WAFER-SCALE INTEGRATION OF SEMICONDUCTOR MEMORY

A thesis submitted by
Russell Croston AUBUSSON, BSc
in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy
of the Council for National Academic Awards

April 1979

Sponsoring Establishment:-
Middlesex Polytechnic.

WAFER-SCALE INTEGRATION OF SEMICONDUCTOR MEMORY.

by Russell Croston Aubusson.

ABSTRACT

This work is directed towards a study of full-slice - or "wafer-scale integrated" - semiconductor memory. Previous approaches to full slice technology are studied and critically compared. It is shown that a fault-tolerant, fixed-interconnection approach offers many advantages; such a technique forms the basis of the experimental work. The disadvantages of the conventional technology are reviewed to illustrate the potential improvements in cost, packing density and reliability obtainable with wafer-scale integration (W.S.I).

Iterative chip arrays are modelled by a pseudorandom fault distribution; algorithms to control the linking of adjacent good chips into linear chains are proposed and investigated by computer simulation. It is demonstrated that long chains may be produced at practicable yield levels. The on-chip control circuitry and the external control electronics required to implement one particular algorithm are described in relation to a TTL simulation of an array of 4 X 4 integrated circuit chips. A layout of the on-chip control logic is shown to require (in 4ϕ dynamic MOS circuitry) an area equivalent to ~ 250 shift register stages - a reasonable overhead on large memories.

Structures are proposed to extend the fixed-interconnection, fault-tolerant concept to parallel/serial organised memory - covering RAM, ROM and Associative Memory applications requiring up to ~ 2M bits of storage. Potential problem areas in implementing W.S.I are discussed and it is concluded that current technology is capable of manufacturing such devices. A detailed cost comparison of the conventional and W.S.I approaches to large serial memories illustrates the potential savings available with wafer-scale integration.

The problem of gaining industrial acceptance for W.S.I is discussed in relation to known and anticipated views of new technology. The thesis concludes with suggestions for further work in the general field of wafer-scale integration.

ACKNOWLEDGEMENTS

This work was supported during the period May 1975 - October 1977 by ACTP contract K/A72b/383; it is a pleasure to recall the interest of David Clayden, Ted Newman, Jim Sherlock and Richard Wolski during this period.

Thanks are due to many people in the industry for enlightening discussions on specific detail of 4ϕ MOS circuit design and more general aspects of wafer-scale integration. Several colleagues within Middlesex Polytechnic contributed to the project; I would like in particular to acknowledge the invaluable assistance of Dick Gledhill in developing the programs for spiral simulation and to thank the technican staff of the Electronics and Microelectronics laboratories for their help in the wiring of the 30 pc boards for the TTL simulation of the multiple-chip array. The skilled advice of Ivor Catt during this simulation was also much appreciated.

It is a pleasure to acknowledge the continual help and able supervision of Dr John Butcher of Middlesex Polytechnic and Dr David Wilcock of the Allen Clark Research Centre of the Plessey Company.

I would also like to record sincere gratitude to my wife, Elizabeth, for her encouragement and the forbearance of my children, John, James and Susan, during the long evenings, weekends and Christmas holiday spent in self-imposed exile preparing this thesis.

Last, but by no means least, I would like to thank Sylvia Flood for typing this thesis.

CONTENTS

LIST OF ILLUSTRATIONS

GLOSSARY

As some of the terms used in this work have different interpretations in other fields, their meanings in this thesis are now defined.

CHIP. The circuit element step-and-repeated in the iterative W.S.I array - equivalent to the (unscribed) chip of the conventional technology.

CHIP A. The W.S.I chip of the serial memory approach of (10). It contains a serial shift register for data storage and the control logic required to detect and implement externally applied commands.

CHIP Z. The external control electronics which governs the generation of a chain of chips in (10).

FAILURE TOLERANCE. The capability of automatic reconfiguration as an apparently perfect device in the event of chip failure during operation.

FAST LINE. The data line which bypasses all serial shift registers in the chain of chips described in Section 3.4.2.

FAULT TOLERANCE. The capability of bypassing faulty chips during initial commissioning of the memory, but not having the ability ior automatic reconfiguration in the event of a chip failure during use.

GRACEFUL DEGRADATION. An attribute of failure-tolerant devices; after chip failure the memory will reconfigure to either its original specification or a downgraded function dependent on the number and/or distribution of redundant chips on the wafer.

SLOW LINE. The data line which passes through all "barrelling" shift registers in the chain of chips described in Section 3.4.2.

SOFTWIRING. The routing of signals by means of logic gating, thus enabling the selection of various data paths under external control when using a fixed-interconnection metallisation pattern.

SPIRAL. The linear chain of connected chips configured in the iterative array by Chip Z.

1. INTRODUCTION

Wafer-scale integration (W.S.I) is an alternative technology which may be applied to certain types of semiconductor device to create full-slice arrays. It is particularly suited to circuits having a high degree of regularity and which will benefit in the areas of cost, packing density and reliability from the considerable increase in device complexity which, as will be seen, are afforded by this approach. It will be shown that certain classes of semiconductor memory are ideal candidates for wafer-scale integration, both serial and parallel access structures being compatible with the technology. Although other potential application areas such as analogue to digital converters are identified, the present work is restricted in scope to a consideration of the wafer-scale integration of semiconductor memory.

The work described in this thesis commenced in May 1975 as a Department of Industry ACTP (Advanced Computer Technology Project) contract to investigate the technical feasibility and commercial viability of a novel type of semiconductor memory. This concept, first described by Catt[10], represented a new approach to full-slice technology. The essential feature of this invention - more fully described in Chapter 3 - was the capability of linking together, under external control, many of the good chips on a wafer without requiring a customised interconnection layer and without any prior knowledge of which chips were good and which were faulty.

The experimental work of Chapter 4 and the computer simulations of Chapter 5 developed from the outline in the patent specification[10]. Catt was regularly consulted especially in the early stages. Several potential problem areas had been suggested during his discussions with workers in the integrated circuit industry and elsewhere; these were to be investigated, together with any others which might be identified, as part of the assessment of technical feasibility. They are examined in Chapter 6.

