Several hundred bonding pads must be connected. Each transmission point should be able to address several receiver locations; it is also desirable for channels to cross without interference.

An imaging system can provide this connection mechanism. Consider the arrangement of Fig. 6. A semiconductor emitter illuminates a holographic optical element coded to distribute radiation to one or more image points. Photodiodes convert optical to electrical signals, which are then decoded by a digital electronic circuit.

Advantages of using holographic elements include their adaptability to decentered layouts by using off-axis recording geometries and to fan out by using sequentially exposed multiple holograms.

A number of factors must be considered in a practical system of this type. The most attractive sources and detectors are those made from materials which are compatible with integrated electronics. Semiconductor sources developed for optical communications have emission wavelengths from 780 nm to 1.6 μm. To date only a few holographic recording materials are responsive at these wavelengths and these are not very sensitive.[9]

Other considerations are the emission profile and polarization characteristics of the source. Laser diodes have an emission profile corresponding to the diffraction pattern of the junction geometry. Planar stripe junction diodes have transverse mode divergence angles which have typical values of 60° by 10°. Therefore only a portion of the volume above the source will be illuminated. The hologram need only occupy this region above the source to be effective.

The polarizations of these two directions are orthogonal. Kogelnik[10] has shown that polarization vectors oriented in the plane of incidence of the grating

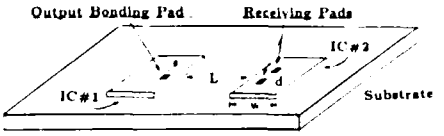


Fig. 5. Geometrical layout of a chip-to-chip connection. Two integrated circuits are mounted on a common substrate with L = 1.5 cm, w = 1 cm, and bonding pad widths and separations = 100 μm.

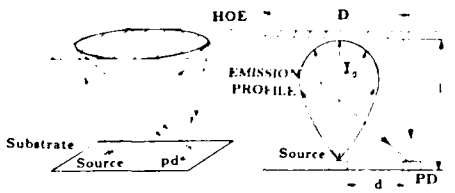


Fig. 6. Imaging system for chip-to-chip communication. Light emitting sources and detectors replace transmitting and receiving bonding pads. A hologram is used as the imaging element. Design must include f/No. or $1/D$ ratio, intensity emission profile of the source, and source-detector separation.

produce a reduced coupling constant and diffraction efficiency which results in lower image intensity.

A LED is also a potential semiconductor source. It has the advantage of being a surface emitter and is much easier to fabricate than a laser diode. In addition they can be made to emit in the visible by introducing traps in the band gap. However, they are inefficient in comparison to laser diodes and have spectral bandwidths of ~20 nm. Also they emit unpolarized light which results in lower diffraction efficiency for the reason mentioned above. Their intensity emission profile is cosinusoidal in angle and therefore illuminates a larger region of a hologram than would a laser diode. Image reconstructions with this type of emission profile are brighter when the hologram occupies large solid angles relative to the source.

IV. Holographic Optical Element Characteristics

The requirement for a compact system implies that the element must have a small f/No. This also improves flux collection. The meridional angles for $f/1$ and $f/3.5$ elements are 26.5° and 8.1° in air. A model for diffraction efficiency must be valid for grating vectors covering this angular range. A relatively simple description of grafting diffraction efficiency is Kogelnik's coupled two-wave treatment.[10] The expression of efficiency for reflection holograms with absorption is given by

$$\eta = \left[\xi/v + \left(1 + \frac{\xi^2}{v^2}\right)^{1/2} \coth(v^2 + \xi^2)^{1/2}\right]^{-1},$$

where η is the diffraction efficiency,

$$v = j\pi n_1 \frac{d}{\lambda(c_r c_s)^{1/2}},$$

$$\xi = 1/2\, D_0(1 - c_r/c_s),$$

$$D_0 = \frac{\alpha d}{\cos\theta_0},$$

n_1 is the refractive-index modulation, d is the grating thickness, c_s and c_r are obliquity factors, α is the absorption/length, and θ_0 is the Bragg angle. The planar grating treatment can be extended to curved surfaces by assuming that the surface is locally plane in the region where the ray intersects the grating.[11]

A number of planar volume phase holograms were formed in bleached photographic film. The thickness, refractive index, and postbleached absorption were measured to obtain average values for these parameters. The results were then used in Kogelnik's model to predict the diffraction efficiency curves and were compared with measured curves. Although slight changes to the values for absorption and emulsion thickness change had to be used, the agreement was very good. Figure 7 shows two measured diffraction efficiency curves from gratings with **K** orientations approximately equal to the meridional angles of *f*/1 and *f*/3.5 systems. High diffraction efficiency is maintained over a large range of playback angles. The holographic optical element (HOE) field of view is essentially this angular range and is ~30° for 25° grating slant angles and 60° for 10° slant angles corresponding to the *f*/3.5 system.

A single grating element can interconnect a number of sources and their conjugate receiver locations over the angular range of high efficiency. When source reconstruction coordinates differ significantly from formation positions, hologram image aberrations reduce image irradiance. Aberrations can be evaluated with ray tracing techniques. For thick holograms these expressions may be derived from the reflected ray components which are perpendicular and tangent to the grating vector: $\mathbf{r} = (\mathbf{K}\cdot\mathbf{r})\mathbf{K} - \mathbf{K} \times (\mathbf{K} \times \mathbf{r})$, where **r** is a unit vector along the reconstruction ray, and **K** is the grating vector given by $\mathbf{K} = \mathbf{r}_o - \mathbf{r}_c$, and $\mathbf{r}_o$ and $\mathbf{r}_c$ are unit vectors along the object and reference ray directions, respectively. The reconstructed or image ray is $\mathbf{r}' = -(\mathbf{K}\cdot\mathbf{r})\mathbf{K} - \mathbf{K} \times (\mathbf{K} \times \mathbf{r})$.

The spot diagram generated by ray tracing should be adjusted for the variation in efficiency at different locations in the aperture of the volume HOE. However it has been shown that a close relationship exists between the observed image field and the density of rays traced through the element.[12] Figure 8 shows the spot diagram of rays from a source point displaced 0.5 cm perpendicular to the axis and 0.1 cm along the axis from the source formation positions for *f*/1 and *f*/3.5 elements. It is clear that off-axis imaging degrades much more rapidly for smaller *f*/No. HOEs.

A computer program coding the grating equation can be used to generate a spot diagram at any desired image plane. When used in conjunction with Kogelnik's efficiency model, both the aberrations and the efficiency of the rays forming the image can be determined. This gives a better indication of the distribution of flux at the receiver location and the detector current produced from a source of given size, output power, and location relative to the HOE. Such a program is currently under development in our lab for use with multiple image reflection hologram design.

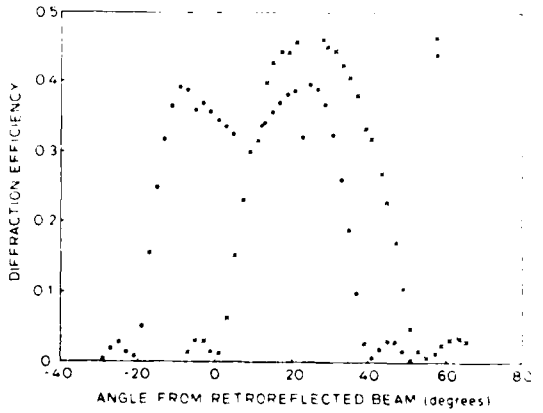


Fig. 7. Measured diffraction efficiency curves for gratings with **K** approximating those formed by the meridional rays in (O) an *f*/1 system, i.e., 25°, and (×) an *f*/3.5 system, i.e., 10°. Significant efficiency exists over an angular range of 30–60°.

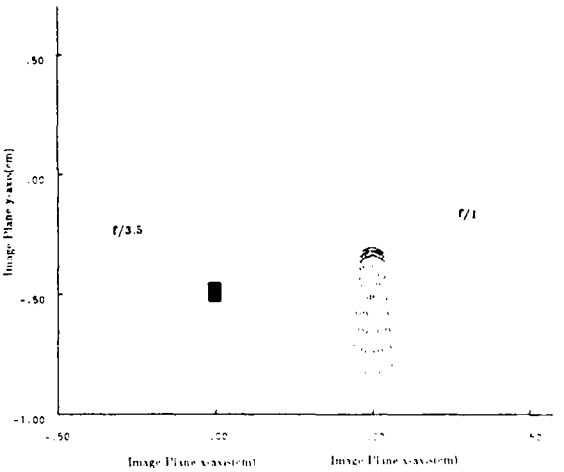


Fig. 8. Spot diagrams for *f*/1, and *f*/3.5 systems with a reconstruction source point 0.5 cm from the axis of the element at $x = 0$, $y = 0$. Computations are based on the grating vector equation.

The effective HOE aperture and reflection losses also restrict the usable source power. The solid angle subtended by the HOE relative to a source point is

$$\Omega = \left(\pi \frac{D^2}{4}\right) \frac{\cos\theta}{r^2},$$

where D is the diameter of the hologram aperture, θ is the angle from the source point to the center of the hologram, and r is the distance from the source point to the hologram center.

If the source is a Lambertian emitter, the flux collected by the aperture of the HOE is $\Phi = (I_0 \cos\theta)\Omega$. When the source and optical element are on-axis, 12.5% of the available source power is collected with an *f*/1 system and only ~1.0% for an *f*/3.5 system.

If the hologram recording medium is not index matched to the source and detector surfaces, Fresnel

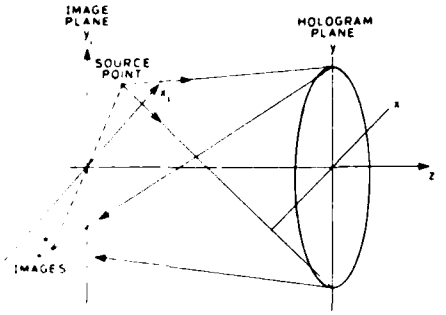


Fig. 9. Simplest holographic configuration for imaging interconnects. Light from a point source is imaged to a diametrically opposite point. Several sequential exposures can be encoded and used to produce an invariant pattern of images. This can be used for invariant fan-out configurations.

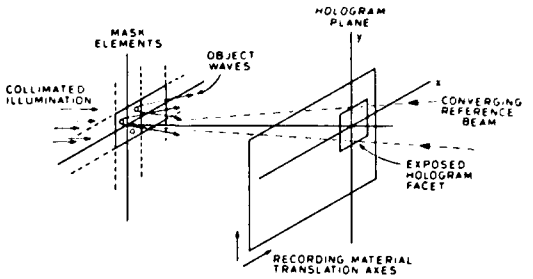


Fig. 10. Combined multifacet hologram and variable image mask. A separate hologram facet is formed with each fan-out pattern encoded on the mask. The mask and hologram are translated with respect to each other. Each hologram is formed with a converging reference wave to allow playback with an expanding beam.

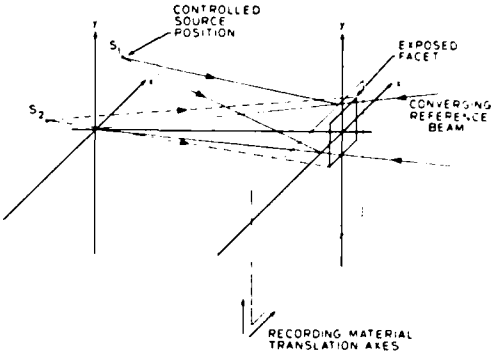


Fig. 11. Multifacet hologram formed with selective object source points. Source points are encoded in sequential fashion. This is the most flexible configuration but also the most difficult to implement.

reflection losses also reduce the flux entering the grating. The recording medium used has a refractive index of 1.64, resulting in transmitted intensities ranging from 91 to 93% for incident angles of 0–30°. Therefore 7–9% of the available source power is lost by reflection. If a fixed amount of flux Φ_{dt} is required at the detector, axially located sources must have output powers Φ_{dt} exceeding this value by $\Phi_{st} = \mu\Phi_{dt}$, with $\mu = 1/(0.125 * 0.08) = 9$ for an $f/1$ system, and $\mu = 1/(0.01 * 0.08) = 109$ for an $f/3.5$ system. Therefore considerable power is required from a Lambertian source even when a 100% efficient hologram is used.

The divergence angle from a laser diode is approximately matched to the meridional angle of an $f/1$ element (~30° for the laser and 26° for the optical element. This implies that all the power from a laser diode can be collected by a smaller aperture than for a Lambertian source. A laser diode can therefore have much lower input power and still produce the required detector current and perform the switching task.

After considering the above HOE and semiconductor source characteristics, three types of hologram configuration appear to offer a solution. The first arrangement is a large aperture reflecting lens with one or multiple gratings (Fig. 9). This element is relatively easy to fabricate and position and uses a point source for reconstruction. Multiple grating formation allows a single reconstruction source to address several locations simultaneously. It does however restrict the locations of sources and detectors to positions along diameters which pass through the optical axis, and fan out can only be accomplished in an invariant pattern. This restriction may preclude this arrangement from practical application but it is important for optical system evaluation. The second and third configurations utilize the multifacet or aperture partitioning concept recently discussed by Haugen *et al.*[13] for transmission holograms and requires directed beam reconstruction either from a laser diode or a directed LED emission pattern. In one of these arrangements a mask with the address pattern serves as the object wave and a converging beam as the reference wave (Fig. 10). This method has the attractive aspect of having an IC compatible technique (i.e., mask making) used for generating an address pattern. The drawbacks of this arrangement are the intermodulation terms which limit the efficiency of the reconstruction images.[14] It is not obvious where this becomes restrictive for this application. In the last hologram configuration proposed each facet is illuminated sequentially with a number of diverging object beams and a converging reference wave (Fig. 11). The positions of the object beams can be moved automatically with a computer-controlled stepper motor drive and beam ratios can be adjusted for maximum diffraction efficiency. This configuration appears to offer the most flexible arrangement for fabricating an interconnect pattern since it satisfies requirements for both a large number of independent channels and spatially variant fan out. The difficulty with this HOE fabrication technique is the mechanical complexity of the mount; however there appears to be no fundamental restriction to its implementation.

V. Experimental Results

To evaluate some of the above ideas a number of experimental systems were fabricated and tested. Only the first hologram design described above is discussed here. The other two hologram types will be presented in future papers.

Table I: Source and Detector Characteristics

Sources	Litronix	Hewlett-Packard (HP)
Input power	155 mW	140 mW
Output power	70 μW/str	80 μW/str
Intensity profile	Lambertian	Lambertian
Size	250 μm^2	150 μm^2
λ_{peak}	660 nm	635 nm
$\Delta\lambda$	20 nm	20 nm
Detectors	**HP EO Coupler PD**	**HP4205 PIN**
Size	400 μm^2	200 μm (diameter)
Responsivity (630 nm)	0.1 A/W	0.4 A/W

The effects of image degradation and power loss were determined by mounting a number of sources and detectors at increasing separations and measuring the received detector photocurrent. The operating characteristics of the sources and detectors are given in Table I. The sources are surface emitting GaP LEDs. The primary reason for using these devices is their peak emission in the visible (635 and 655 nm) making them compatible with a number of available holographic recording materials. They have about a 20-nm spectral bandwidth and a near Lambertian intensity emission profile. Their main disadvantage is their poor electrical to optical conversion efficiency. Measured efficiency of both the 635- and 655-nm LEDs is ~0.5%. Sources and detectors used were in chip form with cross-sectional dimensions of the same order of magnitude as the size of the bonding pads (see Fig. 12).

Two source–detector mounts were used. On the first, the devices were set on the common conducting plane of a dual in-line IC package. This arrangement allowed evalution of both electrical coupling and direct optical scattering on the detector signal received from the source image. The second mount had source and detector on different substrates and was optically isolated to allow examination of the effects of image degradation and aperturing at large source–detector separations.

Figure 13 is a plot of the ratio of photodiode current with the image of the source focused onto the detector to the current with the image focused just off the detector. Response with source–detector separations from 86 μm to 4 mm was obtained with the source and detector mounted on the same conducting substrate. It appears that optical scattering and electrical coupling greatly reduce the effective signal response at separations <100 μm. At separations from 2 to 4 mm, contrast ratios increase more slowly than at closer separations. With source and detectors on separate substrates and isolation from optical scatter, the contrast ratio improves by an order of magnitude at 1.0-cm distances, then falls by a factor of 2 as separation increases to 2 cm. The falloff at larger separations results from aberrations which reduce image irradiance.

The image of the source was also observed on a CCD line scanner to directly evaluate the image irradiance pattern. Figure 14 shows these profiles when the 635-nm LED is 0.45, 0.60, 1.00, and 1.50 cm from the linear scanner. The hologram used for these mea-

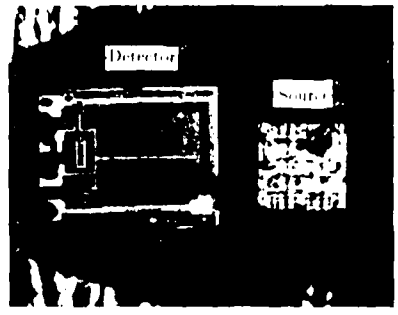


Fig. 12. Litronix LED with 250-μm^2 emission area and a Hewlett-Packard photodiode from an electrooptic coupler with 400 μm^2 active area. The separation of the two chips is ~60 μm.

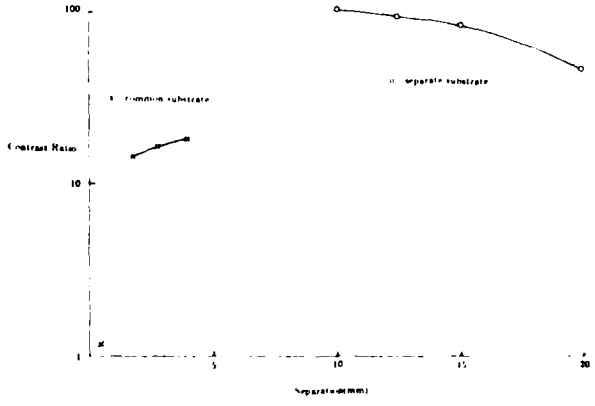


Fig. 13. Plot of the ratio of photodiode current with image focused on the detector to the current with the image focused off the detector. The equipment used did not allow measurements with source–detector separations from 4 to 10 mm; (×) indicates measurements obtained with sources and detectors on the same substrate; (○) on separate substrates.

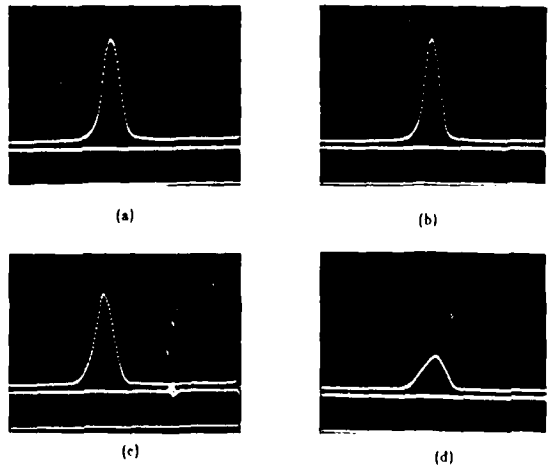


Fig. 14. CCD line scan traces of images of the 635-nm LED produced with the f/1.9 HOE. The CCD has 256, 13-μm elements. Oscilloscope scale is 330 μm per 1 cm. Source–CCD separations are (a) 0.45 cm; (b) 0.60 cm; (c) 1.00 cm; and (d) 1.50 cm.