Once the basic concept of the invention had been grasped it was then possible to relate this design to earlier techniques aimed at selectively interconnecting chips to form very large memory structures on a silicon wafer. These were studied as part of the literature survey of Chapter 2 and it was then possible to establish the general principles of wafer-scale integration, as developed in Section 3.3. It will be seen that certain structures based on a fault-tolerant, fixed interconnection approach to W.S.I. offer the advantage of being reconfigurable as apparently perfect devices in the event of limited failure of part of their structure during operation. The serial memory here investigated has this property.

The major problem associated with this study was known from the outset; the likely prohibitive cost of developing a prototype package and assembly techniques to house any completed wafer-scale integrated devices would almost certainly preclude the manufacture of full-scale working samples. The assessment of technical feasibility was therefore directed towards a study of the likely problem areas of this design and full slice technology generally, reinforcing this with experimental work where practicable and appropriate. The major areas of practical work in this respect have been to verify a proposed logic design, to assess the problems of implementing this as on-slice control circuitry and to propose algorithms and investigate their efficiency and yield requirements for the interconnection of the good chips on flawed wafers. Aspects of the work which have, of necessity, remained more theoretical in nature include an assessment of commercial viability of the proposed arrays, a design study of suitable packages and a consideration of potential problem areas (perhaps new to the technology) specific to fault-tolerant, fixed interconnection approaches to Wafer-Scale integration.

2. LITERATURE SURVEY

This chapter commences with a brief survey of fundamental limits on W.S.I technology imposed by gate power dissipation. The rest of the chapter is directed specifically towards a study of approaches to full slice technology.

Other aspects of the general field of memory devices are discussed at appropriate points in this work; these studies are by no means an exhaustive treatment but are selective with relevance to the main theme of this thesis. No attempt has been made either to compare the many technologies available for semiconductor memory devices or to consider in detail the suitability of each for wafer-scale integration. Each major technology now has its own extensive literature - for example the 268 papers on charge-coupled devices listed by Agajanian[1]. The major technologies available prior to 1972 are discussed (with reprints of selected papers) by Hodges[65]. Wilcock[64] has recently reviewed the currently available semiconductor memory device types.

2.1 POWER AND SPEED LIMITS FOR W.S.I.

Techniques for removing heat dissipated in W.S.I. memories are discussed in Section 6.7.4. It will be seen that, in view of the large size (> 1M bit) anticipated for W.S.I. arrays, it will be necessary to utilise a low power dissipation circuit technology. Rather than attempt to consider the power requirements and attendant noise limitations of the various technologies, this section attempts to indicate the areas of importance and provide an entry point to this literature.

Substantial improvements of circuit technology - e.g. the advent of Integrated Injection logic (I^2L) and Complementary Metal-Oxide-Semiconductor logic (CMOS) have reduced gate sizes and power-delay products to levels which, while still generally far short of fundamental limits, represent vast advances on the mass-production capabilities of a decade ago. Stein[52] notes that the energy required per logical operation has been reduced over the last 25 years by nine orders of magnitude. Fig.2.1, reproduced from this paper, shows that the best "minimum energy per logical operation" achievable with current CMOS designs is still more than two orders of magnitude greater than the limitation imposed by an (arbitrary) maximum permissible noise-induced error rate of 10^{-19} (1 error per year with 300 elementary operations per nanosecond).

This diagram also illustrates the drastic increase in power delay product as gate delays are reduced to ~1ns in practical devices. This suggests that ultra-fast W.S.I RAMs of multi-megabit capacity will require quite impracticable power levels with existing circuit technology. This problem is further exemplified by the characteristics of 10ns ECL RAM's predicted for 1981 by Herndon et.al.[23] Here it is suggested that 4096-bit ECL RAMs of < 10ns access time will be available on 11,000 sq.thou. chips of 1W dissipation. 500 such chips could be formed on a 3" slice, providing up to ~ 2M bits of memory, but, unless the power per bit can be reduced on going to W.S.I, the wafer power consumption (minimum of 500W) would require the input of 100A at 5V.

Power supply and thermal dissipation problems are discussed in Section 6.7.

Other fundamental physical limitations in integrated circuit technology of relevance to W.S.I are discussed by Wallmark[62].

2.2 REVIEW OF FULL-SLICE TECHNIQUES.

The general field of full slice technology has an extensive literature, but the vast majority of this relates to the restricted area of programmed interconnection procedures and, in particular, to the specific techniques of discretionary wiring and fusible links as approaches to this end. Within this general field of full slice technology the types of structure investigated have varied from pure memory, through logic-in-memory, to programmable logic arrays where the characteristics of the cell may be radically changed to suit a wide range of applications by suitable choice of external control signals.

These general aspects of full slice arrays have been extensively studied by Shoup[50] (48 references) who considers the relationship between the various structures proposed prior to 1970, by Seth[49] (15 references) (1970) who discusses the problems of test and fault diagnosis of such arrays and by Manning[34] (62 references, 1975) whose work is discussed in greater detail in Section 2.2.4.

In view of the existence of these excellent reviews this survey does not review the whole field in detail but concentrates on those publications of especial relevance to the present work and, in particular, on the literature published since 1975. Not all the techniques described in this section have yet been applied to full slice technology, but all are of relevance to W.S.I.

2.2.1 Discretionary Wiring.

The principle of "discretionary wiring" is the selective interconnection, by means of a second level custom-designed metallisation pattern, of functional elements on the wafer. The position of these is determined by prior testing and it is assumed that all such elements will survive the subsequent metallisation process which, in turn, will be flawless. The elements have ranged in complexity from single gates[8] to large-scale integrated chips[4].

The most complete account of the formative work on the technique of discretionary wiring is probably contained in (57-59). However, as these reports are not readily available - a full set could not be obtained even from DRIC* - the technique will now be described with reference to more generally available literature, most of which dates from the period 1966-69.

Early work on discretionary wiring was directed towards the connection of individual gates (e.g. (8,28)). The high complexity of testing and interconnection techniques for this approach are well illustrated by (8) in particular. The technique was later extended to the discretionary interconnection of more complex cells or circuits. Tammaru and Angell[54] considered various approaches to discretionary

* Defence Research Information Centre, St Mary Cray, Kent.

wiring and analysed the yield requirements of such redundant arrays. For an array of N + S elements, where N is the number of needed elements and S the number of spare elements, the probability that the array is flawless is quoted as

$$P_A = \sum_{j=o}^{S} \binom{N+S}{j} P_E^{(N+S-j)} (1-P_E)^j$$

where $\binom{N+S}{j} \triangleq \frac{(N+S)!}{(N+S-j)!\ j!}$, the number of possible combinations of N + S items taken j at a time and P_E is each cell's probability of working. This analysis is then extended to the yield improvement factor available in "Repair Processing", taking note of the additional defects arising during the second level metallisation.

In the late 1960's it was by no means clear whether to develop such techniques for avoiding faulty devices or to attempt to increase practicable chip sizes by reducing the defect density. Many workers recognised the serious limitations of Discretionary Wiring from the outset. Dingwall[13], for example, notes

> "Because of the need to allow space for probing pads (required in electrical testing) and potential wiring paths, high levels of compactness are not as readily obtained with this method ..... Moreover, the remaining technically significant portion of the multilevel process wherein gates are interconnected, packaged, and tested require 100 percent yields."

These problems are discussed further in Section 3.3.

Perhaps the most recent (and most successful) application of discretionary wiring to Wafer-scale integration is the "Functional Wafer" approach of IBM[4]. In this structure a matrix of 8 x 6 N-channel MOSFET array chips, each of 2048 bits, was formed on a 2¼" wafer. The chips were individually tested after first level metallisation and a programmed via hole mask ensured that only good chips were connected to the global wiring. Defective chips were then "repaired" by "flip-chipping" a mirror image chip directly over the (disconnected) faulty chip. A printed circuit board assembly technique was used, forced air cooling being adequate to remove the 5-7 watts dissipated by the wafer. In addition to the N-channel MOSFET array a set of bipolar support chips were flip-chip joined on top of the wafer "thus forming the complete functional entity of a 8k x 9 writeable control store of the IBM system 370/125".

A photograph of the completed assembly is reproduced in Fig. 2.2.

2.2.2 Post-Programmed Interconnection Techniques.

It will be shown in Section 3.3 that many of the disadvantages of discretionary wiring can be overcome by postponing the customisation stage until after metallisation is completed: Techniques for such "post-programming" of the interconnection pattern to isolate faulty elements from an array are now discussed.

Post-programming is usually provided in the form of fusible links which can be selectively blown to isolate elements or blocks from an array; other techniques include laser evaporation of metal and mechanical scribing.

Although this technique has not yet been successfully applied to full slice technology it does represent a possible route to this goal and is discussed in this context.

Fusible links, often based on nichrome of carefully selected track width and thickness, may either be blown selectively during device test or may be blown automatically by parts of the structure attempting to draw excessive current. This latter approach is utilised in large V.H.F power transistors. To increase device yield such chips are provided with more emitters than are required for normal operation; faulty emitters will draw excessive currents which blow series nichrome fuses - thus achieving self-isolation. Read only Memory (ROM) devices may also be based on fusible links(e.g. Schroeder et al[48])although preprogramming by contact hole or metallisation mask modification is more suited to quantity production. Memory structures, however, use post-programming merely to define the required state in a perfect device rather than to isolate defects in their structure. Although these objectives and constraints are quite different, certain of the techniques may well be applicable to wafer-scale integration and so are now summarised.

Programmable ROM's (PROMs) often permit alterable customisation. The FAMOS device[16] (Floating gate Avalanche injection MOS), for example, permits a transistor to be held permanently conducting through the injection of carriers from an avalanching junction into the silicon dioxide where they may charge an isolated poly-silicon electrode. Charge leakage is negligible (over periods of years) as the electrode is completely surrounded by silicon dioxide so a permanent channel is induced in the under-lying substrate. This charge may be (nonselectively) dispersed by ultraviolet light and the memory subsequently reprogrammed.

Selective programming is possible in the EAROM (electrically alterable ROM) device first described by Wegener et al[63]. Here the charge is stored at the interface of two dielectrics - usually a layer of silicon nitride on top of a (thin) layer of silicon dioxide - and may be selectively erased and reprogrammed by suitable voltages applied to the single overlying gate electrode. The potential application of this technique to full slice technology is illustrated by the Honeywell "Superchip" device[6]. Here, a 1.1 x 1.2 inch chip contained 256 arrays of 256 bit p-channel silicon gate MOS shift registers divided into four groups of 64 arrays connected via a common bus; these were then connected into the superchip array by means of a FAMOS PROM. Working devices of ~ 40,000 operating bits were produced by early 1975.

The largest structure to which post-programmed interconnection is believed to have been applied to date is a 92,160 bit shift register[15], also by Honeywell. Good and bad blocks (each containing ten parallel 256 bit registers) were identified and the bad blocks disconnected from the power supply by laser-burned fuses. Each good block then had its shorted address line fuse burned out to put it in series with the other good blocks to produce a CCD shift register of up to 92,160 bits. A module containing ten such chips was also designed.

In principle the fault tolerant fixed interconnection route - to be discussed in Section 2.2.4 - can be achieved with on-slice reprogrammable memory of the FAMOS or EAROM types. However this technique is not yet known to have been applied to memory devices; the 64k bit RAM recently reported[36] from IBM, for example, is said to be based on fusible links. The chip actually contains over 66k bits before the 65,536 bit (2^{16}) pattern is burned in.