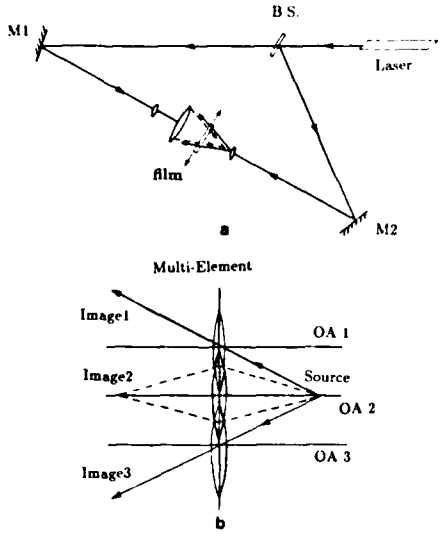


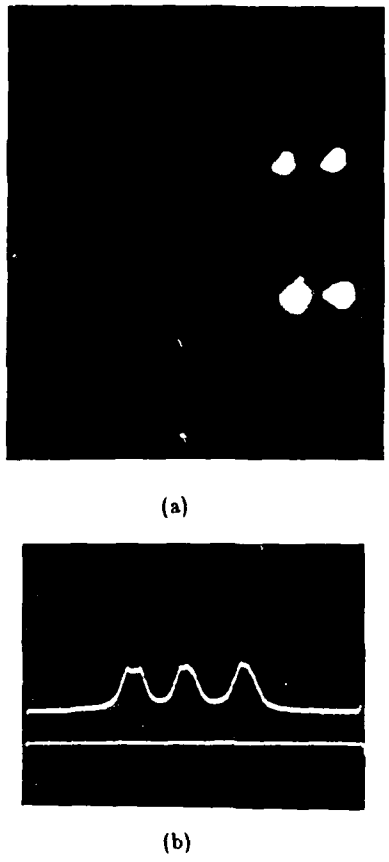
Fig. 15. (a) Schematic of hologram construction arrangement to form a multiple image with a single reconstruction source. The film plane is translated through fixed construction beams. (b) The resulting element is in effect a set of reflecting lenses with displaced optical axes which image the source relative to their respective axes. The lenses in (b) are shown unfolded for clarity.

surements has a diameter of 1.5 cm and is 3.10 cm from the source-detector plane. The CCD scans indicate that the falloff in effective signal response at large separations results from an increase in the image area and a corresponding decrease in image irradiance illuminating the detector.

A number of multiple exposure holograms were made to examine the potential of optical fan out. Elements were made with the arrangement shown in Fig. 15. Converging and diverging wave fronts overlap to form an on-axis reflecting lens type hologram. The film plane is then translated in this overlap region to form a number of holographic lenses with their optical axes displaced by the amount of translation. A single reconstruction source has a different displacement from the optical axis of each encoded element and therefore images the source at a different position in space. Figures 16(a) and (b) show images produced from two such elements. In the first, film translations of 0.7 by 0.25 cm were used, while in the second 0.5-mm movements were made. Both situations give well-resolved images with full width at half-intensity maxima (FWHM) of $\sim$300 μm. The LED emission surface is 150 μm in length.

VI. Conclusions and Future Research

Reactive power considerations indicate that with current electrooptic technology an optical chip-to-chip interconnect requires approximately the same amount of power to transmit high speed signals as electrical connections but without the need to devote large sections of the circuit substrate to communication channels. This would allow the use of more input output ports and increase the information capacity of the IC.



(a)

(b)

Fig. 16. (a) Photograph of multiple images formed with an element having 0.25-cm horizontal and 0.70-cm vertical displacements using a LED reconstruction source. The diode is 1 cm from the center of the image pattern. (b) Photograph of a CCD line trace of the LED imaged by a HOE with three 500-μm translations. Scale is 330 μm per 1 cm.

It could also reduce electrical coupling difficulties of conventional interconnect schemes. The chip-to-chip interconnect can be recast in terms of an optical imaging system with semiconductor sources as signal transmitters and photodiode detectors as receivers.

The diffraction efficiency characteristics of reflection volume holograms have sufficient angular response to accommodate source-detector separations of a few centimeters. These separations also require that the holographic element be located a comparable distance above the circuit substrate. Other practical considerations are Fresnel reflection losses and flux collection characteristics of a particular f/No. element and source emission profile. Serious limitations also exist in the lack of compatibility between efficient semiconductor sources and holographic recording materials. A match between these components would allow use of much more efficient sources and greatly improved flux collection geometries.

Initial experiments indicate that electrical and optical coupling are serious problems when sources and de-

tectors are <100 μm apart and the image blurring causes the falloff in detector irradiance at separations of a few centimeters and greater.

Experiments also indicate that sequentially exposed holograms have sufficient resolution to address a number of receivers spaced from several hundred micrometers to centimeters. This could be used to implement a number of very flexible interconnect patterns without the drawbacks of conventional electrical systems.

This work was supported by the Air Force Office of Scientific Research. One of us (RKK) would especially like to thank IBM for fellowship support during this period.

References

1. R. W. Keyes, "Fundamental Limits in Digital Information Processing," Proc. IEEE **69,** 267 (1981).
2. A. J. Rainal, "Computing Inductive Noise of Chip Package," AT&T Bell Lab. Tech. J. **63,** 177 (1984).
3. P. M. Solomon, Proc. IEEE **70,** (1982).
4. J. W. Goodman, F. J. Leonberger, S. Y. Kung, and R. A. Athale, "Optical Interconnections for VLSI Systems," Proc. IEEE **72,** 850 (1984).
5. B. K. Jenkins and T. C. Strand, "Computer-Generated Holograms for Space-Variant Interconnections of Optical Logic Systems," Proc. Soc. Photo-Opt. Instrum. Eng. **437,** 110 (1983).
6. A. A. Sawchuk and T. C. Strand, "Digital Optical Computing," Proc. IEEE **72,** 758 (1984).
7. C. Mead and L. Conway, *Introdution to VLSI Systems* (Addison-Wesley, Menlo Park, Calif., 1980).
8. S. E. Miller and A. G. Chynoweth, *Optical Fiber Telecommunications* (Academic, New York, 1979).
9. H. M. Smith, Ed., *Holographic Recording Materials* (Springer, Belin, 1977).
10. H. Kogelnik, "Coupled Wave Theory for Thick Hologram Gratings," Bell Syst. Tech. J. **48,** 2909 (1969).
11. R. R. A. Syms and L. Solymar, "Localized One-Dimensional Theory for Volume Holograms," Opt. Quantum Electron. **13,** 415 (1981).
12. R. Ferrante, M. P. Owen, and L. Solymar, "Conjugate Diffraction Order in a Volume Holographic Off-Axis Lens," J. Opt. Soc. Am. **71,** 1385 (1981).
13. P. R. Haugen, H. Bartelt, and and S. K. Case, "Image Formation by Multifacet Holograms," Appl. Opt. **22,** 2822 (1983).
14. J. Upatnieks and C. Leonard, "Efficiency and Contrast of Holograms," J. Opt. Soc. Am. **60,** 297 (1970).

APPENDIX B

OPTICA ACTA, 1985, VOL. 32, NO. 12, 1489–1496

Fan-in and fan-out with optical interconnections

JOSEPH W. GOODMAN

Department of Electrical Engineering, Stanford University, Stanford, California 94305, U.S.A.

(*Received 18 June 1985*)

Abstract. Optical beams are known to have many desirable properties when used for providing interconnections. Such interconnections would be used in an all-optical computer based on optical gates, but can be used at various levels of architecture in electronic computing systems. The fan-out of optical interconnections from one computing element to N computing elements is accompanied by an N-fold loss of light power for each connection. Less obvious is the fact that fan-in of connections from N computing elements to a single computing element can in some cases also be accompanied by an N-fold loss of power.

1. Introduction

Much attention is now being given to the possible use of optics as a means for providing interconnections in computing structures of various kinds at various levels of architecture [1–3]. The main attraction of optics in this regard is the freedom from interference between adjacent channels of interconnections, arising fundamentally from the fact that most propagation media are linear at the light levels that would be used for such signals. Interconnection paths formed by flows of electrons have a strong tendency to interact, due to the fact that such flows are composed of moving charges.

Optical interconnections can obviously be utilized in an all-optical computer, for which the basic logic operations are performed by optical logic elements, perhaps based on optical bistability. However, they can also play a more immediate role in hybrid opto-electronic computers, in which the tendency of electrons to interact is exploited to produce nonlinear interactions of signals in electronic logic gates, while optics is used to provide interconnections at some levels of architecture. Applications of optics for interconnections at high levels of architecture (machine-to-machine or processor-to-processor) are currently most easy to realize, while optical interconnections at the lowest levels of architecture (e.g. gate-to-gate connections) are most difficult to realize.

In this paper we examine some fundamental properties of optical interconnections related to their fan-in and fan-out properties. The term *fan-out* refers to the splitting of a single node or interconnection into several interconnections, each carrying the same signal. The term *fan-in* refers to the coming together of several interconnections into a single interconnection or node, all of the component signals being added to form a single signal. The two cases are illustrated in figure 1. We will show that optical and electronic interconnections share some properties but also differ in some fundamental ways. In particular, we shall see that electronic and optical interconnections are quite similar with respect to their fan-out properties, but can differ markedly in their fan-in properties.

(a) (b)

Figure 1. Representation of (*a*) fan-out and (*b*) fan-in.

2. Fan-out

It is very common in the construction of complex logic circuits that the output of a single gate must be sent to the inputs of several gates that follow. A simple example is shown in figure 2 [1] in which the output of one inverter drives the inputs of several inverters in parallel. In order to activate the parallel set of inverters, it is necessary that the current supplied by the first inverter charge the input capacitances of the following inverters to the point where the voltages across those capacitances all exceed the logic threshold voltage. Other examples can be found at higher levels of computer architecture. For example, in the construction of a crossbar switch for interconnecting several processors and memory modules (see figure 3), fan-out must be present if the switch is to offer broadcast capability, i.e. the capability of a single module to broadcast a common message to several other modules simultaneously. Again a single output must charge the inputs of a parallel array of capacitances.

It is tempting to believe that optical interconnections offer a distinct advantage *vis-à-vis* electrical interconnections when substantial fan-out is present. However, as we now argue, this is generally not the case. An optical interconnection (figure 4 (*a*)) is established by driving an optical source (a laser diode or an LED) with an electrical current. The optical source converts the flow of electrons into a flow of photons, subject to certain limitations on the efficiency of that conversion. A portion of this flow of photons is incident on a photodetector at the far end of the interconnection.

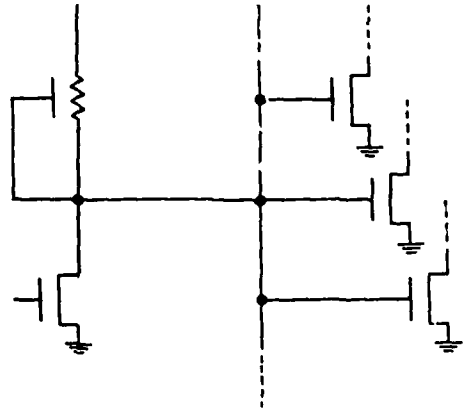


Figure 2. Fan-out of connections from one inverter to other inverters.

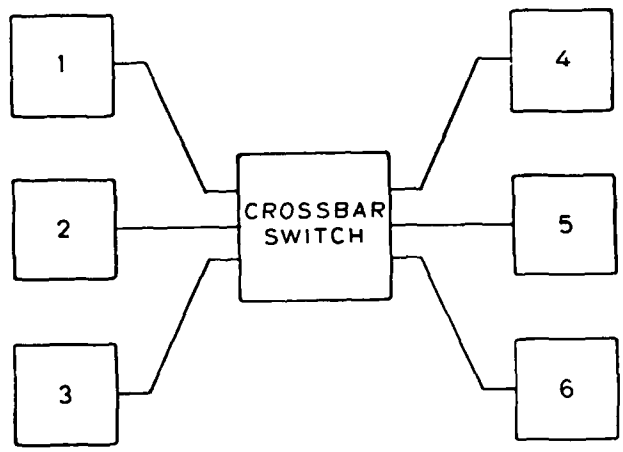
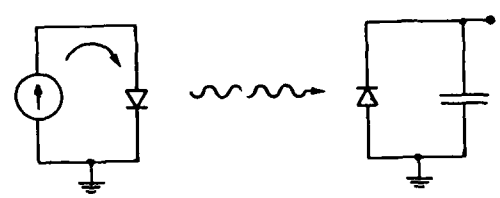
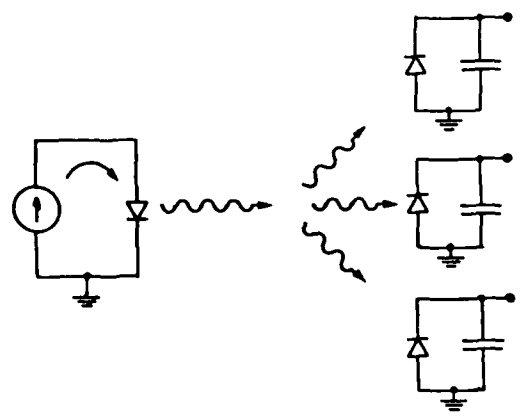


Figure 3. A crossbar switch. Modules 1, 2 and 3 can be connected to modules 4, 5 and 6 in an arbitrary fashion.



(*a*)



(*b*)

Figure 4. Optical interconnections (*a*) for a one-to-one interconnect, and (*b*) with N-fold fan-out.

The photodetector converts the flow of photons into a flow of electrons, again subject to certain limits on conversion efficiency. Finally, the flow of electrons must charge the input capacitance of a gate to its logic threshold voltage. Just as the flow of electrons must be divided N ways if an electronic interconnection with N-fold fan-out is to be established, so too the flow of photons must be divided N ways if an optical interconnection with N-fold fan-out is desired (figure 4(*b*)). In both the electrical and the optical cases, fan-out by a factor N will result in an N-fold increase in the time required to charge the N capacitors at the ends of the N interconnections, unless the rates of electron and photon flows are increased by a factor-N to compensate. Thus an optical interconnection in effect suffers from the same capacitive-loading effects that an electronic interconnection experiences, contrary to what might have been expected at the start.

There is one respect (in addition to the immunity of optical interconnections to interference mentioned earlier) in which optical interconnections do offer a potential advantage. If the length of a metallized electronic interconnection is substantial, then the capacitance of the interconnection itself may become comparable to or even greater than the capacitance of the gate at the far end. The increased capacitance will result in slower charging times and lower transmission speeds for the interconnection. An efficient optical interconnection does not possess any characteristics similar to the capacitance of the interconnection line itself. Therefore when long interconnections are required, optics may have a distinct advantage. However, a recent examination of the chip-to-chip interconnection problem [2], for which interconnection lengths of only a few centimetres were assumed, showed that the capacitances of the metallic interconnection lines were small compared with the capacitances of the bonding pads, indicating that this potential advantage of optics may not be important for short-distance communication between chips.

One important difference between optical and electronic interconnections becomes evident when further optical consequences of fan-out are fully considered. Such consideration requires the use of the principle of conservation of generalized *étendue* [3, 4], often referred to as the *constant radiance theorem* [5]. According to this theorem, the product of the cross-sectional area and the square of the numerical aperture of an optical beam must remain constant under any lossless linear transformation of that beam. Thus the fan-out of a single optical beam of cross-sectional area A into N beams, each of cross-sectional area A, must be accompanied by a reduction of the numerical apertures of the new beams by a factor $\sqrt{N}$. Such will be the case whether the optical interconnections propagate in free space or in multimode waveguides and fibres. This theorem, which is derived using the principles of geometrical optics, does not hold in the case of single-mode guides, for which geometrical optics is not valid. The fact that fan-out of optical beams changes the beam divergence has no obvious analogue in the case of electronic interconnections. The implications of the constant radiance theorem in the case of fan-out, while important, are overshadowed by those for the case of fan-in, to which we now turn.

3. Fan-in

Just as fan-out of multiple connections from a single logic gate is common, so too fan-in of multiple connections to a single logic gate is often required. Fan-in is also required at higher levels of architecture. For example, some forms of crossbar switch are constructed in such a way that all input lines can simultaneously address a single

output line. Therefore, it is important to consider the consequences of fan-in for both electronic and optical interconnections.

Figure 5 illustrates a generic kind of fan-in connection. When the connections are electrical, all sources of current must be capable of charging the input capacitance of the final node to the logic threshold voltage. However, if the output impedances of the devices driving the lines are finite, a portion of the current generated by one source will flow back through all other lines, causing the rate of charging of the desired capacitance to be slower than would be the case with no fan-in. The degree to which the speed of the circuit is limited depends on the output impedances of the sources and on the number of such lines being fanned-in to a common point.

It might appear at first glance that optical interconnections do not suffer from fan-in limitations of the above kind. Indeed in some cases they do not, but in other cases there is a very important limitation associated with optical fan-in, which, while different in origin than the effect encountered with electrical interconnections, none the less has similar or even worse consequences. The optical effect can again be viewed as a consequence of the constant radiance theorem, and its seriousness depends on the relationship between the cross-sectional areas and the numerical apertures of the beams that are being fanned in, and the same parameters of resultant beam after fan-in. If the fan-in of N identical optical beams is onto a detector with N times the cross-sectional area of the individual beams, or with an acceptance numerical aperture that is $\sqrt{N}$ times the numerical aperture of one of the incident beams, then there need be no penalties associated with fan-in (aside from the fact that a large optical detector generally has a high capacitance and a correspondingly slow speed). On the other hand, if the fan-in requires that N identical and mutually incoherent beams be combined to form a single beam with the same cross-sectional area and the same numerical aperture as those of any one of the incident beams, then the constant-radiance theorem implies that *the optical power delivered into the resultant beam cannot exceed* $1/N$*th of the total incident optical power carried by all the interconnections.*

The above conclusion can have profound effects on the design of optical interconnections. For example, one possible way to attempt an N-fold fan-in of optical beams is by means of a holographic optical element used as a beam combiner.

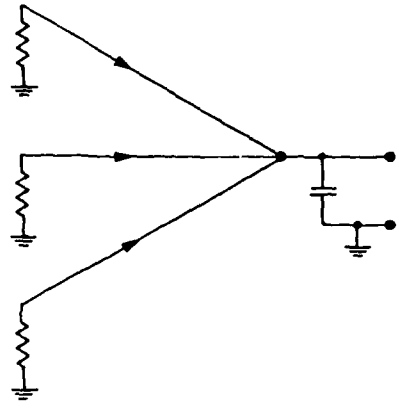
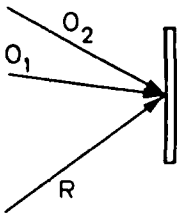


Figure 5. Fan-in of electrical connections.

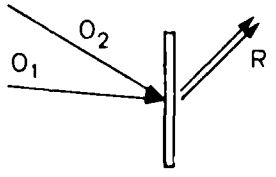
The hologram is recorded by sequentially or simultaneously recording the interference patterns between a single reference plane wave and N object plane waves travelling at different angles with respect to the reference wave (figure 6 (*a*)). When N beams are incident on the resulting hologram at angles duplicating those of the original object beams, there will be generated a beam propagating in the direction of the original reference wave, carrying contributions from all of the incident beams (figure 6 (*b*)). It has been assumed for simplicitly that the wavelength of the light exposing the hologram is identical to that of the light incident during the beam-combining operation. The cross-sectional area and the divergence angle of the combined beam should be identical with those of the beams incident on the combiner. However, the constant-radiance theorem implies that, on the average, the new beam can contain no more than $1/N$th of the power from each of the incident beams, the average being over all possible relative phases of the incident beams. The light not carried by the combined beam can be shown to appear in other orders of transmitted light.

An intuitive argument confirming the prediction of the constant radiance theorem can be reached by considering the same holographic element illuminated by a backwards travelling version of the original reference wave. The hologram can send at most $1/N$th of the incident light into each of the back-propagating versions of the object waves. Thus any single grating in the hologram can be at best $100/N$ per cent efficient, and in general will be even less efficient.

It has been implicitly assumed in the above arguments that the beams to be combined have random phases with respect to one another. Such will be the case if the beams to be combined originate from different optical sources. It will also be the case when all beams originate from the same source unless the entire optical



(*a*)



(*b*)

Figure 6. A holographic beam combiner: (*a*) construction and (*b*) utilization

interconnection system is stabilized to maintain absolutely constant paths in all arms. Such a *stabilization* seems unlikely in practice. For any one realization of relative phases between beams, the amount of optical power transferred to the resultant beam would be influenced by interference between the various contributing beams, and could be greater than or less than the power predicted by the constant-radiance theorem.

Similar conclusions also apply if the beams are travelling in single-mode waveguides. It is known [6] that the amount of power coupled into a single monomode waveguide from a Y-junction of two identical monomode waveguides (figure 7) carrying identical optical powers may be as great as twice the power carried by one of the input guides, or may be as small as zero, depending on the relative phases of the two incident beams. If the phase difference between the two beams varies randomly and uniformly over 2π radians, then the predictions of the constant-radiance theorem are obtained, namely on the average, one half of the incident power will be transferred to the outgoing guide.

4. Concluding remarks

The assumption that optical interconnections are superior to electronic interconnections from the viewpoint of fan-out and fan-in is in general unwarranted. The fan-out properties of optical beams are essentially the same as those of electrical connections. The fan-in properties of optical beams are somewhat more complex than those of electronic interconnections. If N identical incoherent beams are to fan-in to a single beam with the same cross-sectional area and the same angular divergence as the input beams, then there must be a significant and fundamental loss of power associated with the fan-in operation. If the N beams are mutually coherent, then the amount of power transferred to the resultant beam depends on the relative phases of the component beams, but averaged over all possible relative phases, results identical to those of the incoherent case will be obtained.

There are still good reasons to be interested in optics for interconnections, principally the relative immunity of optical beams to mutual interference effects. A second important reason for interest rests on the potential for constructing dynamic optical interconnection networks, which would allow rapid reconfiguration of interconnections and thereby offer a new degree of freedom for computer design.

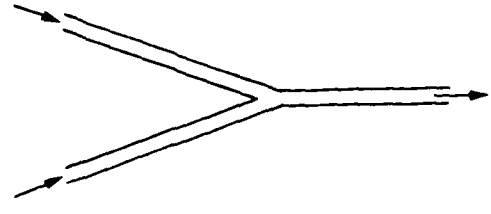


Figure 7. A monomode waveguide Y-junction.

References

[1] MEAD, C., and CONWAY, L., 1980, *Introduction to VLSI Systems* (Reading, Massachusetts: Addison-Wesley), p. 11.
[2] KOSTUK, R. K., GOODMAN, J. W., and HESSELINK, L., 1985, *Appl. Optics*, **24**, 2851–2858.
[3] MARCUSE, D., 1972, *Light Transmission Optics* (Princeton, New Jersey: Van Nostrand-Reinhold).
[4] WELFORD, W. T., and WINSTON, R., 1978, *The Optics of Nonimaging Concentrators* (New York: Academic Press, Inc) (see Appendix A).
[5] BOYD, R. W., 1983, *Radiometry and the Detection of Optical Radiation* (New York: John Wiley & Sons) (see § 5.2).
[6] REDIKER, R. H., and LEONBERGER, F. J., 1982, *IEEE Jl. quant. Electron.*, **18**, 1813–1816.

APPENDIX C

Real-time enhancement of defects in a periodic mask using photorefractive $Bi_{12}SiO_{20}$

Ellen Ochoa, Joseph W. Goodman,
and Lambertus Hesselink

a reprint from Optics Letters
volume 10, number 9, September, 1985

Real-time enhancement of defects in a periodic mask using photorefractive $Bi_{12}SiO_{20}$

Ellen Ochoa, Joseph W. Goodman, and Lambertus Hesselink

Department of Electrical Engineering, Stanford University, Information Systems Laboratory, Stanford, California 94305

Received May 2, 1985; accepted June 26, 1985

The first known experimental results of real-time optical defect enhancement of a periodic mask are reported. A low-intensity reference wave interferes with the Fourier transform of an object beam to form a hologram in a photorefractive crystal. The nonlinear properties of the crystal perform a filtering operation, and phase-conjugate readout results in a defect-enhanced image. Defects of size 10 μm × 100 μm have been easily detected with high signal-to-noise ratio, and a discussion of performance limitations is presented.

We consider the problem of selectively enhancing defects in a mask that consists of mostly periodic structure. This type of problem in image processing occurs, for example, in the inspection of integrated-circuit masks. Digital techniques for inspection of a two-dimensional field, generally utilizing a dual-scanning microscope system and sophisticated algorithms for comparison and detection, are complicated and time consuming.[1-6] Optical systems, however, offer the advantage of parallel processing. Furthermore, there is no excessive requirement for accuracy in the output in terms of the actual intensity at each point. It is sufficient that the signal associated with the defect be much larger than the signal associated with the surrounding periodic structure, so that, for example, a thresholding operation can be used to determine the defect location.

Optical spatial-filtering techniques to perform defect enhancement have been examined in the past with regard to such applications as inspection of the electron-beam collimating grid and the silicon-diode-array target for a television camera tube as well as for inspection of photomasks used in the manufacture of integrated circuits.[7-9] These systems used a filter in the Fourier plane to attenuate the discrete spatial frequencies of the periodic portion of the mask, so that, on retransformation, only defects were present in the output. Although the results of such systems were promising, the usefulness of the technique was limited by the fabrication time or difficulty of the filter and by the need to use high-quality, low-f-number lenses when inspecting objects of large dimensions. Recently the second constraint was removed by employing holographic recording of the output combined with phase-conjugate readout.[10] Although this method has been used to detect submicrometer defects, it requires two processing steps: for each mask to be inspected, a new hologram must be recorded, and for each different type of mask, a new photographic filter must be made.

We present a method to enhance defects in real time, using a photorefractive crystal. Use of the crystal allows holographic recording, filtering, and phase-conjugate readout processes to be performed simultaneously. The mask to be inspected is placed in the input plane, and the defect-enhanced image appears at the output plane, in a time limited only by the time constant of the photorefractive material. This time constant, which depends on the material used and the incident light intensity, ranged from about 50 to about 250 msec for our experimental parameters. This method also differs from that described above in that all operations are carried out in the Fourier domain. To our knowledge, this work is the first demonstration of a real-time system for enhancing defects in a periodic mask.

The technique for performing real-time defect enhancement is based on two observations. The first is that the Fourier transform of a periodic object is an array of discrete spikes whose width depends inversely on the input field size and whose spacing depends inversely on the period of the mask. In contrast, the Fourier transform of a small defect is a continuous function that is several orders of magnitude less intense than the periodic spikes. The second observation is that the diffraction efficiency of a volume phase hologram formed in a photorefractive medium is maximized when the intensities of the two writing beams are approximately equal and decreases as the difference in intensity increases. For a reference plane-wave intensity (I_r) more intense than the object-beam intensity (I_o), the output is proportional to the object-beam intensity; for an object beam more intense than the reference beam, the output is proportional to the intensity inverse of the object beam. A typical diffraction-efficiency versus beam-ratio curve is plotted in Fig. 1 on a log–log scale, assuming that beam ratio $R = I_r/I_o$ is varied by changing I_o while keeping I_r fixed.[11] This curve was generated by using the standard Kogelnik expression for diffraction efficiency[12]

$$\eta = \exp[-(\alpha d/\cos\theta)]\sin^2\left(\frac{\pi\Delta n d}{\lambda\cos\theta}\right)$$

and substituting in parameters appropriate for a $Bi_{12}SiO_{20}$ (BSO) crystal of thickness $d = 1$ cm under conditions of illumination of $\lambda = 514$ nm and an applied field of 5 kV/cm. The details of the determination of the refractive index modulation Δn are discussed elsewhere, as well as experimental verification of

0146-9592/85/090430-03$02.00/0

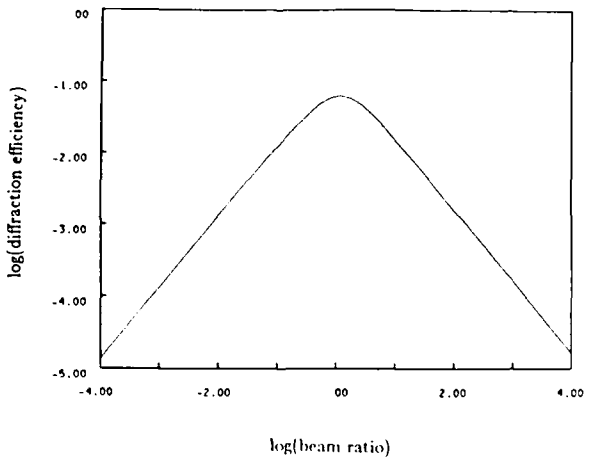


Fig. 1. Diffraction efficiency versus beam ratio. Shown is a typical curve for photorefractive BSO or BGO.

portion of the curve, are given in Ref. 11.) Therefore, a defect can be enhanced by focusing the Fourier transform of the mask onto the photorefractive crystal and making the intensity of the peak spectral component that is due to the defect less than or equal to the intensity of the reference beam. The intensity of the spikes that is due to the periodic structure will be so much greater than the reference-beam intensity that the corresponding diffraction efficiency will be very small. Thus the refractive-index pattern formed inside the crystal performs both recording and filtering operations.

The technique of using a weak reference beam and a strong object beam to perform optical processing is not new. Ragnarsson recorded filters in photographic film with this technique in order to perform division.[13] This technique has also been used in photorefractives to obtain edge enhancement of binary images, by both Huignard and Herriau in BSO[14] and by Feinberg in $BaTiO_3$.[15] However, to our knowledge, this is the first use of the technique in photorefractives to enhance selected features in an object beam and suppress others.

A Fourier-optics analysis can be used to describe the propagation of light from the object to the crystal. Suppose that the mask has dimensions $W \times L$ and that a small transparent defect, located at (x_0, y_0), has dimensions $w \times l$. Let $p(x, y)$ represent one unit cell of the periodic structure, which is spaced at intervals of length a. The intensity of the Fourier transform at the crystal, assuming $W, L \gg a$ and unit illumination, is

$$|T(u,v)|^2 \cong \frac{1}{(\lambda f)^2}\left\{(WL)^2\frac{1}{a^4}\sum_n\sum_m P^2\left(\frac{n}{a},\frac{m}{a}\right) \times \operatorname{sinc}^2\left[W\left(u-\frac{n}{a}\right)\right]\operatorname{sinc}^2\left[L\left(v-\frac{m}{a}\right)\right] + (wl)^2\operatorname{sinc}^2(wu)\operatorname{sinc}^2(lv)\right\},$$

where the sinc function is as defined by Bracewell.[16] The spatial frequencies' variables are related to spatial variables as $u = x/\lambda f$ and $v = y/\lambda f$, and $P(u, v)$ is the Fourier transform of $p(x, y)$. $P(0, 0)$ represents the transmitting area of one period of the pattern, and $P(0, 0)/a^2$ is the fraction of the mask area that is transmitting. At the crystal, the hologram should be recorded such that the intensity of the periodic portion of $|T(u, v)|^2$ is greater than the reference-beam intensity I_r, and the intensity of the defect portion of $|T(u, v)|^2$ is less than I_r. Mathematically, if D is defined as the relevant dynamic range of the periodic portion and I_i is the intensity incident upon the mask, then the two conditions are

$$I_i\left(\frac{WL}{\lambda f}\right)^2\frac{P^2(0,0)}{a^4}\frac{1}{D} > I_r, \qquad I_i\left(\frac{wl}{\lambda f}\right)^2 < I_r.$$

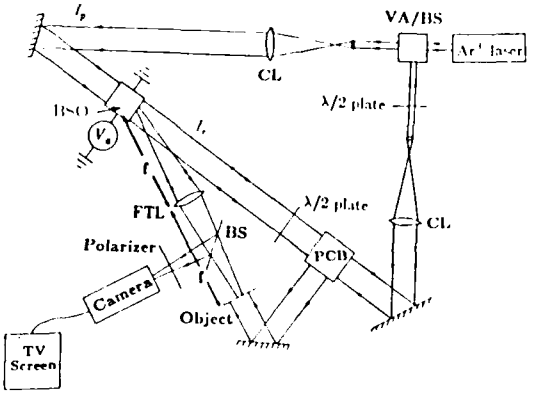
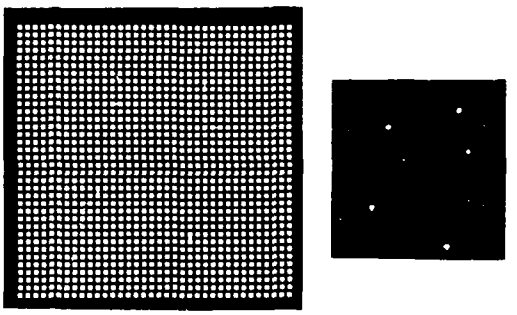


Fig. 2. Experimental setup. VA/BS, variable attenuator/beam splitter; BS, beam splitter; CL, collimating lens; PCB, polarizing-cube beam splitter; FTL, Fourier-transform lens.



(a) (b)

Fig. 3. Input mask and output-defect-enhanced image. The coordinates of the seven defects, measured in units of numbers of squares and taking the center of the lower left-hand square to be (0, 0) are

Defect size (μm^2)	Coordinates (hor., vert.)
100 × 100	(22, 7.5)
50 × 100	(10, 13.5)
100 × 50	(12.5, 25)
10 × 100	(15, 20.5)
25 × 100	(24, 27.5)
100 × 25	(25.5, 22)
100 × 10	(25.5, 15)

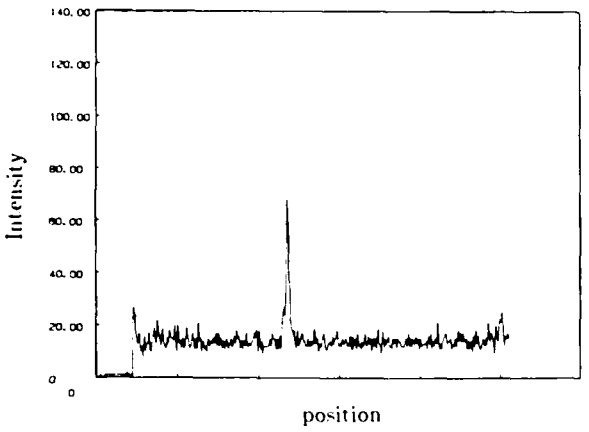


Fig. 4. Intensity line scan of 10 μm × 100 μm defect. Graph illustrating the signal-to-noise ratio obtained for the smallest defect.

If the defect is opaque rather than transmitting, then the second condition should be modified:

$$I_i\left(\frac{wl}{\lambda f}\right)^2\left|\frac{1}{a^2}\sum_n\sum_m P\left(\frac{n}{a},\frac{m}{a}\right)\mathrm{sinc}\left[w\left(\frac{n}{a}\right)\right]\mathrm{sinc}\left[l\left(\frac{m}{a}\right)\right]\right|^2 < I_r.$$

The experimental setup used to obtain defect enhancement is shown in Fig. 2. An argon-ion laser (λ = 514.5 nm) was collimated and split to form the two writing beams as well as the probe (readout) beam. A BSO crystal, of size 8 mm × 8 mm × 8 mm, was oriented with the x direction shown in Fig. 1 along a [110] axis. An f/4.9 lens was used to perform the Fourier transform, and the output was detected by a charge-coupled device (CCD) camera. The combination of a half-wave plate, a polarizing-cube beam splitter (PCB), and a second half-wave plate allowed the beam ratio to be changed while the polarizations were kept the same. To improve the signal-to-noise ratio, a polarizer was placed in front of the output to reduce the scattered light.[17]

The object mask consisted of a 36 × 36 array of squares, each with sides of 150 μm. The spacing between the squares was 100 μm, so the period a was equal to 250 μm. The total mask size was 9 mm × 9 mm. Within this array were placed seven transmitting defects of sizes 100 μm × 100 μm down to 100 μm × 10 μm, as shown in Fig. 3(a). The output of the optical system, obtained using an applied voltage of 4 kV, is shown in Fig. 3(b). The periodic background has been quite effectively suppressed, leaving the defects clearly visible. Figure 4 shows an intensity scan of one line of the output image, illustrating the worst-case signal-to-noise ratio obtained. The defect represented is one of the two 10 μm × 100 μm spots; thus the system appears easily capable of detecting smaller defects.

In recording the hologram, the object-beam intensity at the mask (I_i) was 16 mW/cm^2, and the reference-beam intensity was 3.0 mW/cm^2, which led to beam ratios at the crystal of 0.014 to 0.00014, depending on the size of the defect. Thus the experimental results indicate that enhancement occurs even for values of R much less than one. Because the inverse properties shown in Fig. 1 were derived under conditions of plane-wave illumination, the filtering properties of the crystal cannot be described by simply a beam-ratio dependence. Further investigation into the actual behavior of the crystal is currently being undertaken.

The resolution obtained in the output was constrained by two factors. The primary constraint was the size of the crystal. Given the f-number of the system, the crystal captured only the central fifth of the primary lobe of the sinc function that was due to the smallest defect; therefore the output of the system produced the defect convolved with a smoothing function. Thus reducing the f-number of the optical system (and using a crystal of larger dimensions) will greatly improve the resolution capability. The second constraint on the resolution was the size of the imaging elements of the CCD camera, each of which measured 23 μm × 13.4 μm.

In summary, a method to enhance defects in a periodic mask in real time has been presented. A photorefractive crystal is used to perform holographic recording, filtering, and readout process simultaneously. Preliminary experimental results show detection of defects down to 10 μm × 100 μm in size. Detection of smaller defects should be possible by using an optical system with a smaller f-number and a camera with smaller resolution elements.

This research was supported by the National Science Foundation–Materials Research Laboratory program through the Center for Materials Research at Stanford University and by the U.S. Air Force Office of Scientific Research. The assistance of Mike Smith and Zora Norris in the mask preparation is greatly appreciated.

References

1. J. D. Knox, P. V. Goedertier, D. Fairbanks, and F. Caprari, Solid State Technol. **20**(5), 48 (1978).
2. K. Levy, *Solid State Technol.* **21**(5), 60 (1978).
3. KLA Instruments Corp., Solid State Technol. **26**(1), 45 (1983).
4. D. B. Novotny and D. R. Ciarlo, Solid State Technol. **21**(5), 51 (1978).
5. R. A. Simpson and D. E. Davis, Proc. Soc. Photo-Opt. Instrum. Eng. **334**, 230 (1982).
6. B. Tsujiyama, K. Saito, and K. Kurihara, IEEE Trans. Electron Devices **ED-27**, 1284 (1980).
7. L. S. Watkins, Proc. IEEE **57**, 1634 (1969).
8. N. N. Axelrod, Proc. IEEE **60**, 447 (1972).
9. R. A. Heinz, R. L. Odenweller, Jr., R. C. Oehrle, and L. S. Watkins, Western Elect. Eng. **17**, 39 (1973).
10. R. L. Fusek, K. Harding, L. H. Lin, and S. C. Gustafson, Proc. Soc. Photo-Opt. Instrum. Eng. **523**, 54 (1985).
11. E. Ochoa, L. Hesselink, and J. W. Goodman, Appl. Opt. **24**, 1826 (1985).
12. H. Kogelnik, Bell Syst. Tech. J. **48**, 2909 (1969).
13. S. I. Ragnarsson, Phys. Scr. **2**, 145 (1970).
14. J. P. Huignard and J. P. Herriau, Appl. Opt. **17**, 2671 (1978).
15. J. Feinberg, Opt. Lett. **5**, 330 (1980).
16. R. N. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1978).
17. J. P. Herriau, J. P. Huignard, and P. Aubourg, Appl. Opt. **17**, 1851 (1978).

APPENDIX D

Neural Networks for Computation: Number Representations and Programming Complexity

Mitsuo Takeda[*] *and Joseph W. Goodman*

Department of Electrical Engineering

Stanford University

Stanford, CA 94305

Abstract

Methods for using neural networks for computation are considered. The success of such networks in finding good solutions to complex problems is found to be dependent on the number representation schemes used. Redundant schemes are found to offer advantages in terms of convergence. Neural networks are applied to the combinatorial optimization problem known as the "Hitchcock problem", and signal processing problems, such as matrix inversion, and Fourier transformation . The concept of programming complexity is introduced. It is shown that for some computational problems, the programming complexity may be so great as to limit the utility of neural networks, while for others the investment of computation in programming the network is justified. Simulations of neural networks using a digital computer are presented.

[*] When this work was done, Mitsuo Takeda was on scholarly leave from University of Electro-communications, 1-5-1, Chofugaoka, Chofu, Tokyo, 182, JAPAN

I. Introduction

Even the fastest modern computer cannot compare with the brain of an infant in the performance of intelligent information processing such as image processing and pattern recognition. This well quoted fact suggests the possibility of a quite different type of computer. The fundamental difficulty in creating artificial intelligence on conventional digital computers comes from the large difference in architectures of information processing between digital computers and human brains, i.e., the sequential processing in von Neumann machines and the massively parallel computation in human brains[1]. Neuroscientists have revealed that the massive parallelism and the computational richness in the human brain lie in the global and dense interconnections among a large number of identical logic elements or neurons which are connected to each other with variable strengths by a network of synapses[2]. An artificial neural network system that can perform parallel computation and the function of natural intelligence is extremely attractive as a future-generation computer.