2.2.3 The 100% Yield Approach.

Apart from such sporadic re-investigation of the programmed interconnection technique as discussed in the previous two sections, the vast majority of manufacturers have followed the conventional fixed interconnection 100% yield route to larger chips. While this route cannot be expected to lead to 100% perfect full slice arrays the techniques of yield enhancement are nevertheless of direct application in wafer-scale integration - which must also benefit from improved yields.

Improvements have occurred across the entire field of integrated circuit technology; silicon quality, epitaxy, doping techniques, oxide defect density and interface charge, photoengraving (fewer pinholes and smaller dimensions), metallisation and assembly technology have all improved with greater understanding of materials and processes. Reliability (and yield) have benefited from advances in process monitoring equipment - the scanning electron microscope being perhaps worthy of special mention in this context. Advances in circuit technology have drastically reduced gate areas and power dissipation.

However, each of these areas represents many engineer years of development and it would be impossible to do justice here to the many thousands of papers which have been published in the area of yield improvement. A modern textbook on IC Engineering - e.g.[23] perhaps provides the best entry to this literature. Certain points of particular relevance to the present work are now noted.

A study of the literature of the latter part of the 1960's (the era of discretionary wiring) shows many companies pinning their faith for future LSI designs on improved processing technology; Dingwall and Herzog[14] proposed redundant stages in the processing - e.g. double-photo-engraving involving two masks of (different) random defect distributions to virtually eliminate oxide pinhole problems. Four basic techniques to improve gate yields were discussed by Dingwall[13]. These were

"(1) Use of minimum-area devices;

(2) Non-contact photolithographic methods to provide perfection in the photoresist operation;

(3) Extensive use of redundant processing sequences;

(4) Use of special-care handling and environmental controls to ensure a continuous quality product. "

The "Polycell" approach of Motorola, as described by Hazlett[22], was essentially a fixed interconnection 100% yield technology, the major advantage being in the use of computer assisted layout to reduce design costs rather than any yield improvement by programmed interconnection. The discretionary wiring and Polycell techniques are compared by Stern[B10].

To assist in minimising the chip area, multiple interconnection layers were to be used in both the RCA and Motorola approaches to more complex chips. The Polycell approach of (22) proposed the use of no less than four metallisation layers; such a procedure would have been uneconomic in production at this time owing to the lack of understanding of aluminium step coverage problems which, being beyond observation by optical microscopy, only became fully appreciated with the advent of the scanning electron microscope in the late 1960's. This obsessive idea that multilayer metallisation would provide the key to LSI is widespread in the literature. Even Stern[B10], normally providing a shrewd assessment of IC technology, states

> "The further development of multilayer metallisation techniques appears to hold the key toward eventual implementation of large-scale integration. As component sizes continue to decrease, full advantage of the increasing component density potential can be derived only if the interconnecting wiring patterns do not demand a significant portion of the available die space. Yet, increasing circuit complexity is accompanied by a corresponding increase in component and circuit interconnections. This problem can be surmounted only by stacking the interconnecting patterns in successive layers one atop the other, each one insulated from the one beneath by a layer of insulating material ...
>
> The possibility of up to four layers of metallisation has already been demonstrated in the laboratory."

He does, however, sound a cautionary note (ibid):-

> "It remains a considerable challenge, however, to convert this laboratory capability into a high yield production process."

Moving forward from this era, we find that yields have been raised dramatically over the past decade by simplification of process technology. Many processes, based on as few as three masks (e.g. the "Trimask" process), have been developed; the main objective has been to raise device yields by process simplification without reducing device performance and circuit versatility to unacceptable levels. For example the p-channel MOS process, requiring only four masks is often less suited to complex devices than complementary MOS which requires five to six masks and one to two additional diffusion schedules.

The development of the five mask bipolar CDI (Collector Diffusion Isolation) process has been achieved by Ferranti[19] to full VLSI compatibility such that the complex F100L microprocessor comprising ~ 1500 gates can be manufactured at a tolerable yield.

Ordered arrays generally require far simpler metallisation, resulting in higher packing densities. The current practicable limit in complexity for chips based on the 100% yield approach is probably ~ 64K bit memory (CCD technology).

2.2.4 Fault-Tolerant, Fixed Interconnection Techniques.

Although, as will be discussed in Section 3.3, the fault tolerant fixed interconnection approach appears to offer considerable advantages for wafer-scale integration and provides the basis for the present work, very little has been published on this approach. One procedure, described by Manning[33 34], is summarised later in this section.

Another structure, having some resemblance to Manning's "Twist Repair" procedure (to be presented later in this section) is described by Goldberg et.al.[18] In this array, illustrated in Fig.2.3, switching logic on the control lines permits faulty chips to be replaced by adjacent ones thus allowing the faulty locations to be shuffled towards the array edge. Three horizcntal control lines run along chip rows; all chips may be switched to either the line above or the line below their actual row. Similarly two vertical data lines permit chips to contact either the line to their left or the one to their right. The procedure required to reconfigure the array is summarised in relation to Fig. 2.3, reproduced from (18).

The normal setting of switches is such that chips serve the data lines to their right - this condition being represented by upper case letters. On switching to the data line to its left a chip is represented by lower case letters. Asterisks represent faulty chips and hyphens are good unused ones.

The single faulty chip on row D is readily replaced by causing all chips to its right to switch to the data line to their left (as indicated by the letters "d"). To cope with the three faults on row F the following procedure may be adopted :-

1) Switch all good F chips on the right of the faulty chips to their left hand data line, thus eliminating the right-hand faulty chip.

2) Use the two chips labelled "F" from row G to replace the remaining two faulty ones in row F and switch the G chips to the right of these to their left hand data line - as indicated by "g". This leaves one empty site on row G.

3) Use one chip ("G") from row H to fill this empty position and switch the chips to its right onto their left-hand data line - as shown by "h".

Although this procedure is highly efficient in terms of array usage and repairability at high yield levels - when the number of spare chips need not exceed the number of faulty ones - it becomes a problem to optimise the shuffling procedure at higher fault levels. Also the technique is unable to cope with large clusters of faults - for example the block of six flawed chips illustrated in Fig.2.3 cannot be repaired as no chip is available to correct the left-hand fault in row L. The array therefore suffers from the problem of having key chip positions which, for any particular fault distribution, will cause the array to be non-reconfigurable in the event of failure of that chip. It will be seen that this problem is eliminated with the parallel/serial array structure proposed in Section 6.4.1 as any array chip address may be routed to any redundant chip site.

Review of Related Work by Manning

This theoretical study[34] of full slice arrays, entitled "Automatic Test, Configuration and Repair of Cellular Arrays" was unknown to us until the publication of (33). It is of particular relevance to this present work for two reasons; firstly because it is directed towards fault tolerant approaches to full slice technology and secondly because it confirms the results of spiral generation on rectangular arrays developed in Section 5.4.

Manning has placed equal emphasis on a theoretical study of "arm", "tree" and "grid" structures (summarised in Fig. 2.4) while the present work concentrates mainly on a more complete investigation of the first of these (which is equivalent to the "spiral" of Chapter 5).

A precis of Manning's work - which runs to nearly 250 pages - cannot be attempted here. Instead, various extracts of particular relevance are noted with comment when appropriate. These are taken in the sequence presented in (34).

i) "In most flawed arrays with N percent flawed cells, $0 \leqslant N \leqslant 25$, this program embeds an arm machine containing (100-2.2N) percent of the total cells in the array. This performance can be slightly improved. When N is greater than approximately 35, only very small arm machines can be embedded in a checkerboard array."

This is compared with the present studies in Section 5.4.

ii) "The most difficult checkerboard array repair involves embedding a high-relcon machine whose essential network is a grid."

This statement is reinforced several times, in particular by the relation:-

"$ORE_G \leqslant ORE_A \leqslant ORE_T$ for any flaw pattern" where ORE ≡ "optimum repair efficiency" and G, A, T refer respectively to grid, arm and tree algorithms. It is discussed further in Section 6.6

iii) "Random-access and track-addressed sequential-access memories may be efficiently realized as tree machines in flawed arrays."

This statement is of particular importance because the major reason for not investigating tree structures in the present work is the apparent difficulty of organising a tree structure having an arbitrary number of branches of arbitrary length as either a serial shift register or a random access memory.

iv) "The checkerboard array's interconnection structure is highly compatible with the two-dimensional, step-and-repeat nature of IC production. Further more, this structure is relatively easy to understand and manipulate compared to, for instance, hexagonal two-dimensional structures."

It is noted in Section 3.4.4 how the hexagonal array may also be made compatible with step-and-repeat techniques. It has also been successfully manipulated in the computer simulations of Chapter 5.

v) In quoting Bell[5] Manning notes

"In contrast to technology, system design costs have risen; this shift is demonstrated by, for instance, the decreased emphasis on minimisation in logic design, but on the other hand, reliability, mass producibility, and maintainability are now the important design criteria."

This point has become increasingly important since 1972 and indicates strongly in favour of a fault tolerant WSI procedure.

vi) "The fact that the considerable difficulties inherent in this (Discretionary Wiring) approach were even attempted points up the desirability of high-yielding high-integration IC's."

This statement reinforces the comments of Section 3.3.

vii) "The closest precursor of our testing and repair approach seems to be < Kukreja 73 >[37]. "

This overlooks British Patent 1,377,859,

viii) "Very small arrays, such as the 100-cell arrays in our experiments, tend to have lower repair efficiencies and lower cutoff points; because a higher percentage of cells are edge cells. An edge is a barrier that restricts the growth of an arm."

A consideration of the effects of the array boundary led to the proposal of the toroidal array configuration of Section 5.3.

ix) Manning describes at length the problems of spiral branching. This is considered in detail in Section 6.7.3 where it is concluded that the control logic design investigated in Chapter 4 will reduce the probability of branching to an acceptably low level.

x) The possibility of multiple input chips and the siting of these away from the array edges to improve the chances of the first chip being good is discussed. The use of multiple input chips has also been considered in the present approach but here the first chip is away from the array edge anyway and it is believed that the i/o interface chip (discussed in Section 6.8.4) provides a better solution.

xi) "Hughes developed a 50-watt package for a 3" slice as part of the Navy's All Applications Digital Computer program; unfortunately we haven't learned any details about this yet."

This point is noted in relation to device assembly in Section 6.8.2.

xii) "For technologies that require power lines connecting many cells, increases in array size increase the probability of array-destroying power problems. The probability of a power bus being open-circuited can be made very small by making the bus wide. Layout care can lower the chance of shorts between a power bus and a signal line; most such shorts would probably not be catastrophic anyway. Nevertheless very large arrays should perhaps include protection devices in each cell or block of cells. This circuitry could cut a shorted, or even overheated, cell off from its power source, before the malfunction blew the power line's fuse or sucked down the power line. The protection devices could be a fuse, or could be semiconductor circuitry, such as common transistor - SCR protection circuitry."

"Another power-handling approach would make a cell's supply of power controllable by the cell's neighbors. For instance, any of a cell's neighbors could command that the cell's power supply be switched on or off. This could save power in an array, and reduce the danger of faulty cells, by channelling power only to the cells in an embedded machine. Indeed a "power arm" could be "grown" in parallel with a processing arm into an initially quiescent array of cells."

These and other forms of global power supply protection are considered in Section 6.7.1.

xiii) The next major section is devoted to the construction of grid sub-arrays on orthogonal grids. The major techniques are now briefly summarised.

"Simple Repair" embeds a grid structure as shown in Fig. 2.5. Even with the best possible distribution of N faults the resulting grid must contain m fewer rows and n fewer columns, where $mn \geqslant N$; the worst case distribution of N faults eliminates N rows and N columns, as shown in Fig. 2.5.

"Twist Repair" embeds a grid structure of the form illustrated in Fig. 2.6. It will be observed that this is always able to embed at least as large a grid as "Simple Repair".

"Blockoff" breaks the array up into segments in which "Twist Repair" then embeds grids. These segments may, or may not, be reconnected to form a larger grid. "Blockoff" is illustrated in Figs. 2.7 and 2.8.