However, there exist two major problems that must be attacked before the realization of such a neural computer. The first is a hardware problem of how to implement those global and dense interconnections among many neuron-like logic elements, and the second is a software problem of how to program such highly parallel computation on a neural network system. We may take two different approaches to the first problem, VLSI-based interconnections and optical interconnections[3]. Neurons in the human brain are interconnected in three-dimensional space since it is the most natural and efficient way of interconnection, but VLSI-based interconnections are inherently two-dimensional in nature. Optical signals, on the other hand, can flow through three-dimensional space to achieve the required interconnects between neuron-like logic elements. Based on this idea, several schemes of optical computing have been

proposed[4-7]. Among them, Psaltis and Farhat[4,7] recently reported an optical implementation of the Hopfield neural network[8,9] using an optical vector-matrix multiplier[10] as a programmable interconnector, and demonstrated the feasibility of optical content addressable associative memory.

Extensive studies have been done on the basic characteristics of the neural networks themselves[11], but the second problem of how to program them to do various computations of practical interest has not been fully studied except in their application to associative memory[12]. Quite recently, Hopfield and Tank[13] showed that a certain class of optimization problems can be programmed and solved on their neural network model. They demonstrated the computational power and speed of their neural network by solving one of the NP-complete problems[14] known as the "Traveling-Salesman problem." The purpose of this paper is to extend their idea and explore new possibilities of programming and solving on neural networks other various non-biological problems of practical interest. We emphasize that our goal is not to propose mechanisms that might actually be utilized by the brain, but rather to apply neural network ideas to computational problems, and thereby to open some new avenues for realizing powerful man-made computers.

We first review briefly the Hopfield neural network model, and describe some minor modifications. Next, we propose a new scheme to represent numbers by neuron state variables, which is essential in solving numerical problems on neural networks. Based on this number representation scheme, we show how we can program and solve combinatorial optimization problems[15] known as network flow problems[16] or more specifically as the "Hitchcock problem,"[17] and simulate its computational performance on a digital computer. Then, we give a programming scheme to perform signal processing for signal recovery, such as the computations of matrix inversion and Fourier

transformation. The performance is again simulated on a digital computer.

The important idea of programming complexity is then introduced, and it is shown that for some problems the data-dependent programming complexity is so great that computations invested in finding the right neural interconnection and bias patterns may equal the complexity involved in solving the problem directly without a neural network. For such problems, neural networks, as we now understand them, may not be an appropriate architecture for computational problem-solving.

We conclude with the discussion of the limitations and the problems that remain to be solved in future.

II. The Hopfield Model and Its Modifications

A. The Hopfield model

The Hopfield model[8,9] consists of a number of mutually interconnected nonlinear devices called "neurons" whose states are characterized by their outputs V_i (which may take values between 0 and 1). The dynamics of neurons in the Hopfield model can be described in both discrete and continuous spaces.

The discrete model is illustrated in Fig.1. At fan-in terminals Σ_i, each neuron i receives inputs $T_{ij}V_j$ from other neurons j and a bias input I_i associated with itself;

$$U_i = \sum_{j=1}^{N} T_{ij}V_j + I_i, \tag{1}$$

where N is the number of neurons, and T_{ij} are elements of an interconnection matrix representing the strengths of connections. At discrete times, switches SW_i turn on, and the inputs U_i are fed back to corresponding neurons to change their states or to leave their states fixed according to a threshold rule determined by nonlinear operators NLR_i, such that

$$V_i(k+1) = \mathrm{stp}[\ U_i(k)\ ], \tag{2}$$

where k is discrete time, and stp(x) is a unit step function which is 1 for $x \geq 0$, and 0 for $x<0$. Thus, neurons take binary values either 1 or 0, and the binary outputs are sent out from fan-out terminals Ω_i and distributed through the interconnection network to re-generate new inputs at the fan-in terminals Σ_i.

In the continuous model, neurons change their states according to the following equations of dynamics:

$$dU_i/dt = \sum_{j=1}^{N} T_{ij}V_j + I_i \tag{3}$$

$$V_i = g(\ U_i\ ), \tag{4}$$

where t is continuous time, and $g(x)$ is a nonlinear function whose form can be taken to be

$$g(x) = (1/2)[\ 1 + \tanh(x/x_0)\ ], \tag{5}$$

which approaches a unit step function as x_0 tends to zero.

Hopfield[9] has shown that if $T_{ij}=T_{ji}$, neurons in the continuous model always change their states in such a manner that they minimize an energy function defined by

$$E \equiv -(1/2)\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij}V_iV_j - \sum_{i=1}^{N} I_iV_i, \tag{6}$$

and stop at minima of this function. The same is also true[8] for neurons in the discrete model if we further assume that $T_{ii}=0$.

B. Neuron transition modes

We adopt the discrete-time model because it is much easier to simulate on a digital computer. But when $T_{ii}\neq 0$, the model sometimes shows an oscillatory behavior or keeps wandering around the state space near the minima of the energy function. Most problems of practical interest require self feedbacks ($T_{ii}\neq 0$) when programmed on a

neural network. We therefore need to *design* transition modes that reduce such phenomena. Without claiming any similarity to natural neuron transition rules, we choose four different discrete-time transition modes for examination.

(a) Direct synchronous transition mode

All the transitions occur simultaneously when the switches SW_i turn on in synchronism at discrete times k. The fan-in inputs are directly fed back to generate new neuron states. A continuous nonlinear function $g(x)$ allows neurons to take state values between 0 and 1. The following equations are assumed to hold:

$$U_i(k) = \sum_{j=1}^{N} T_{ij} V_j(k) + I_i \tag{7}$$

$$V_i(k+1) = g[\ U_i(k)\ ]. \tag{8}$$

(b) Differential synchronous transition mode

The differential equations in the continuous model are approximated by difference equations. Transitions occur synchronously. In this case,

$$U_i(k) - U_i(k-1) = \sum_{j=1}^{N} T_{ij} V_j(k) + I_i \tag{9}$$

$$V_i(k+1) = g[\ U_i(k)\ ].$$

This mode requires one memory cell for each neuron to keep its previous input.

(c) Direct asynchronous transition mode (random delays)

This mode is similar to mode (a), but the switches SW_i turn on and off asynchronously, i.e. with random delays. In this case,

$$U_i(k - \Delta t_i) = \sum_{j=1}^{N} T_{ij}\ V_j(k - \Delta t_i) + I_i \tag{10}$$

$$V_i(k - \Delta t_i + \varepsilon) = g[U_i(k - \Delta t_i)]$$

where Δt_i are skews caused by time delays in the network, and are fractions of one

clock time, while ε is a small positive constant. Without loss of generality we can assume

$$\Delta t_1 \leq \Delta t_2 \leq \ldots \leq \Delta t_N,$$

because the numbering of neurons is arbitrary. In this mode, one particular neuron i need not wait for the last neuron N for synchronization, and when it decides its new state, it can make use of information about new states of other neurons that have already renewed their states.

(d) Differential asynchronous transition mode (random delays)

This is an asynchronous version of mode (b). In this case,

$$U_i(k - \Delta t_i) - U_i(k - \Delta t_i - 1) = \sum_{j=1}^{N} T_{ij} V_j(k - \Delta t_i) + I_i, \tag{11}$$

$$V_i(k - \Delta t_i + \varepsilon) = g[U_i(k - \Delta t_i) - U_i(k - \Delta t_i - 1)]$$

Using simulations on a digital computer, we found that the synchronous transition modes (a) and (b) gave rise to large oscillations in the energy function when $T_{ii} \neq 0$, but that the asynchronous transition modes (c) and (d) have greatly reduced oscillatory or wandering behavior, though the reduction is not complete. While mode (c) is quicker in minimizing the energy function, mode (d) has more reduced oscillations. Depending on the characteristics of the problems of interest, we shall make a proper choice of a mode from (c) and (d).

III. Number Representation Schemes

In most problems of practical interest, solutions are described by a set of numbers. Therefore we must have a means to encode numbers on neuron state variables V_i. While allowing neurons to take continuous state values during the process of energy function minimization, we demand that they take binary values of 1 or 0 at the

final stage so that we can obtain digital solutions like those given by digital computers. For simplicity, we first assume the numbers are positive integers including 0, though we can also represent general bipolar and complex numbers by using additional neurons. We consider three different ways of mapping the positive integer space Z^+ onto the neuron state space V.

A. Binary scheme

A common way of representing numbers in digital computers is to use binary digits. For example, 5 is expressed by 0101. This scheme uses $\log_2(N+1)$ bits to express a number N. If we let one neuron represent one bit, we have a one-to-one correspondence between elements in the number space Z^+ and those in the neuron state space V. Despite the economy in the number of bits or neurons used, a system based on the binary scheme is not fault-tolerant. In other words, even a single failure in a highly significant bit gives rise to a large error in the number represented.

B. Simple-sum scheme

In this scheme, a number is represented by a simple sum of the neuron state variables V_i, i.e., the total number of firing (V_i=1) neurons. For example, 5 is expressed by 0011111, 0101111, 1101011, etc., all of which have five 1-bits. This is a one-to-many mapping from Z^+ to V, and the numbers have degenerate representations. This scheme requires N bits to express a number N, and is not economical in the number of bits or neurons. However, it is highly fault-tolerant because an error in a single bit does not cause a large error in the number represented. The fault-tolerance of the human brain is believed to come from this type of averaging over a large number of neurons[11].

So far, we have compared the binary scheme and the simple-sum scheme from the viewpoint of their fault-tolerance. More important is their difference in problem-

solving capability. As will be seen later, problems are solved through a spontaneous energy minimization process in a neural network, and the solution is given by a point in the neuron state-variable space that is reached after this minimization process. In the binary scheme, there is only one point in the state variable space that gives a correct solution. In the simple-sum scheme, on the other hand, multiple points give the correct solution. Because of this degeneracy and the clustering of quasi-minimum energy points in the neuron state-variable space, the simple-sum scheme offers more chances to reach the correct solution. Suppose, for example, 3 is the correct solution. In the simple-sum scheme, we can get a correct solution when the final state is either 00111, 10110, 11100, or 10101, etc., whereas we can get the correct solution in the binary scheme only when the final state is 00011. Simulation results reported later in this paper support the hypothesized superiority of the simple-sum scheme.

C. Group-and-weight scheme

Despite its merit in fault-tolerance and computational capability, the simple-sum scheme requires too many neurons when solutions include large numbers. We propose the group-and-weight scheme which lies between the binary and the simple-sum schemes. In this scheme, we divide the total q bits into K groups each of which has M bits ($q=KM$), and interpret the groups as digits whose numbers are given by simple sums of the bits in the corresponding groups. For example, with q=6, K=2, M=3, 5 is expressed either by 100 100 ($4^1\times(1+0+0) + 4^0\times(1+0+0) = 5$), 010 001, 001 010, or 100 001 etc. A number expression for the simple-sum scheme is given by

$$\sum_{k=1}^{K}[(M+1)^{k-1}\sum_{i=1}^{M}V_{(k-1)M+i}]. \tag{12}$$

This expression includes the binary and the simple-sum schemes as special cases. When we put M=1 and $K=q$, we obtain a number expression for the binary scheme

$$\sum_{k=1}^{q} 2^{k-1} V_k \,, \tag{13}$$

and when we put $M=q$ and $K=1$, we obtain a number expression for the simple-sum scheme

$$\sum_{i=1}^{q} V_i \,. \tag{14}$$

The group-and-weight scheme requires $M \log_{M+1}(N+1)$ bits to express a number N. This also gives the number of bits required in the binary scheme when we put $M=1$, and that required in the simple-sum scheme when we put $M=N$.

D. Bipolar and Complex Integers

So far, we have restricted our number representations to positive integers, but they can easily be extended to include bipolar and complex integers. A bipolar expression can be obtained simply by adding a negative bias integer to the expression for positive integers given by Eq. (12):

$$\sum_{k=1}^{K} \left[(M+1)^{k-1} \sum_{i=1}^{M} V_{(k-1)M+i} \right] - \left\lceil (1/2) \left[(M+1)^K - 1 \right] \right\rceil \tag{15}$$

where $(1/2)\left[(M+1)^K - 1\right]$ is half the largest positive integer that can be expressed by Eq. (12), and the floor operation $\lceil x \rceil$ gives the nearest integer value less than x. Eq. (15) can express bipolar integers ranging over $\pm\left\lceil (1/2)\left[(M+1)^K - 1\right]\right\rceil$.

To express complex integers, we need twice as many neurons, i.e., neurons $V_i^{(R)}$ and $V_i^{(I)}$ that represent real and imaginary parts, respectively. Complex integers are expressed by

$$\begin{aligned} &\sum_{k=1}^{K} \left[(M+1)^{k-1} \sum_{i=1}^{M} V^{(R)}_{(k-1)M+i} \right] - \left\lceil (1/2) \left[(M+1)^K - 1 \right] \right\rceil \\ &+ j \left\{ \sum_{k=1}^{K} \left[(M+1)^{k-1} \sum_{i=1}^{M} V^{(I)}_{(k-1)M+i} \right] - \left\lceil (1/2) \left[(M+1)^K - 1 \right] \right\rceil \right\} \end{aligned} \tag{16}$$

where $j^2 = -1$.

E. General Real and Complex Numbers

We can also express numbers with fractional digits, e.g. 13.26, 3.14, etc., by using more neurons and labeling them with negative subscripts ($i<0$), e.g. V_{-4}, V_{-12}, etc., so that the parameter k in the first summation in Eq. (12) can run from a negative integer $-K'$; the number representation becomes

$$\sum_{k=-K'}^{K}\left[(M+1)^{k-1}\sum_{i=1}^{M}V_{(k-1)M+i}\right]. \tag{17}$$

Equation (17) can express numbers ranging from 0 to $(M+1)^K - (M+1)^{-(K'+1)}$, with a minimum digit of quantization being $(M+1)^{-(K'+1)}$. Just as we did in subsection D, we can easily modify Eq. (17) to a form similar to Eq. (16), so that it can express general complex numbers. Again here, the group-and-weight scheme includes the binary and simple-sum schemes as special cases. If we put $M=1$ and $K=q$, Eqs. (15), (16), and (17) give the expressions for the binary scheme. Likewise, the expressions for the simple-sum scheme can be obtained by substituting $M=q$ and $K=1$ into Eqs. (15) and (16), and $M=q$ and $K=-K'$ into Eq. (17).

Finally, it should be noted that the number representation schemes we proposed here are all based on linear mapping of the number space onto the neuron state space. In other words, numbers are represented by linear combinations of neuron state variables. This is an important point in designing number representation schemes for the Hopfield neural network, since the energy function Eq. (6) has a quadratic form with respect to neuron state variables. Other nonlinear mapping schemes, like floating point expressions, cannot form the energy function required by the Hopfield model, because the floating-point expressions need to have neuron state variables in exponents. This certainly limits the possibility of covering a wide range of numbers using a small

number of neurons, but for a neural computer it is not a fatal disadvantage because the use of ample neurons with much redundancy is the key to improving its computational capability and system stability.

IV. The Hitchcock Problem

Based on the number representation schemes described in the previous section, we show how a combinatorial optimization problem known as the Hitchcock problem[17] can be programmed and solved on a neural network.

Suppose there are m sources (X=1, . . . , X=m) for a commodity, with S_X units of supply at X, and n sinks (Y=1, . . . ,Y=n) for the commodity, with a demand D_Y at Y, as shown in Fig. 2. If C_{XY} is the unit cost of shipment from X to Y, the Hitchcock problem is to find a flow f_{XY} that satisfies demands for supplies and simultaneously minimizes flow cost. Thus the problem is to minimize

$$\sum_{X=1}^{m} \sum_{Y=1}^{n} C_{XY} f_{XY}, \tag{18}$$

under the constraints

$$\sum_{Y=1}^{n} f_{XY} = S_X \qquad (X=1,2, \ldots ,m), \tag{19}$$

and

$$\sum_{X=1}^{m} f_{XY} = D_Y \qquad (Y=1,2, \ldots ,n). \tag{20}$$

In Table 1, (a) is an example of a unit cost table, and (b) is an example of a solution represented in the form of a flow matrix or a transportation matrix. The flow matrix describes, for example, that from the source at X=2, two units of the commodity should be sent to the demand at Y=1, and one unit to the demand at Y=2.

A. Flow Matrix Representation

Table 2 shows how the flow matrix can be represented by neurons. We assign q neurons to each matrix element to represent its content f_{XY}, so that we use $N=qmn$ neurons in total for the complete representation of the flow matrix. For the convenience of mathematical treatment, we specify each neuron by a set of three subscripts $V_{XY,i}$, where XY specifies the matrix element the neuron belongs to, and i specifies the position of the neuron in that matrix element. Since the group-and-weight number representation scheme includes the binary and simple sum schemes as special cases, we express the flow matrix elements f_{XY} by the group-and-weight scheme:

$$f_{XY} = \sum_{k=1}^{K}\left[(M+1)^{k-1}\sum_{i=1}^{M}V_{XY,(k-1)M+i}\right]. \tag{21}$$

B. Energy Function

We use the spontaneous energy minimization process of a neuron network to solve optimization problems. Since the energy function defined by Eq. (6) has a quadratic form with respect to neuron state variables V_i, we find a quadratic function of $V_{XY,i}$ such that the minimization of the function corresponds to minimizing the flow cost and minimizing violations of the constraints. An energy function that satisfies such requirements is given by

$$\begin{aligned} E = &-(A/2)\sum_{X=1}^{m}\sum_{Y=1}^{n}\sum_{k=1}^{K}\sum_{i=1}^{M}(M+1)^{k-1}\left[1-2V_{XY,(k-1)M+i}\right]^2 \\ &+(B/2)\sum_{X=1}^{m}\left[S_X-\sum_{Y=1}^{n}\sum_{k=1}^{K}\sum_{i=1}^{M}(M+1)^{k-1}V_{XY,(k-1)M+i}\right]^2 \\ &+(C/2)\sum_{Y=1}^{n}\left[D_Y-\sum_{X=1}^{m}\sum_{k=1}^{K}\sum_{i=1}^{M}(M+1)^{k-1}V_{XY,(k-1)M+i}\right]^2 \\ &+(D/2)\left[\sum_{X=1}^{m}\sum_{Y=1}^{n}\sum_{k=1}^{K}\sum_{i=1}^{M}C_{XY}(M+1)^{k-1}V_{XY,(k-1)M+i}\right]^2 \end{aligned} \tag{22}$$

where A, B, C, and D are positive weight factors. The first term weighted by A is

introduced for the binarization of the neuron state variables $V_{XY,i}$, i.e. $V_{XY,i} = 1$ or 0. Because the function $F(V) = -(1-2V)^2$, $(0 \le V \le 1)$ takes minimum values at $V=0$ and $V=1$, minimizing this term assures that the final solution is given by binary numbers. The second term, weighted by B, is introduced to minimize violations of the source constraints given by Eq. (19). Likewise, through minimization of the third term with a weight C, we can satisfy the demand constraints given by Eq. (20). The last term, weighted by D, is for minimization of the total flow cost. The total cost is squared in Eq. (22), but we may also introduce it without squaring, because the cost is always positive. Note that the way we define the energy function is not unique, so that we can solve the same problem by using different programs on the neural network, just as is often the case in solving problems on conventional digital computers.

Considering the various terms represented in Eq. (22), it can be seen that solutions with low energy do not necessarily correspond to solutions with low cost. However, if the weighting constants are properly chosen, then the binarization, source and demand constraints will eventually all be perfectly satisfied, resulting in a one-to-one relation between energy and cost. Thus eventually low energy solutions will correspond to low cost solutions.