Such grid embedding, it is noted:-

"involves using some good cells purely as links between essential neighbors."

Many good cells are not used even for this purpose (see, for example, Fig. 2.8).

xiv) It is shown that grid embedding computations are comparatively simple for slightly flawed and extremely flawed arrays. They become very complex however, at moderate yield levels owing to the vast number of possible routes to be tried. It is noted that

"The time to repair an array varies widely, even for a constant array-size and % flawed."

xv) In comparing his repair procedures, Manning notes: -

"For arrays with more than a few (approximately five) flaws, Twist Repair is far superior to Simple Repair."

By way of example he notes that, in an array of 625 cells containing 20 flawed cells Twist Repair embedded a 14 x 14 square grid whereas Simple Repair embedded a 4 x 4 grid in the same array. This result is surprising - surely a 5 x 5 grid must be possible with even a worst-case fault distribution?

xvi) In relation to "Blockoff" it is stated that:

> ".... there is a fairly predictable, optimum block-size for a given flaw density."

xvii) A vast computation time for optimisation of repair using "Blockoff" is noted and it is further stated:

> "We wish we could offer more experimental results from Repair. However, Repair's large computation-time demands have made further experiments unfeasible."

xviii) Suggestions for improvement of "Repair" performance include:

> ".... very large arrays might benefit by interconnection strips wider than one line between large blocks", and

> ".... perfect machines should probably be designed in a modular fashion, with relatively few communication paths between modules."

xix) In comparing applications for arm, tree and grid machines, Manning notes the latter are more suited to:

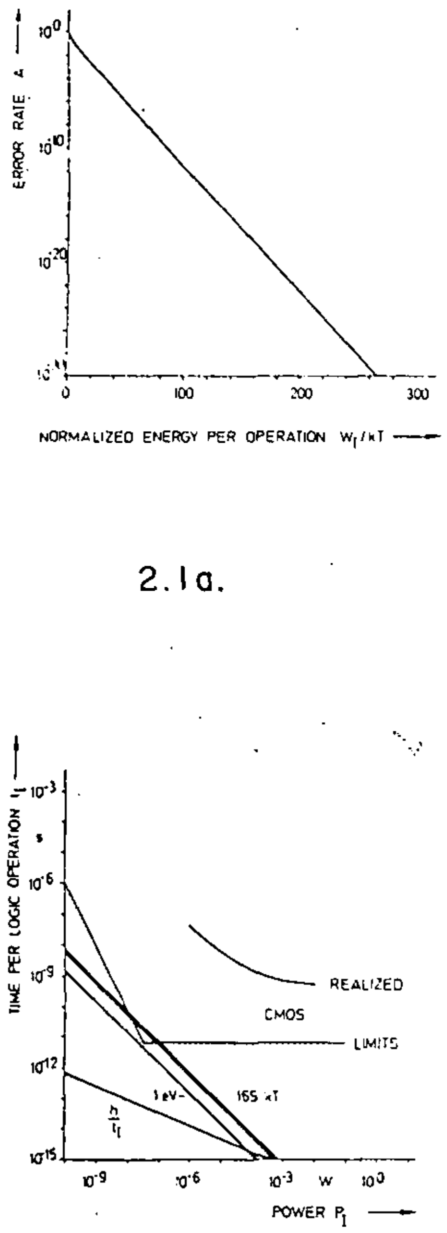
> ".... modules which communicate different information to three or more other modules at the same time."

xx) A tree organisation is suggested where

> ".... access time is minimised by placing a tree's base cell at the centre of a square region of cells; one diagonal of the square is a row, the other is a column, and the diagonals cross at the tree base cell."

Manning does not present a sketch for this but an interpretation fitting this description is given in Fig. 2.9.

To summarise, Manning's thesis presents many points which are in agreement with the ideas of the present work. However the majority of this work has diverged from that of Manning in several different ways, thus providing considerably greater coverage of the field of wafer-scale integration than would have been possible in a single work. The isolation of the two projects until a very late stage has ensured two virtually independent approaches to the subject although a prior knowledge of this work would have permitted a more advanced take-off point for the present study.



2.1a.

2.1b.

FIG.2.1. Dependence of error rate on the energy per logical operation. (2.1a)

Power input and time interval per logical operation of inverters - 165kT. Numerical example for a threshold given by the noise-induced error rate $A=10^{-19}$ (T=300K). Ref. 52.