C. Interconnection Matrix

By analogy with digital computers, if we regard the expression for the energy function Eq. (22) as a source program, then the next step is to compile or map it onto the interconnection strengths T_{ij} of the neural network. This can be done by comparing Eq. (22) with the energy function Eq. (6), which is now written as

$$E = -(1/2)\sum_{X=1}^{m}\sum_{Y=1}^{n}\sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{X'=1}^{m}\sum_{Y'=1}^{n}\sum_{k'=1}^{K}\sum_{i'=1}^{M} T_{XY,(k-1)M+i;\,X'Y',(k'-1)M+i'}\, V_{XY,(k-1)M+i}\, V_{X'Y',(k'-1)M+i'}$$

$$-\sum_{X=1}^{m}\sum_{Y=1}^{n}\sum_{k=1}^{K}\sum_{i=1}^{M} V_{XY,(k-1)M+i}\, I_{XY,(k-1)M+i} \qquad (23)$$

where $T_{XY,(k-1)M+i;X'Y',(k'-1)M+i'}$ denotes the strength of the interconnection between the neuron at the $\left[(k-1)M+i\right]$th position in the flow matrix element at XY, and the neuron at the $\left[(k'-1)M+i'\right]$th position in the flow matrix element at $X'Y'$. By equating the corresponding coefficients of the two quadratic equations (22) and (23), we can determine the interconnection strengths and the biases:

$$T_{XY,(k-1)M+i;\,X'Y',(k'-1)M+i'} = 4A(M+1)^{k-1}\delta_{XX'}\delta_{YY'}\delta_{kk'}\delta_{ii'} \\ - B(M+1)^{k+k'-2}\delta_{XX'} - C(M+1)^{k+k'-2}\delta_{YY'} - D(M+1)^{k+k'-2}C_{XY}C_{X'Y'}, \tag{24}$$

and

$$I_{XY,(k-1)M+i} = -2A(M+1)^{k-1} + B(M+1)^{k-1}S_X + C(M+1)^{k-1}D_Y, \tag{25}$$

where $\delta_{ZZ'}$ is a Kroneker delta defined by

$$\delta_{ZZ'} = \begin{cases} 1 & (Z=Z') \\ 0 & (Z\neq Z') \end{cases}.$$

In Eq. (24), the first term describes self-feedbacks, the second and third terms represent local interconnections between neurons in the same row ($X'=X$) and in the same column ($Y'=Y$), respectively. The last term describes the global interconnections between all neurons. If we put $M=1$ and $K=q$, we obtain the interconnection strengths and the biases for the binary number representation scheme:

$$T_{XY,k;\,X'Y',k'} = 4A2^{k-1}\delta_{XX'}\delta_{YY'}\delta_{kk'} \\ - B2^{k+k'-2}\delta_{XX'} - C2^{k+k'-2}\delta_{YY'} - D2^{k+k'-2}C_{XY}C_{X'Y'}, \tag{26}$$

and

$$I_{XY,k} = -A2^{k}+B2^{k-1}S_X+C2^{k-1}D_Y. \tag{27}$$

Likewise, the interconnection strengths and the biases for the simple-sum scheme can be obtained by putting $M=q$ and $K=1$:

$$T_{XY,i;X'Y',i'} = 4A\delta_{XX'}\delta_{YY'}\delta_{ii'} - B\delta_{XX'} - C\delta_{YY'} - DC_{XY}C_{X'Y'}, \tag{28}$$

and

$$I_{XY,i} = -2A+BS_X+CD_Y. \tag{29}$$

D. Numerical Experiments

To examine the computational performance of a neural network, we simulated state transitions of neurons by using a digital computer. We used the unit costs and the source and demand constraints listed in Table 1. Based on these data, we determined the interconnection strengths and the biases. Since at present we have no systematic methods for finding the best combination of the weighting factors A, B, C, and D, they were found empirically through the observation of several experimental results. The lack of a systematic method for finding the weighting factors should not be too disturbing. Such a situation is commonly encountered in solving multiple-target optimization problems (on a conventional digital computer), such as lens design problems and color matching problems. However, it should be emphasized that the ability to obtain a good solution depends strongly on making good choices for A, B, C, and D. Throughout the experiments with the Hitchcock problem, we used the direct asynchronous transition mode and the nonlinear function given by Eq. (5) with $0.1 \le x_o \le 1$.

Figure 3 shows an example of the reduction of energy performed by a network with N=60 neurons that represent the flow matrix based on the *binary* number representation scheme ($N=qmn=3\times4\times5=60$, M=1, K=3). Table 3 shows the flow matrices obtained at several points on the curve of Fig. 3. The weight factors were chosen as A=27, $B=C$=80, and D=0.2. Since we have no a priori knowledge about the solution, uniformly distributed random numbers between 0 and 1 were generated and assigned to the initial states of the neurons. Starting from a very high energy state, the neural network reduced its energy spontaneously by changing its state so that the flow matrix could satisfy the constraints while minimizing the total cost. After six iterations, we reached feasible solutions (marked by open circles) that satisfied all the

constraints and gave 40 as the total cost.

After arriving at a solution using the neural network, it is important to develop some understanding of how good that solution might be. To achieve this end, one could enumerate all the feasible solutions that satisfy the constraints, and from this set determine the best solution. However, since it is very hard to enumerate all the solutions of under-determined simultaneous integer equations, Eqs. (19) and (20) (which belong to a family of Diophantine equations), we used a Monte Carlo method and found 50,000 feasible solutions. (Note that this calculation was performed simply to check how well the neural network had performed.) Figure 4 shows a cost histogram of the feasible solutions found. The solution with cost 40 is found to be one of the very good solutions, which would be reached only with a probability of 6×10^{-5} if we searched randomly among the feasible solutions. Yet it is still not the best solution, which was confirmed to be 38 by using a stepping stone algorithm. Figure 5 and Table 4 show another example, for which we assigned 0.5 to the initial states of all neurons, so that they started evolving from the fuzziest states. In this example, we reached a feasible solution with cost 49 at the seventh iteration, but we could not reach any other feasible solutions by further iterations. The oscillatory behavior of the energy function arises from using a discrete model with self-feedback. The solution with cost 49 is fairly good but not as good as in the previous example. Experiments performed with different initial values and/or weight factors gave solutions most frequently with costs around 50, and could not pick up the best solution. In worst cases, no feasible solution could be reached. These results are indicative of the limitations of the problem-solving capability of the binary number representation scheme. As we now show, much better results can be obtained with a degenerate number representation scheme.

To examine the problem-solving capability of the degenerate number representation schemes, we programmed the same problem on a 140-neuron network using the simple-sum scheme ($N = qmn = 7\times4\times5 = 140$, M=7, K=1). Figures 6 and 7 and Tables 5 and 6 show the computational performance of the 140-neuron network with its initial states all set equal to 0.5, the fuzziest states. Weight factors were chosen to be A=29, B=80, C=80, and D=0.55. Through the first several iterations, the source and demand constraints came to be almost satisfied (see Fig. 6 and Table 5), and at the sixth iteration the first feasible solution, with cost 43, was reached (see Fig. 7 and Table 6). The solution was improved further by continuing iterations, passing another feasible solution with cost 40 at the tenth iteration, the best solution with cost 38 was finally reached on the twenty-first iteration. To show the role played by the degeneracy of the number representation, the complete states of the 140 neurons are depicted in Fig. 8 for the iterations from No. 21 through No. 28. Each neuron is represented by a star when it is firing ($V_{XY,i} = 1$) and by a dot when not firing ($V_{XY,i} = 0$). The number of neurons that are firing in each set of seven neurons represents the content of the flow matrix element f_{XY} at the corresponding position. At iteration No. 21, for example, we had f_{25}=1 because only one neuron $V_{25,3}$ was firing ($V_{25,3}$=1) and the rest of the six neurons were not firing. At iteration No. 22, neuron $V_{25,3}$ stopped firing, but the correct solution f_{25}=1 was retained because the next neighbor neuron $V_{25,2}$ started firing, instead of $V_{25,3}$. We can observe a similar phenomenon in other sets of neurons representing f_{35} and f_{45} at iterations No. 21, 22, 23, 25, 26, and 27. In this manner, the neural network can give correct solutions at many different points in its state space, and these points cluster in a particular region of the state space that corresponds to low energy function values. It is because of this characteristic that the degenerate number representation scheme can have better problem-solving capabilities than the

pure binary number representation scheme.

Figure 9 and Table 7 show another example of the computational performance of the 140-neuron network, where uniform random numbers between 0 and 1 were assigned to the initial state variables of the neurons. In this example, we obtained two different solutions with cost 38, showing that the best solution is not unique.

V. Simultaneous Equations

In this section we show how we can program and solve on a neural network simultaneous equations

$$\mathbf{H}\mathbf{x} = \mathbf{y} \tag{30}$$

where $\mathbf{H}$ is a full-rank square matrix with $N \times N$ elements, and $\mathbf{x}$ and $\mathbf{y}$ are vectors with N elements representing, respectively, unknown and given variables. (Note that deconvolution is a special case of this general problem.)

A. Energy Function

In order to use the spontaneous energy-minimization process of the neural network, we reformulate the problem in the form of a minimization problem by introducing an energy function that includes a term

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2, \tag{31}$$

so that the norm of the difference can be minimized through the energy minimization process. For our later demonstration of the Fourier transformation, we allow $\mathbf{y}$ and $\mathbf{H}$ to take on complex values, but, for the sake of simplicity, we restrict $\mathbf{x}$ to only positive integer values, although we could include complex numbers by using additional neurons labeled by a more complicated set of subscripts. As in Eq. (21), we express the nth element x_n of the unknown vector $\mathbf{x}$ by the group-and-weight scheme:

$$x_n = \sum_{k=1}^{K}\left[(M+1)^{k-1}\sum_{i=1}^{M}V_{n,(k-1)M+i}\right]. \tag{32}$$

By substituting Eq. (32) into Eq. (31), we have an energy function

$$\begin{aligned} E = &-(A/2)\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{M}(M+1)^{k-1}\left[1-2V_{n,(k-1)M+i}\right]^2 \\ &+(B/2)\sum_{l=1}^{N}\left[y_l-\sum_{n=1}^{N}h_{ln}x_n\right]\left[y^*_l-\sum_{n'=1}^{N}h^*_{ln'}x_{n'}\right] \end{aligned} \tag{33}$$

$$\begin{aligned} &= -(A/2)\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{M}(M+1)^{k-1}\left[1-2V_{n,(k-1)M+i}\right]^2 \\ &+ (B/2)\sum_{l=1}^{N}\sum_{n=1}^{N}\sum_{n'=1}^{N}\sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{i=1}^{M}\sum_{i'=1}^{M}(M+1)^{k+k'-2}h_{ln}h^*_{ln'}V_{n,(k-1)M+i}\,V_{n',(k'-1)M+i'} \\ &- B\sum_{l=1}^{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{M}(M+1)^{k-1}\mathrm{Re}[y_l h_{ln}^*]V_{n,(k-1)M+i} \\ &+ (B/2)\sum_{l=1}^{N}|y_l|^2, \end{aligned}$$

where, as in Eq. (22), the first term is for binarization, y_l and h_{ln} are elements of **y** and **H**, and * and Re[ ] denote complex conjugate and real part, respectively.

B. Interconnection Matrix

The energy function is now modified to

$$\begin{aligned} E = &-(1/2)\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{n'=1}^{N}\sum_{k'=1}^{K}\sum_{i'=1}^{M}T_{n,(k-1)M+i;n',(k'-1)M+i'}\,V_{n,(k-1)M+i}\,V_{n',(k'-1)M+i'} \\ &-\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{M}V_{n,(k-1)M+i}\,I_{n,(k-1)M+i}. \end{aligned} \tag{34}$$

By equating the corresponding coefficients of Eq. (33) and (34), we determine the interconnection strengths and the biases:

$$T_{n,(k-1)M+i;n',(k'-1)M+i'} = 4A(M+1)^{k-1}\delta_{nn'}\delta_{kk'}\delta_{ii'} - B(M+1)^{k+k'-2}\sum_{l=1}^{N}h_{ln}h^*_{ln'}, \tag{35}$$

$$I_{n,(k-1)M+i} = -2A(M+1)^{k-1} + B(M+1)^{k-1}\mathrm{Re}\left[\sum_{l=1}^{N}h_{ln}^*y_l\right]. \tag{36}$$

Equation (31) includes the discrete Fourier transform as a special case with

$$h_{ln} = exp[-2\pi j(l-1)(n-1)/N], \tag{37}$$

and the inverse transform is computed by solving the simultaneous linear equations.

In this case, Eq. (35) takes a simple form due to the orthogonality of the Fourier transform matrix:

$$T_{n,(k-1)M+i;n',(k'-1)M+i'} = 4A(M+1)^{k-1}\delta_{nn'}\delta_{kk'}\delta_{ii'} - BN(M+1)^{k+k'-2}\delta_{nn'}. \tag{38}$$

C. Numerical Experiments

Computations of the inverse Fourier transform were programmed on the neural network and the performance was simulated on a digital computer. We used signals with N=15 sample points. Each sample-point x_n was expressed by 24 neurons based on the simple-sum scheme (M=24, K=1), so that 360 neurons were employed in total. We adopted the differential asynchronous transition mode, and chose weight factors as A=28 and B=1. In Fig. 10, (a) and (b) show, respectively, an original signal **x** and its Fourier transform **y** (only absolute values are shown in the figure). The task given to the neural network is to compute **x** from a given **y**.

Assuming no a priori knowledge, we started from the fuzziest initial states $V_{n,i}$ = 0.5 shown in Fig. 10 (c) and got the result shown in Fig. 10 (d) after only two iterations. Another example is shown in Fig. 11, where we used an asymmetric signal and started from random initial states. Again after only two iterations we obtained the result shown in Fig. 11 (d). Although the solutions obtained are not exact, the speed of computation is impressive. In fact, this apparently enormous speed of computation is quite misleading, for reasons that will be revealed later in the following section.

VI. Computational and Programming Complexities

As has been demonstrated in Sections IV, and V, the computational speed of a neural network is very high, solutions (though not always exact) being obtained within several clock times (iterations). At present, we do not know how the computation time

(the number of iterations required) is related to the problem size (the number of neurons employed) and to the algorithm (the choice of the interconnections). We conjecture that the computation time does not grow too rapidly with problem size, because the greater the problem size, the more neurons participate in solving the problem, and the higher the parallelism used. If this conjecture is correct, the computation time is very short for a properly programmed (interconnected) neural network, irrespective of the problem size. It may appear, then, that neural networks would be the computation architecture of choice in most problems that can be included within the energy minimization framework. However, this conclusion is not correct. Although the computation time itself may be very short, it may be necessary to invest very significant computation time simply to program the network, i.e. to determine the proper interconnection strengths and neural biases. The situation is somewhat analogous to the classical analog electronic computer for which a large amount of time must be spent wiring the proper modules together before any problem can be solved. Once the modules are connected, a solution appears almost immediately.

A. Programming Complexity

By analogy with the concept of computational complexity[14, 15] in digital computing, we introduce the concept of *programming complexity* in neural computing. We define programming complexity as the number of arithmetic operations that must be performed to determine the proper interconnection strengths and neural biases for the problem to be solved. Conventional digital computers also need programming, but once the program is compiled and stored in memory, it can be used on many different sets of input data. For this reason, the concept of programming complexity has little significance in the world of conventional digital computers, where programs are completely separable from data. In neural network computers, a program and data are

generally mixed together and stored in the interconnection strengths and/or neural biases. For example, in Eq. (24), the first three terms represent part of the program (since they do not depend on data), and the last term, including the costs C_{XY}, corresponds to the data. Therefore, we must redetermine the interconnection strengths and/or the biases each time we use new data. In such an environment, the programming complexity becomes an important measure of the efficiency of neural computing. We know that it is not meaningful to compare the efficiencies of conventional digital computers and neural computers on the basis of computational complexity and programming complexity, because they mean different things. Digital computers always give exact solutions (within the machine precision) after performing the number of operations specified by the computational complexity, whereas neural computers do not guarantee exact solutions even if they are programmed by performing the number of operations specified by the programming complexity. Nevertheless, a comparison of the computational complexity and the programming complexity does reveal certain interesting aspects of neural computing, as discussed in the following section.

B. Simultaneous Equations

To solve simultaneous equations with N unknown variables, we employed qN neurons, with q being the number of neurons used to represent each unknown variable. We consider q to be a constant factor, since it does not depend on N. The number of interconnections is given by $(1/2)qN(qN+1)=O(N^2)$, and the number of biases is $qN=O(N)$. We need $O(N)$ operations to determine each interconnection strength (see Eq. (35)), and each bias (see Eq. (36)), so that the programming complexity is $O(N^3)$. The computational complexity of this problem is also $O(N^3)$.[14]. This means that solutions of such a problem on either a neural computer or a conventional digital computer would require essentially the same computational load. In the case of the neural computer, the computations must be expended to determine the interconnection stengths and biases, while in the case of the conventional digital computer the computations are expended on solving the problem itself.

This comparison is even more striking in the case of the Fourier transformation discussed earlier. Since Eq. (38) contains no data terms, we need not recompute the interconnection strengths for each different set of data. The programming complexity comes only from computation of the term $\sum_{l=1}^{N} h_{\ln}^{*} y_l$ in the biases, Eq. (36). Noting Eq. (37), we find that to determine the proper biases, we must in fact compute the very same inverse Fourier transform that the neural network was to find! Thus we have already arrived at the solution by the time we finish programming, and it is now no surprise the neural network supplies the answer in only two interations. The answer is in fact pre-programmed into the machine!

C. The Traveling-Salesman Problem

In the previous section we saw an example in which the programming complexity of a neural computer and the computational complexity on a conventional computer are of the same order. The question naturally arises as to whether this is the case with all problems. If so, neural computing loses most of its attractiveness. Hopfield and Tank's paper[13] on the traveling salesman problem provides the best example with which to answer this question. The computational complexity of the traveling salesman problem is an exponential function, $O(N!)$, of the number of cities N. Hopfield and Tank showed that the problem can be programmed on a neural network with N^2 neurons that represent the elements of a permutation matrix. We can show that the programming complexity of this scheme is $O(N^3)$. This large difference of complexities makes neural computing very attractive, even though it does not guarantee the best solution.

C. The Hitchcock Problem

Computational complexity in conventional digital computing depends greatly on the algorithms used, so that a great effort has been made by computer scientists to seek better algorithms and thereby reduce computational complexity. The same can be true with programming complexity in neural computing. The Hitchcock problem provides a good example for demonstrating good and poor algorithms (ways of interconnection) in terms of programming complexity. In Section IV, the Hitchcock problem with m sources and n demands was solved by using $qmn \simeq O(mn)$ neurons. Since Eqs. (24) and (25) include data C_{XY}, S_X, and D_Y, we have to redetermine $(1/2)qmn(1+qmn) \simeq O(m^2n^2)$ interconnection strengths and $qmn \simeq O(mn)$ biases for each new set of data. Each interconnection strength and bias can be determined by a constant number of operations, so that the programming complexity is given by

$O(m^2n^2) = O(n^4)$ for $m \approx n$. In Section IV B, we suggested an alternative definition of the energy function that does not square the total cost in the last term of Eq. (22). If we use this new energy function, the interconnection strengths and biases become

$$T_{XY,(k-1)M+i;X'Y',(k'-1)M+i'} = 4A(M+1)^{k-1}\delta_{XX'}\delta_{YY'}\delta_{kk'}\delta_{ii'} - B(M+1)^{k+k'-2}\delta_{XX'} - C(M+1)^{k+k'-2}\delta_{YY'}, \tag{39}$$

$$I_{XY,(k-1)M+i} = -2A(M+1)^{k-1} + B(M+1)^{k-1}S_X + C(M+1)^{k-1}D_Y - (1/2)D(M+1)^{k-1}C_{XY}. \tag{40}$$

Now the interconnection strengths do not depend on the data C_{XY}, and they need not be re-determined for each new set of data, so that the programming complexity comes only from the biases, Eq. (40), and is given by $O(mn) \approx O(n^2)$ for $m \approx n$. This is a very significant improvement. The computational complexity of the Hitchcock problem depends on the algorithm used by a conventional digital computer. If we search for the best solution randomly among all the possible combinations of the neural states, it becomes $2^{qmn} \approx O(2^{mn})$. Even if we restrict the search to feasible solutions, it can still be exponential $O(n^{m-1}m^{n-1})$ [18]. Of course, these algorithms are worst extremes, and there exist several good algorithms that are in practical use. We do not know exactly what is the computational complexity of the best existing algorithm for the Hitchcock problem, but we estimate it to be a low-order polynomial. If it is still higher than $O(mn)$, then neural computing can have an advantage for this problem.