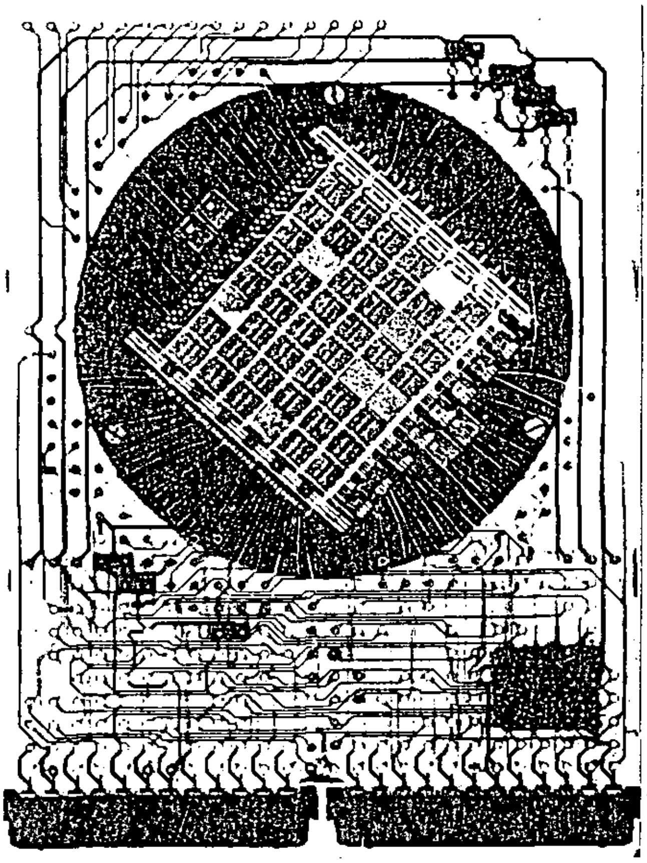
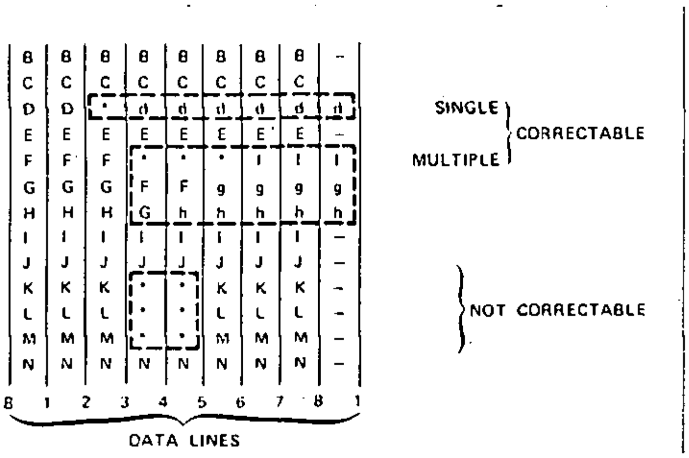
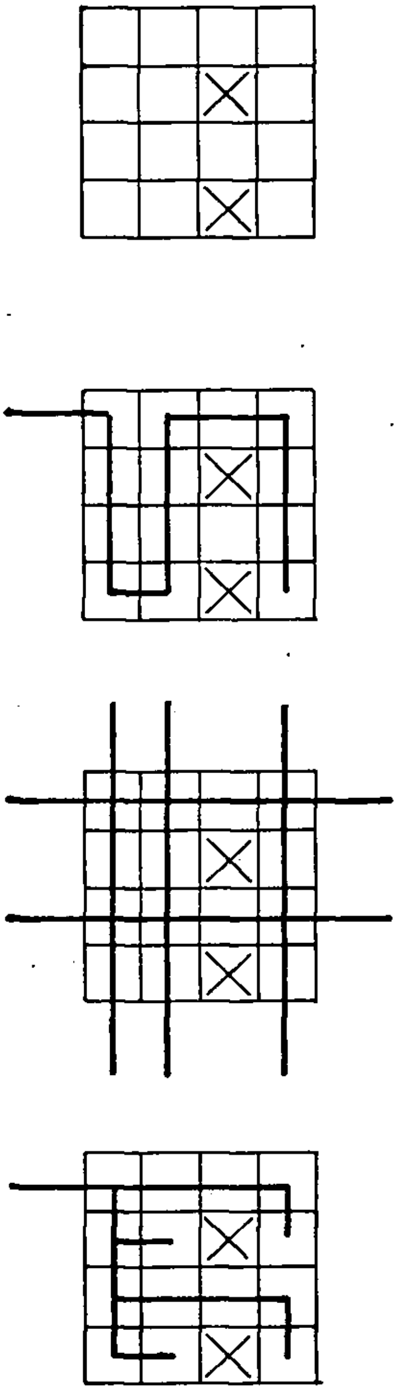


FIG.2.2. Functional Wafer on printed circuit card. Ref. 4.



B	B	B	B	B	B	B	B	–
C	C	C	C	C	C	C	C	
D	D	•	d	d	d	d	d	d
E	E	E	E	E	E	E	E	–
F	F	F	•	•	•	I	I	I
G	G	G	F	F	g	g	g	g
H	H	H	G	h	h	h	h	h
I	I	I	I	I	I	I	I	–
J	J	J	J	J	J	J	J	–
K	K	K	•	•	K	K	K	–
L	L	L	•	•	L	L	L	–
M	M	M	•	•	M	M	M	–
N	N	N	N	N	N	N	N	–

FIG.2.3. A partially fault-tolerant, fixed-interconnection array. Ref. 18.



Flaw pattern.

13-chip arm in flawed array. Repair efficiency (R.E.) is 13/14

6-chip grid in flawed array. R.E. is 6/14 (8 good chips serve only as transmission links.)

14-chip tree in flawed array. R.E. is 14/14

FIG.2.4. Arm, Tree and Grid arrays. Ref. 34.