VII. Conclusion

Following the lead of Hopfield and Tank, we proposed an architecture for programming highly parallel computation on neural networks. In Section III, we described number representation schemes based on linear mapping of the number space onto the neuron space, and pointed out the advantage of the degenerate number representation

schemes. In Sections IV,and V, the validity of the architecture was demonstrated by solving the Hitchcock problem and simultaneous linear equations on neural networks. The dynamics of the neural network were simulated on a digital computer. In Section VI, we introduced the new concept of programming complexity in neural computing, which was used to evaluate the computational efficiency of algorithms performed on neural networks. We compared the programming complexity with the "worst case" computational complexity, simply because the "average" complexity was too hard to estimate. However, we note that programming complexity is better compared with "average" computational complexity, because they have a common characteristic that the solution is not always best or exact, even if we perform the number of operations specified by these complexities.

Finally we point out that there exists a fundamental limitation to the class of problems that can be programmed and solved on the Hopfield neural network. This limitation comes from the requirement that the energy function must be a quadratic function of the neuron state variables. All linear problems, such as discussed in this paper, can satisfy this requirement. However, general nonlinear problems cannot satisfy this requirement. Floating-point number representation is one such nonlinear problem.

Acknowledgement

We wish to thank Dorothy Mighell for careful reading of the manuscript on numerous occasions, and many useful suggestions for improvement.

References

1. See, for example, G.E.Hinton, "Learning in Parallel Networks," *Byte 10,* 265 (1985); J.A.Feldman, "Connections," *Byte 10,* 277 (1985), a special issue for artificial intelligence.

2. See, for example, J.J.Hopfield, "Brain, Computer, and Memory," *Eng.& Sci. 46,* 2 (1982).

3. J.W.Goodman, F.J.Leonberger, S.-Y.Kung, and R.A.Athale, "Optical Interconnections for VLSI Systems," *Proc. IEEE 72,* 850 (1984).

4. D.Psaltis and N.Farhat, "Optical Information Processing Based on an Associative-memory Model of Neural Nets with Thresholding and Feedback," *Opt. Lett. 10,* 98 (1985).

5. G.Eichmann and H.J.Caulfield, "Optical Learning (Inference) Machines," *Appl. Opt. 24,* 2051 (1985).

6. H.Mada, "Architecture for Optical Computing Using Holographic Associative Memories," *Appl. Opt. 24,* 2063 (1985).

7. N.H.Farhat, D.Psaltis, A.Prata, and E.Paek, "Optical Implementation of the Hopfield Model," Appl. Opt. 24, 1469 (1985).

8. J.J.Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proc. Natl. Acad. Sci. USA 79,* 2554 (1982).

9. J.J.Hopfield, "Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons," *Proc. Natl. Acad. Sci. USA, 81,* 3088 (1984).

10. J.W.Goodman, A.R.Dias, and L.M.Woody, "Fully-Parallel High-Speed Incoherent Optical Method for Performing Discrete Fourier Transforms," *Opt. Lett. 2,* 1 (1978).

11. See for example, S.Amari, *Mathematical Theory of Neural Networks,* Sangyotosho Pub., Tokyo (1978), in Japanese; T.Kohonen, *Self-Organization and Associative Memory,* Springer-Verlag, New York (1984).

12. See for example, T.Kohonen, *Content Addressable Memories,* Springer Verlag, New York (1980); K.Nakano, "Associatron - A Model of Associative Memory," *IEEE Trans. Syst. Man Cybern. SMC-2,* 380 (1972).

13. J.J.Hopfield and D.W.Tank, "Neural Computation of Decisions in Optimization Problems," *Bio. Cybern.* (1985), to appear.

14. A.V.Aho, J.E.Hopcroft, and J.D.Ullman, *The Design and Analysis of Computer Algorithms,* Addison-Wesley, Reading Mass. (1974).

15. C.H.Papadimitriou and K.Steiglitz, *Combinatorial Optimization: Algorithms and Complexity,* Prentice-Hall, Englewood Cliffs, N.J. (1982); K.Murty, *Linear and Combinatorial Programming,* John-Wiley & Sons, N.Y. (1976).

16. M.Iri, *Network Flow, Transportation and Scheduling,* Academic Press, N.Y. (1969); D.T.Phillips, A.Garcia-Dias, *Fundamentals of Network Analysis,* Prentice-Hall, N.Y. (1981).

17. F.L.Hitchcock, "The Distribution of a Product from Several Sources to Numerous Localities," *J. Math. Phys., 20,* 224 (1941).

18. S.I. Glass, *Linear Programming,* McGraw-Hill Book Co., New York, NY (1958).

Figure Captions

1. Neural network model.
2. The Hitchcock Problem, with 4 sources and 5 demands.
3. Neural dynamics for the Hitchcock problem, using a binary number representation scheme. The initial states are randomly generated from a seed, and the transition mode is direct asynchronous. The final transportation matrix gives a network flow cost of 40. The constants in the energy function are chosen as A=27, B=C=80, and D=0.2. The constant x_0 is 0.5. See Table 3 for the flow matrices at the iteration numbers indicated by the arrows.
4. Flow cost histogram for the Hitchcock problem. The number of samples is 50000.
5. Second example of the Hitchcock problem using a binary number representation scheme. Uniformly fuzzy states initialized the network, and a "softer" non-linear function was used to give the best solution, with a flow cost of 49. The weights used were A=27, B=C=80, D=0.2. The constant x_0 was 1.0. The open circle represents a solution that satisfied the constraints. See Table 4 for flow matrices at the iteration numbers indicated by the arrows.
6. Network dynamics of the Hitchcock problem using a degenerate (simple sum) number representation scheme. The constants used were A=29, B=C=80, D=0.55, and x_0=0.1. Open circles again represent solutions that satisfy the constraints. Flow matrices corresponding to the arrows are found in Table 5.
7. Continuation of the degenerate network. One of the two-in-50,000 best solutions is found at time 21. Open circles represent solutions that satisfy the constraints (i.e. "consistent" solutions). The cost associated with the solution at the sixth

iteration is 43, that associated with the group of consistent solutions starting at iteration 10 is 40, and that associated with the remaining consistent solutions is 38. See Table 6 for the corresponding flow matrices.

8. Neural state transitions of the degenerate (simple sum) Hitchcock network (Figs. 6 and 7). Iterations 21 through 28 are shown.

9. Second example of the degenerate (simple sum) Hitchcock network. A random initial state drove this network to find both of the best solutions. The two flow matrices are shown in Table 7.

10. Inverse DFT. The transition mode is differential asynchronous. (a) Unknown signal. (b) Known Fourier transform. (c) Uniformly fuzzy initial states. (d) Estimated signal after 2 interations.

11. Inverse DFT, second example. (a) Unknown asymmetric signal. (b) Known Fourier transform. (c) Random initial states. (d) Estimated signal after 2 iterations.

Table Captions

1. (a) Cost matrix for the Hitchcock problem. (b) Sample solution depicting the flow from source X to demand Y.

2. Neural representation of the flow matrix for the Hitchcock network flow problem. q neurons are used to represent one element of the flow matrix.

3. Flow matrices for the specified numbers of iterations,corresponding to the points indicated on Fig. 3.

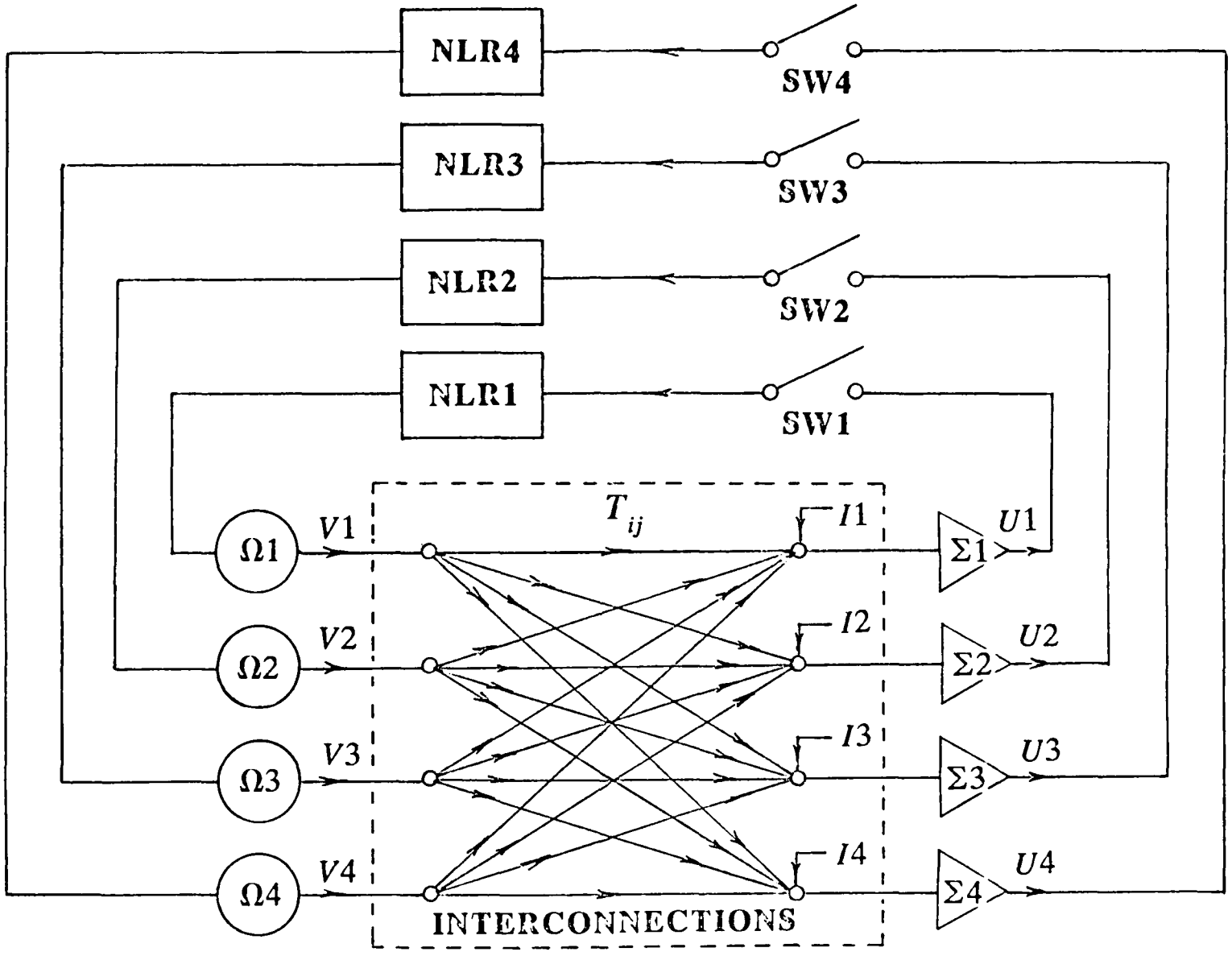
4. Flow matrices for the specified numbers of iterations, corresponding to points indicated on Fig. 5.

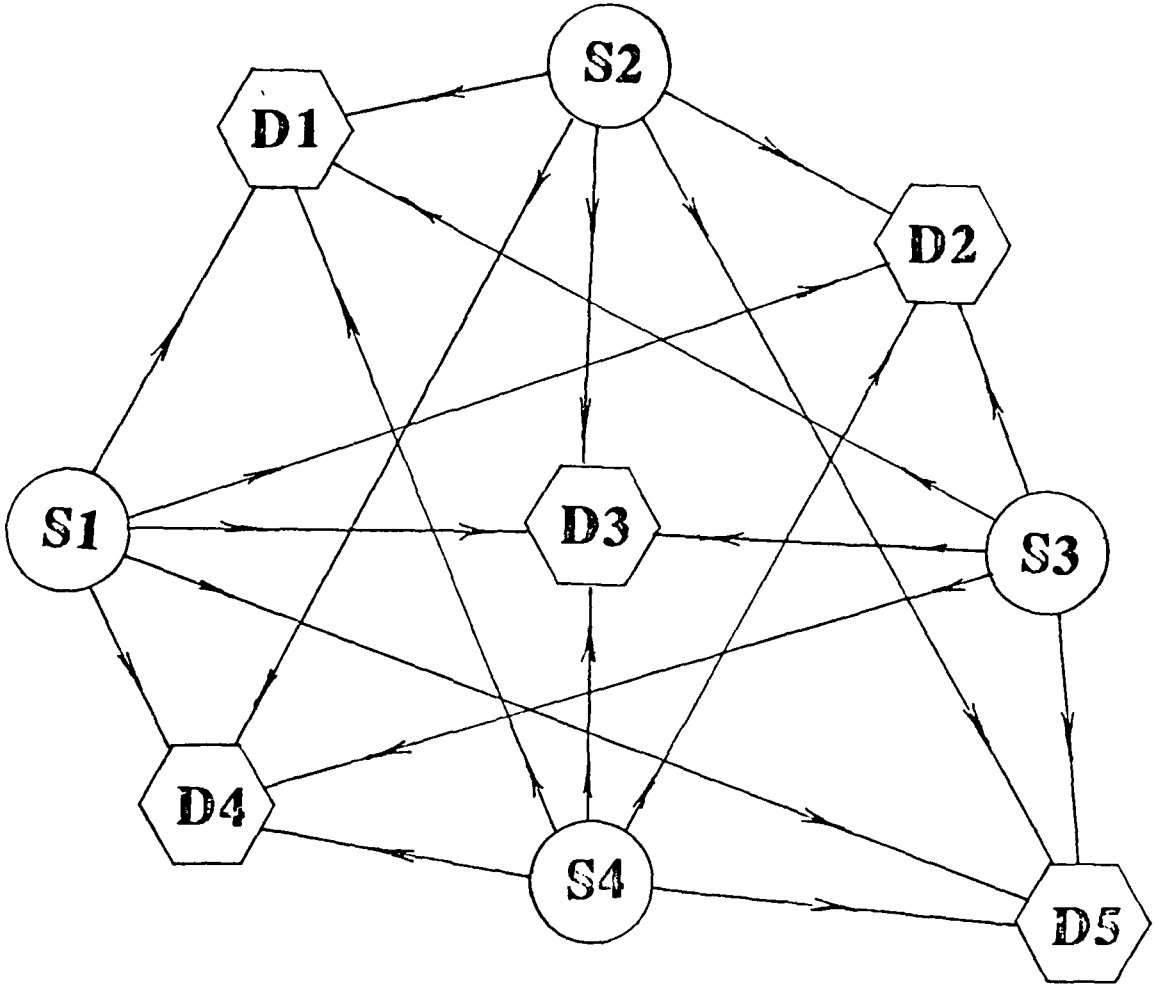
5. Flow matrices for the specified numbers of iterations, corresponding to the points

indicated on Fig. 6.

6. Flow matrices for the specified numbers of iterations, corresponding to the points indicated in Fig. 7.

7. Flow matrices for the specified numbers of iterations, corresponding to the points indicated on Fig. 9.


NLR4
NLR3
NLR2
NLR1
SW4
SW3
SW2
SW1
T_{ij}
Ω1
Ω2
Ω3
Ω4
V1
V2
V3
V4
I1
I2
I3
I4
Σ1
Σ2
Σ3
Σ4
U1
U2
U3
U4
INTERCONNECTIONS



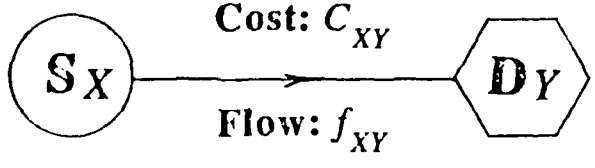
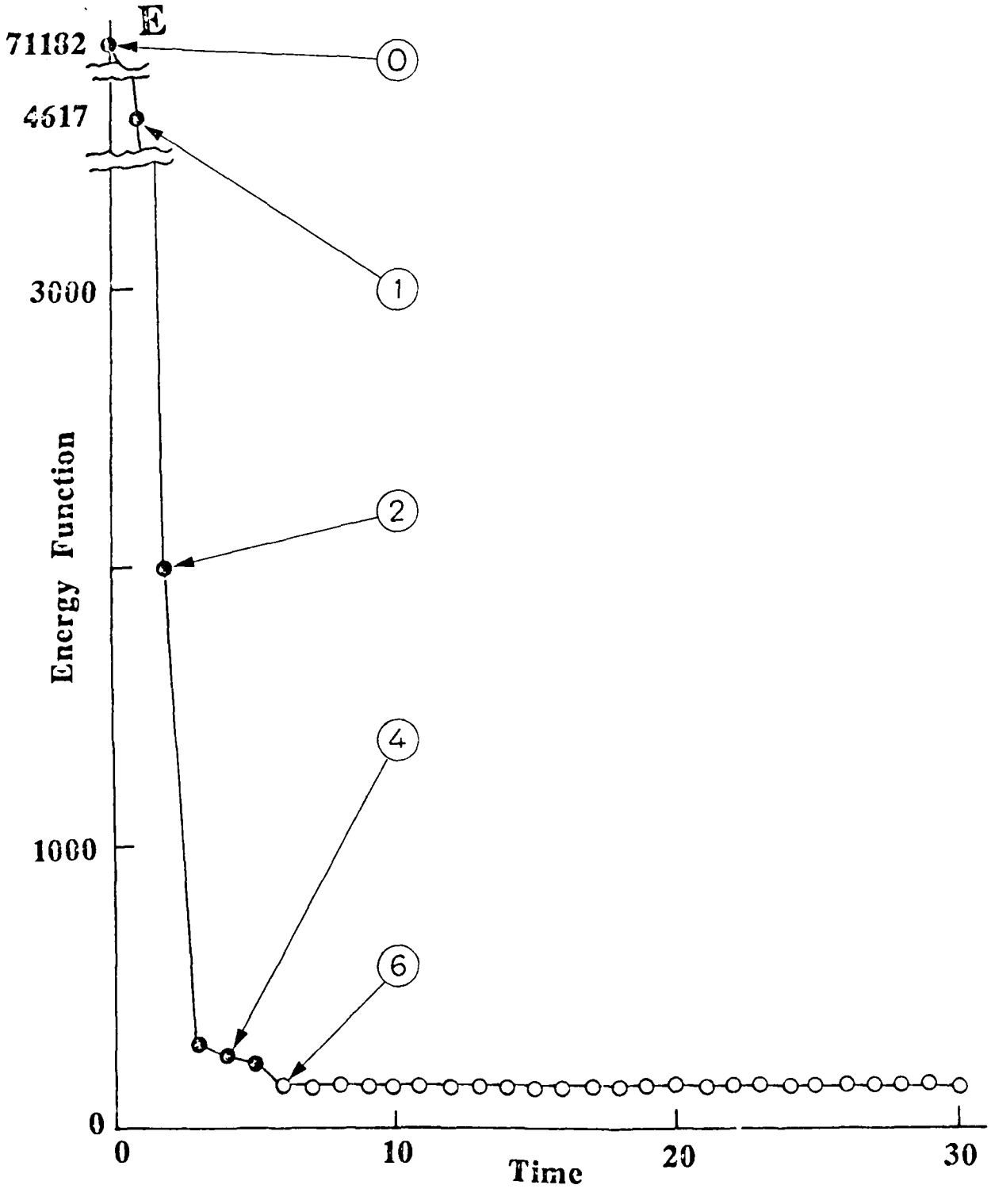
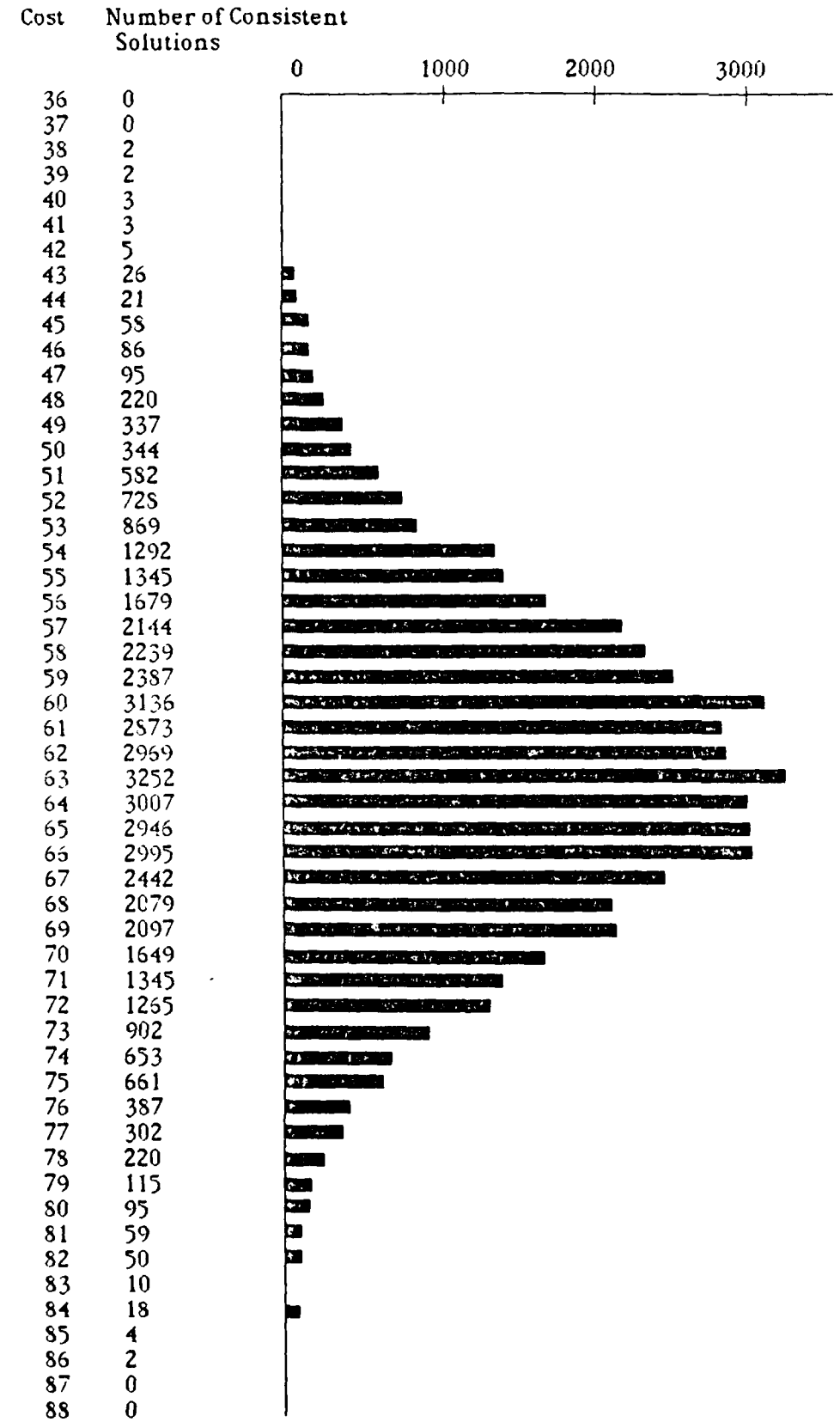
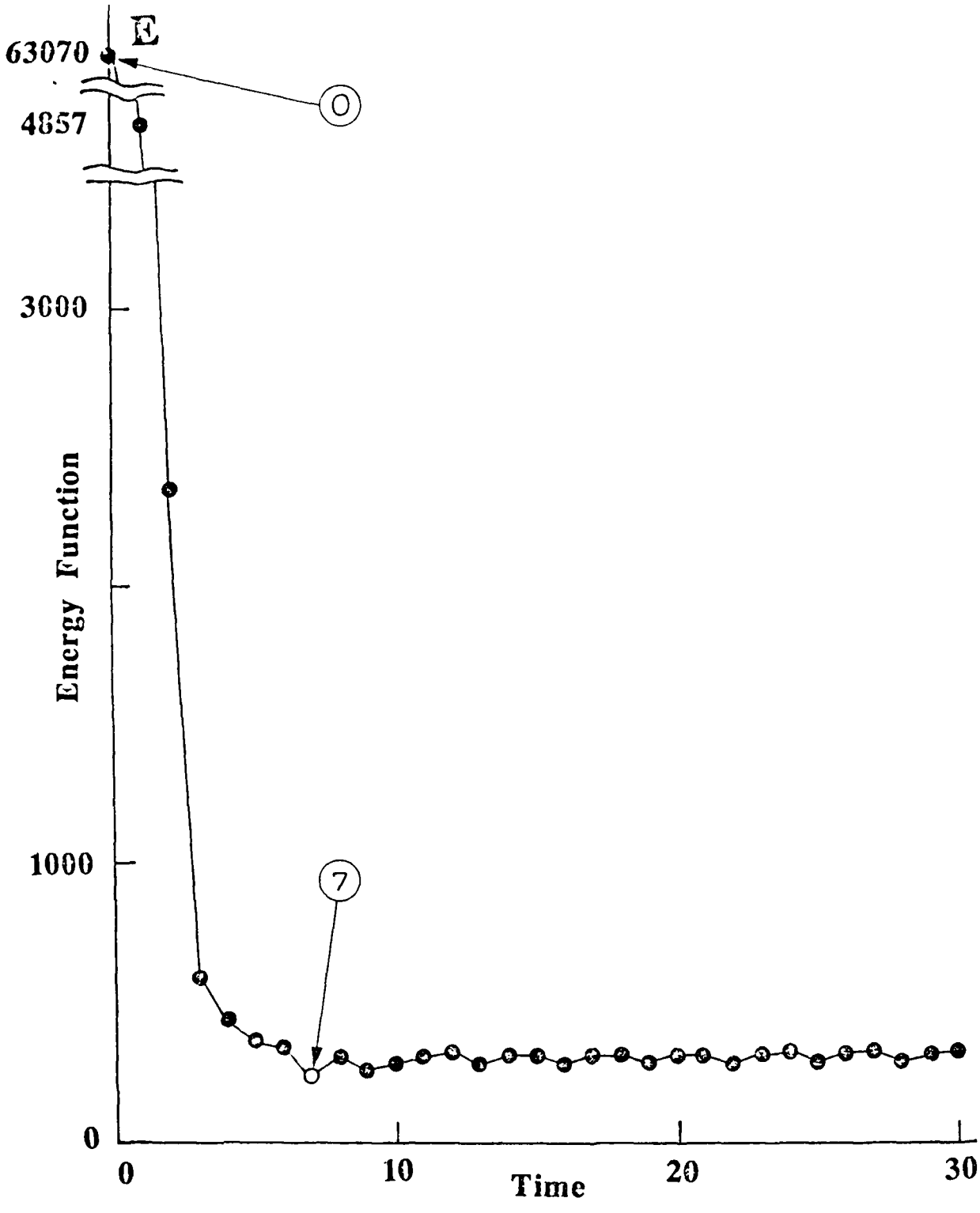
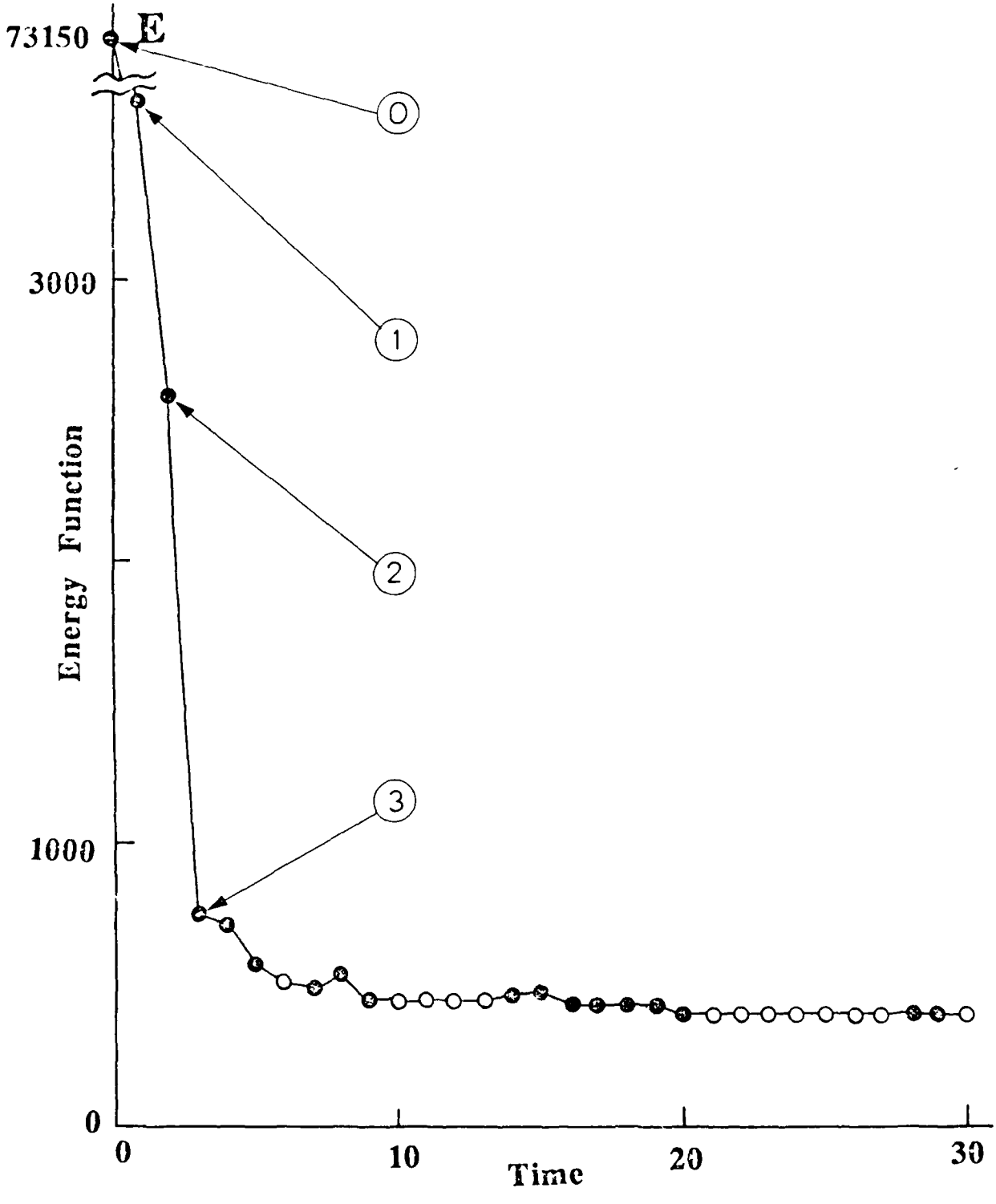


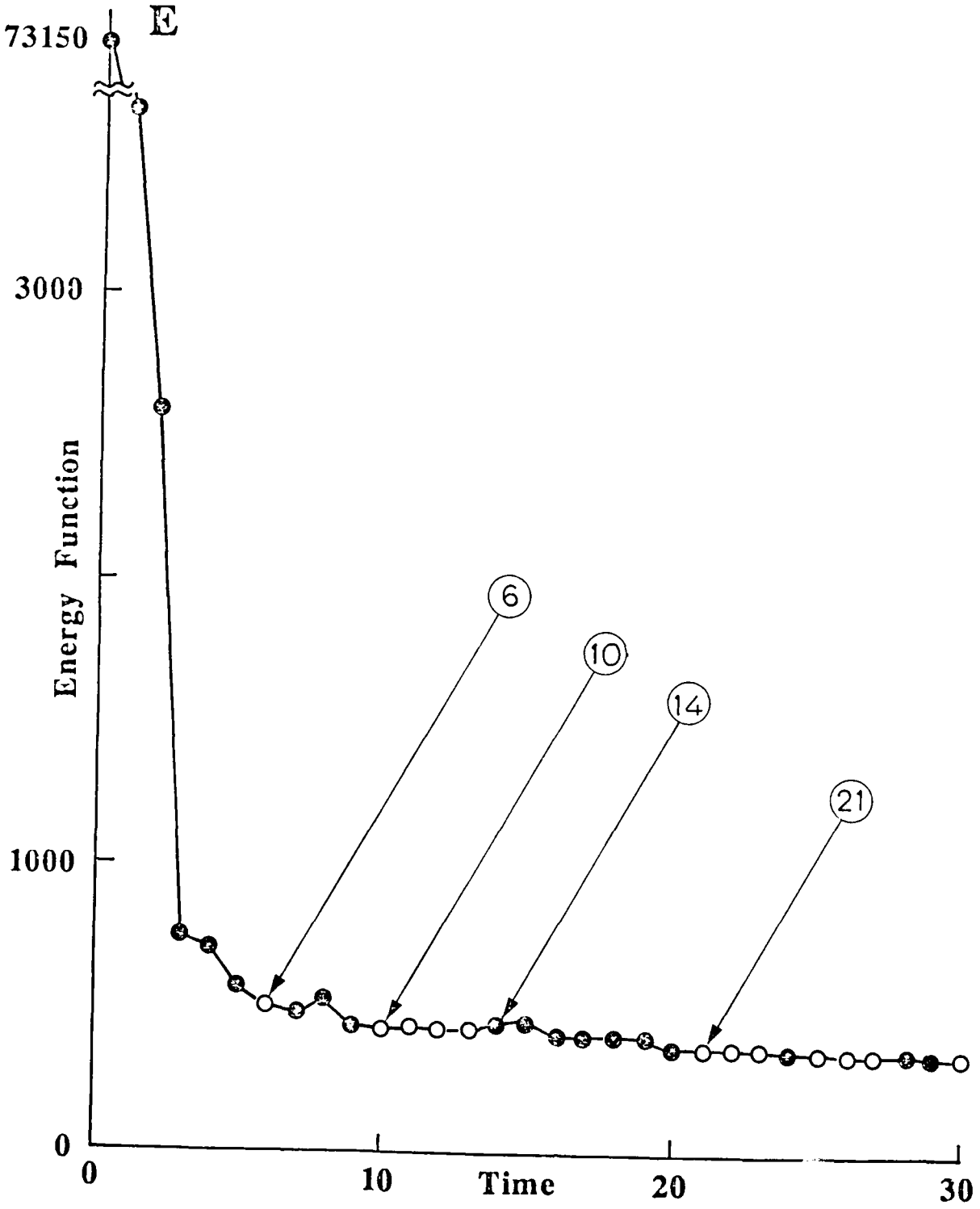
Fig. 2 TAKEDA / GOLDMAN


E
71182
4617
3000
1000
0
Energy Function
0
10
20
30
Time
0
1
2
4
6

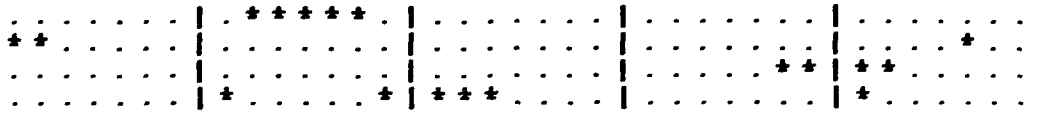

Cost
Number of Consistent Solutions
0
1000
2000
3000
36 0
37 0
38 2
39 2
40 3
41 3
42 5
43 26
44 21
45 58
46 86
47 95
48 220
49 337
50 344
51 582
52 728
53 869
54 1292
55 1345
56 1679
57 2144
58 2239
59 2387
60 3136
61 2873
62 2969
63 3252
64 3007
65 2946
66 2995
67 2442
68 2079
69 2097
70 1649
71 1345
72 1265
73 902
74 653
75 661
76 387
77 302
78 220
79 115
80 95
81 59
82 50
83 10
84 18
85 4
86 2
87 0
88 0


E
63070
4857
0
3000
1000
0
Energy Function
7
0
10
20
30
Time

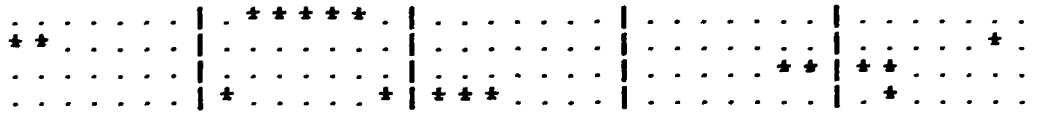

73150
E
0
3000
1
Energy Function
2
3
1000
0
0
10
20
30
Time


E
73150
3000
1000
0
Energy Function
6
10
14
21
0
10
20
30
Time

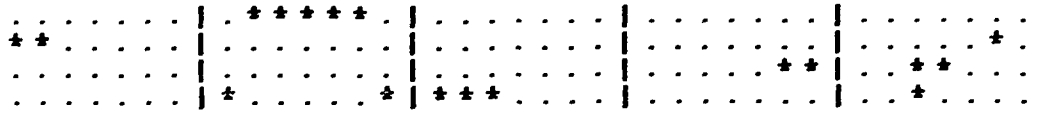
Iteration No. 21



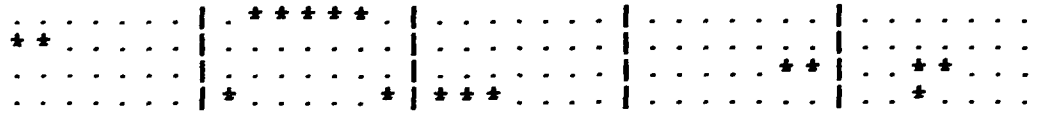
Iteration No. 22



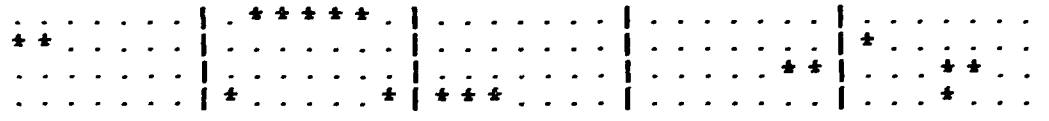
Iteration No. 23



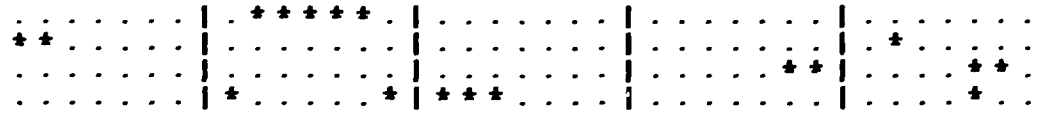
Iteration No. 24



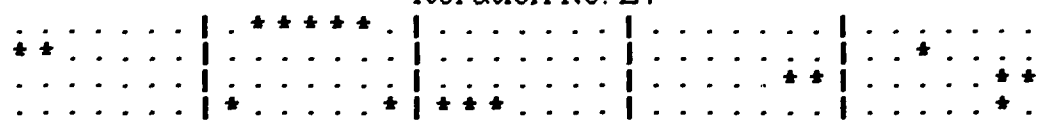
Iteration No. 25



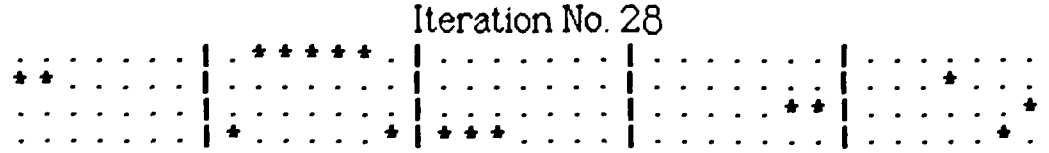
Iteration No. 26

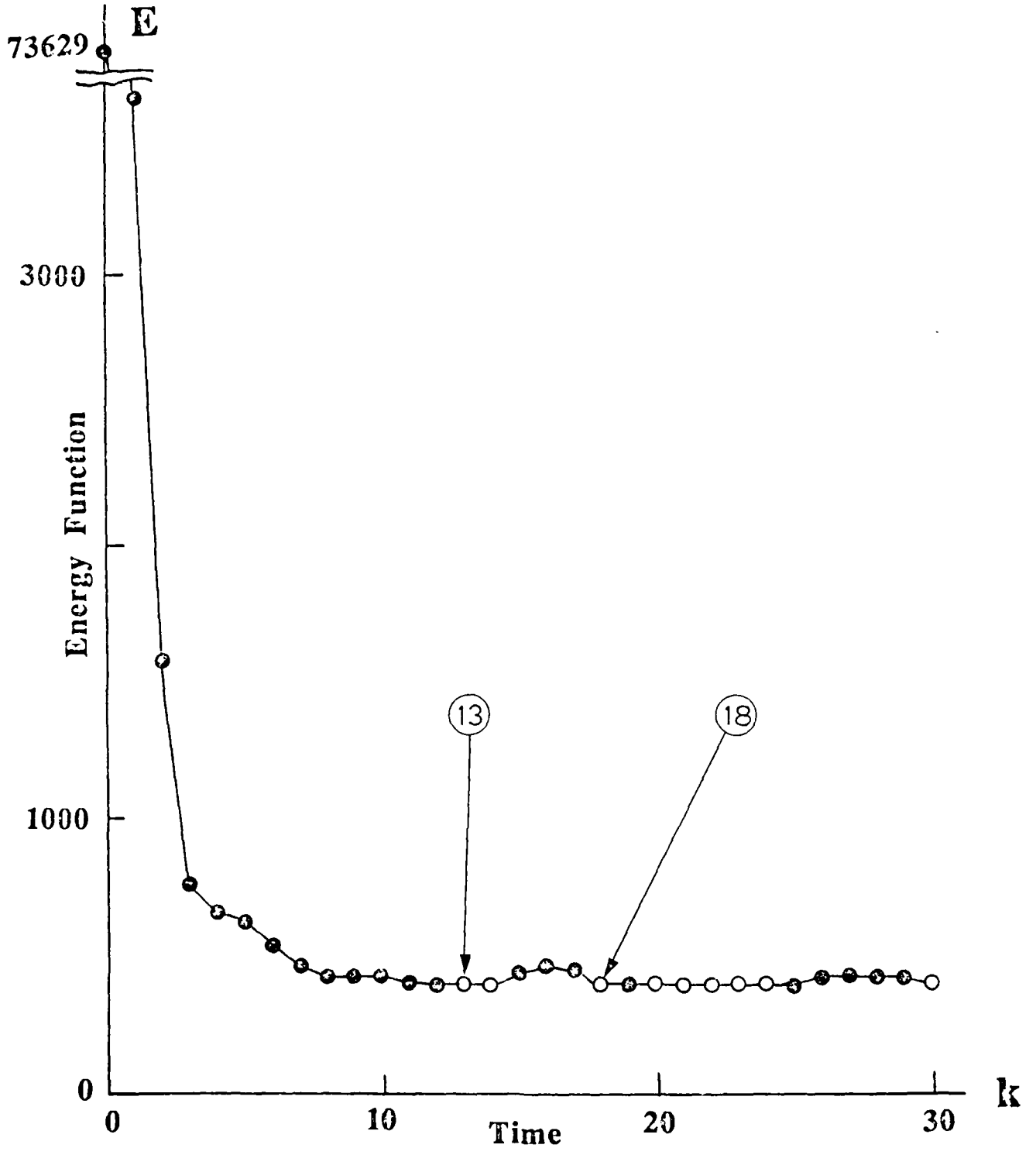


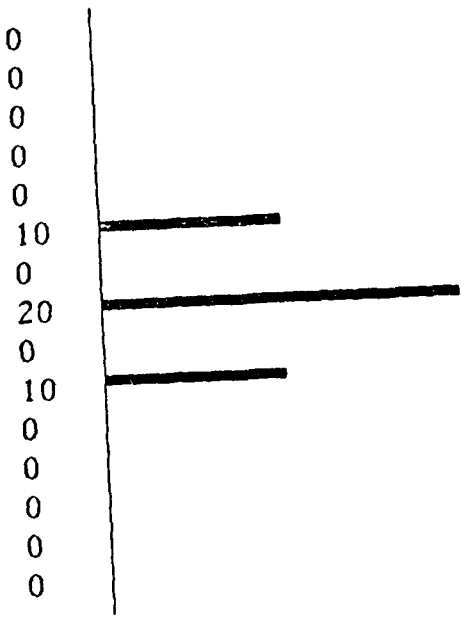
Iteration No. 27



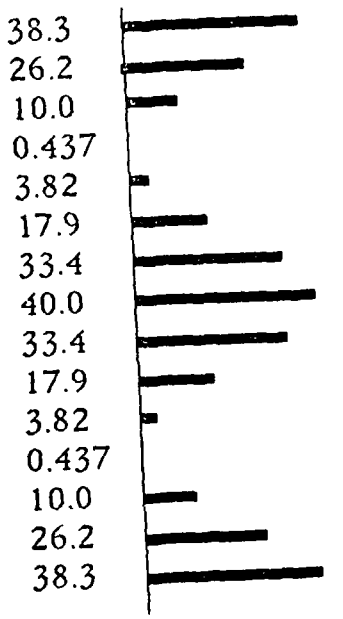
Iteration No. 28



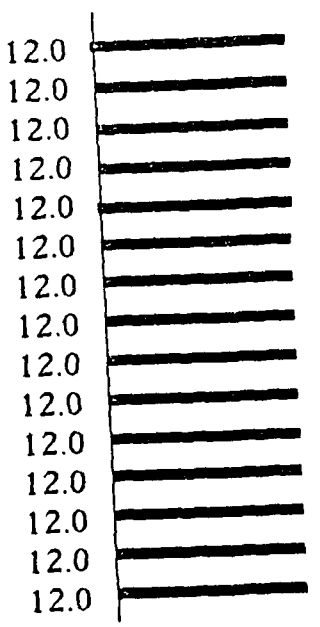

E
73629
3000
1000
0
Energy Function
13
18
0
10
20
30
Time
k


0
0
0
0
0
10
0
20
0
10
0
0
0
0
0

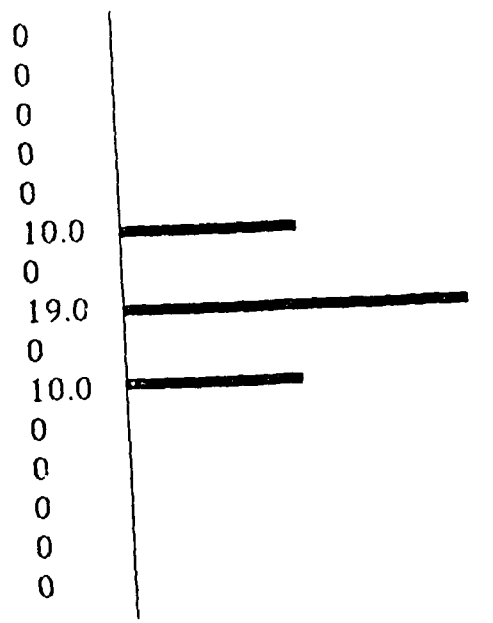
(a)


38.3
26.2
10.0
0.437
3.82
17.9
33.4
40.0
33.4
17.9
3.82
0.437
10.0
26.2
38.3

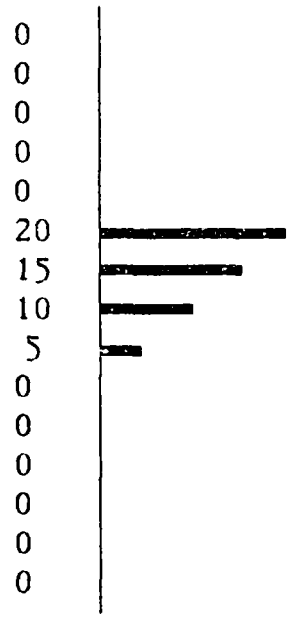
(b)


12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0

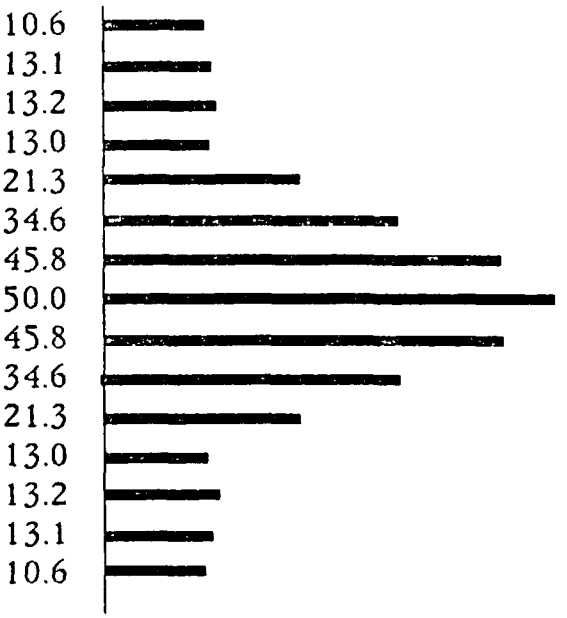
(c)


0
0
0
0
0
10.0
0
19.0
0
10.0
0
0
0
0
0

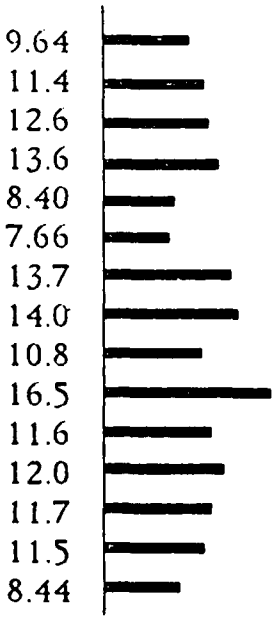
(d)


0
0
0
0
0
20
15
10
5
0
0
0
0
0
0

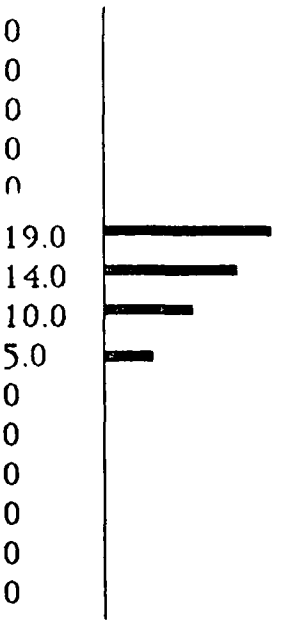
(a)


10.6
13.1
13.2
13.0
21.3
34.6
45.8
50.0
45.8
34.6
21.3
13.0
13.2
13.1
10.6

(b)


9.64
11.4
12.6
13.6
8.40
7.66
13.7
14.0
10.8
16.5
11.6
12.0
11.7
11.5
8.44

(c)


0
0
0
0
0
19.0
14.0
10.0
5.0
0
0
0
0
0
0

(d)

Figure Captions

1. Neural network model.
2. The Hitchcock Problem, with 4 sources and 5 demands.
3. Neural dynamics for the Hitchcock problem, using a binary number representation scheme. The initial states are randomly generated from a seed, and the transition mode is direct asynchronous. The final transportation matrix gives a network flow cost of 40. The constants in the energy function are chosen as A=27, B=C=80, and D=0.2. The constant x_0 is 0.5. See Table 3 for the flow matrices at the iteration numbers indicated by the arrows.
4. Flow cost histogram for the Hitchcock problem. The number of samples is 5000.
5. Second example of the Hitchcock problem using a binary number representation scheme. Uniformly fuzzy states initialized the network, and a "softer" non-linear function was used to give the best solution, with a flow cost of 49. The weights used were A=27, B=C=80, D=0.2. The constant x_0 was 1.0. The open circle represents a solution that satisfied the constraints. See Table 4 for flow matrices at the iteration numbers indicated by the arrows.
6. Network dynamics of the Hitchcock problem using a degenerate (simple sum) number representation scheme. The constants used were A=29, B=C=80, D=0.55, and x_0=0.1. Open circles again represent solutions that satisfy the constraints. Flow matrices corresponding to the arrows are found in Table 5.
7. Continuation of the degenerate network. One of the two-in-50,000 best solutions is found at time 21. Open circles represent solutions that satisfy the constraints (i.e. "consistent" solutions). The cost associated with the solution at the sixth

iteration is 43, that associated with the group of consistent solutions starting at iteration 10 is 40, and that associated with the remaining consistent solutions is 38. See Table 6 for the corresponding flow matrices.

8. Neural state transitions of the degenerate (simple sum) Hitchcock network (Figs. 6 and 7). Iterations 21 through 28 are shown.

9. Second example of the degenerate (simple sum) Hitchcock network. A random initial state drove this network to find both of the best solutions. The two flow matrices are shown in Table 7.

10. Inverse DFT. The transition mode is differential asynchronous. (a) Unknown signal. (b) Known Fourier transform. (c) Uniformly fuzzy initial states. (d) Estimated signal after 2 interations.

11. Inverse DFT, second example. (a) Unknown asymmetric signal. (b) Known Fourier transform. (c) Random initial states. (d) Estimated signal after 2 iterations.

Table Captions

1. (a) Cost matrix for the Hitchcock problem. (b) Sample solution depicting the flow from source X to demand Y.

2. Neural representation of the flow matrix for the Hitchcock network flow problem. q neurons are used to represent one element of the flow matrix.

3. Flow matrices for the specified numbers of iterations,corresponding to the points indicated on Fig. 3.

4. Flow matrices for the specified numbers of iterations, corresponding to points indicated on Fig. 5.

5. Flow matrices for the specified numbers of iterations, corresponding to the points

Cost Matrix, $C_{X,Y}$					
	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$
$X=1$	5	1	7	3	3
$X=2$	2	3	6	9	5
$X=3$	6	4	3	1	4
$X=4$	3	2	2	2	4

Flow Matrix, $f_{X,Y}$						
		D1	D2	D3	D4	D5
		2	7	3	2	4
S1	5	0	5	0	0	0
S2	3	2	1	0	0	0
S3	4	0	0	0	2	2
S4	6	0	1	3	0	2

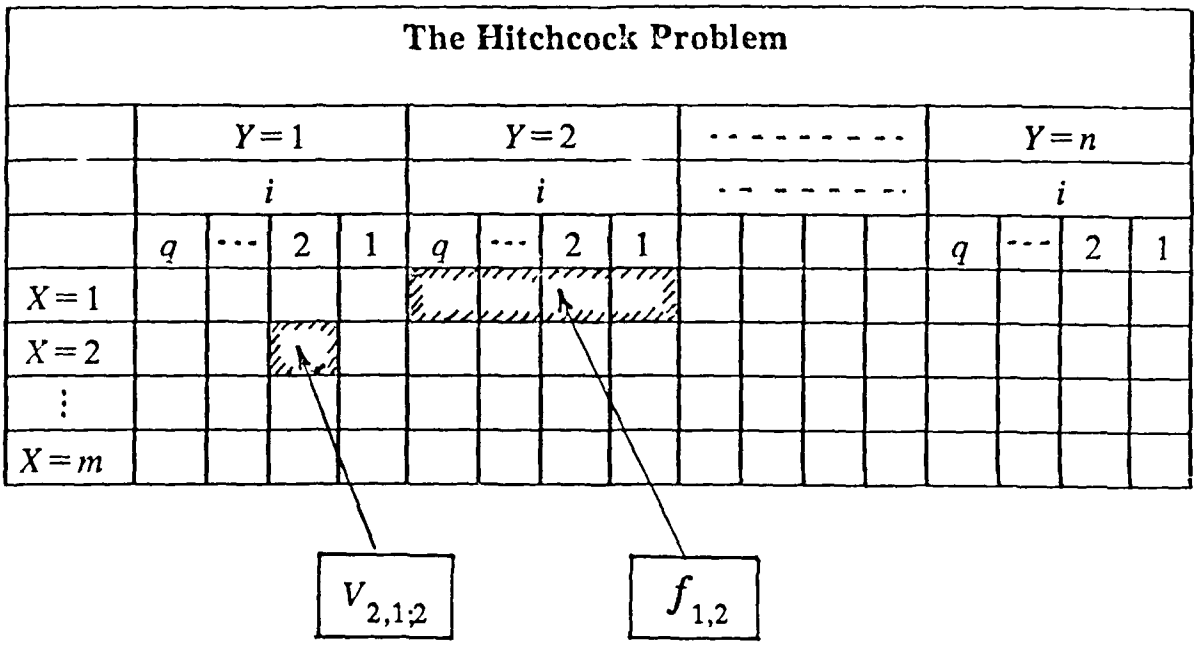


Table 2

TAKEDA & GOODMAN

No.0		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	3.6	5.2	6.3	4.6	2.8
S2	3.0	1.2	3.4	1.4	5.3	2.4
S3	4.0	2.9	3.4	5.8	4.0	2.5
S4	6.0	5.5	1.6	5.9	3.0	1.3

(a)

No. 1		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	0	0	0	0
S2	3.0	0	0	0	0	0
S3	0	0	0	0	0	5.0
S4	6.0	0	1.0	3.0	0	2.0

(b)

No. 2		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	3.0	7.0	0	0	0
S2	3.0	1.0	1.0	1.0	0	0
S3	4.0	0	0	0	0	4.0
S4	6.0	1.0	1.0	3.0	2.0	0

(c)

No. 4		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	1.0	5.0	0	0	0
S2	3.0	1.0	1.0	0	0	0
S3	4.0	0	0	0	0	4.0
S4	6.0	0	1.0	3.0	2.0	0

(d)

No. 6		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	5.0	0	0	0
S2	3.0	2.0	1.0	0	0	0
S3	4.0	0	0	0	0	4.0
S4	6.0	0	1.0	3.0	2.0	0

(e)

No. 0		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	3.5	3.5	3.5	3.5	3.5
S2	3.0	3.5	3.5	3.5	3.5	3.5
S3	4.0	3.5	3.5	3.5	3.5	3.5
S4	6.0	3.5	3.5	3.5	3.5	3.5

(a)

No. 7		D1 2.0	D2 7.0	D3 3.0	D4 4.0	D5 4.0
S1	5.0	0	5.0	0	0	0
S2	3.0	2.0	0	1.0	0	0
S3	4.0	0	1.0	1.0	2.0	0
S4	6.0	0	1.0	1.0	0	4.0

(b)

No. 0		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	3.5	3.5	3.5	3.5	3.5
S2	3.0	3.5	3.5	3.5	3.5	3.5
S3	4.0	3.5	3.5	3.5	3.5	3.5
S4	6.0	3.5	3.5	3.5	3.5	3.5

(a)

No. 1		D1 2.0	D2 7.0	D3 3.0	D4 4.0	D5 4.0
S1	5.0	0	0	0	0	0
S2	3.0	0	0	0	0	0
S3	4.0	0	0	0	0	2.0
S4	6.0	0	1.0	0	2.0	2.0

(b)

No. 2		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	3.0	4.0	0	0	0
S2	3.0	1.0	2.0	1.0	0	0
S3	4.0	0	0	1.0	0	1.0
S4	6.0	0	1.0	1.0	1.0	2.0

(c)

No. 3		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	4.0	0	1.0	0
S2	3.0	1.0	1.0	0	0	0
S3	4.0	1.0	1.0	0	0	1.0
S4	6.0	0	1.0	2.0	1.0	2.0

(d)

No. 6		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	4.0	0	1.0	0
S2	3.0	2.0	1.0	0	0	0
S3	4.0	0	1.0	0	1.0	2.0
S4	6.0	0	1.0	3.0	0	2.0

(a)

No. 10		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	4.0	0	0	1.0
S2	3.0	2.0	1.0	0	0	0
S3	4.0	0	1.0	0	1.0	2.0
S4	6.0	0	1.0	3.0	0	2.0

(b)

No. 14		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	4.0	0	0	1.0
S2	3.0	2.0	0	0	0	1.0
S3	4.0	0	0	0	3.0	0
S4	6.0	0	2.0	3.0	0	1.0

(c)

No. 21		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	5.0	0	0	0
S2	3.0	2.0	0	0	0	1.0
S3	4.0	0	0	0	2.0	2.0
S4	6.0	0	2.0	3.0	0	1.0

(d)

No. 13		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	4.0	0	0	1.0
S2	3.0	2.0	1.0	0	0	0
S3	4.0	0	0	0	2.0	2.0
S4	6.0	0	2.0	3.0	0	1.0

(a)

No. 18		D1 2.0	D2 7.0	D3 3.0	D4 2.0	D5 4.0
S1	5.0	0	5.0	0	0	0
S2	3.0	2.0	0	0	0	1.0
S3	4.0	0	0	0	2.0	2.0
S4	6.0	0	2.0	3.0	0	1.0

(b)


FILMED
2-8


DATE
FILMED
2-8