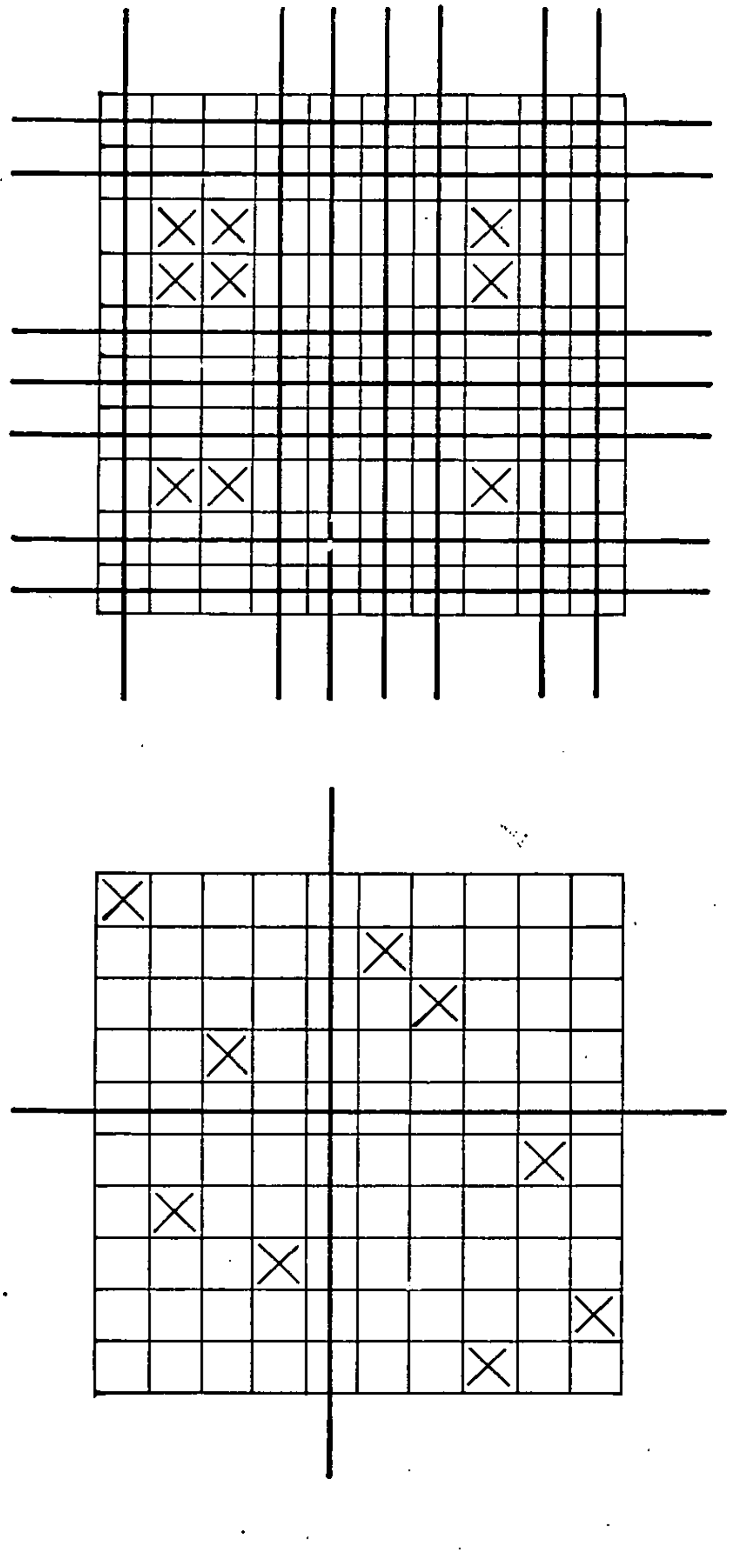


FIG. 2.5. Optimum and worst-case grids in 10 X 10 array having 9 flawed cells.

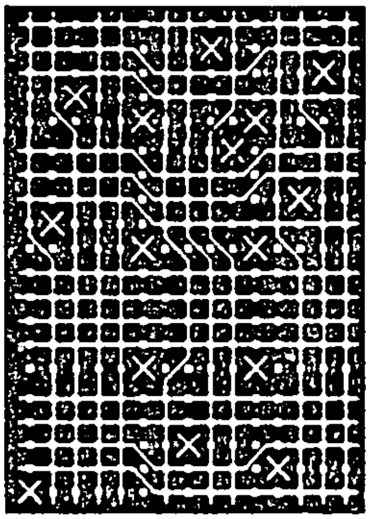
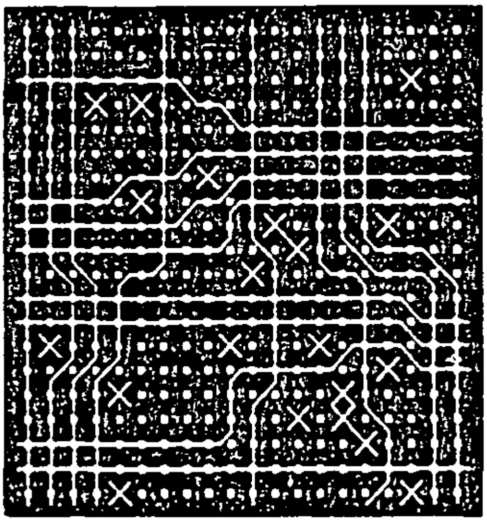
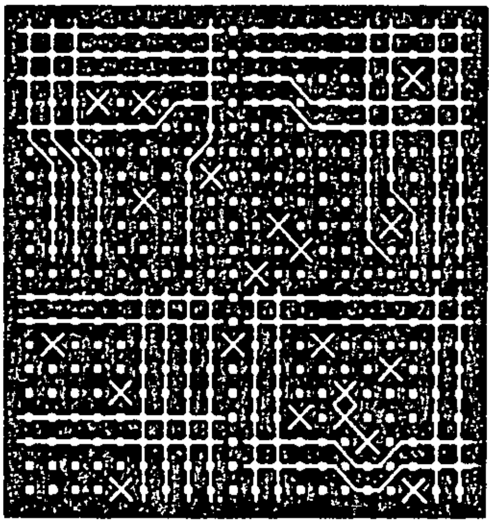


FIG. 2.6. Flawed 15 X 20 array twist-repaired into a perfect 10 X 14 array. Ref. 34.



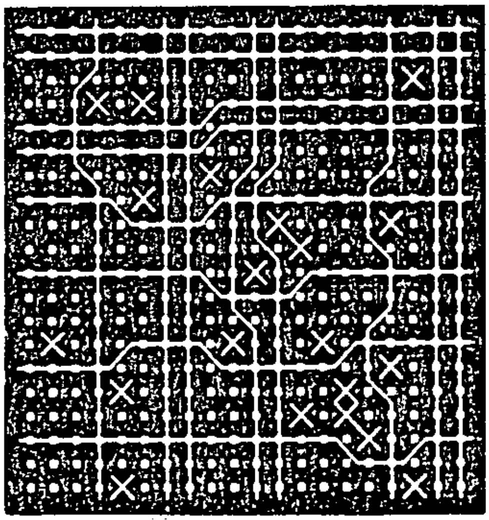
Array is divided into four 10 X 10 blocks. Each block is then twist-repaired to embed a 4 X 4 grid; the four grids are then interconnected to form one 8 X 8 grid.

FIG. 2.7a. "Blockoff's" repair of 20 X 20 array with 5% flawed cells. Ref. 34.



Four unconnected 4 X 4 grids are embedded in the array.

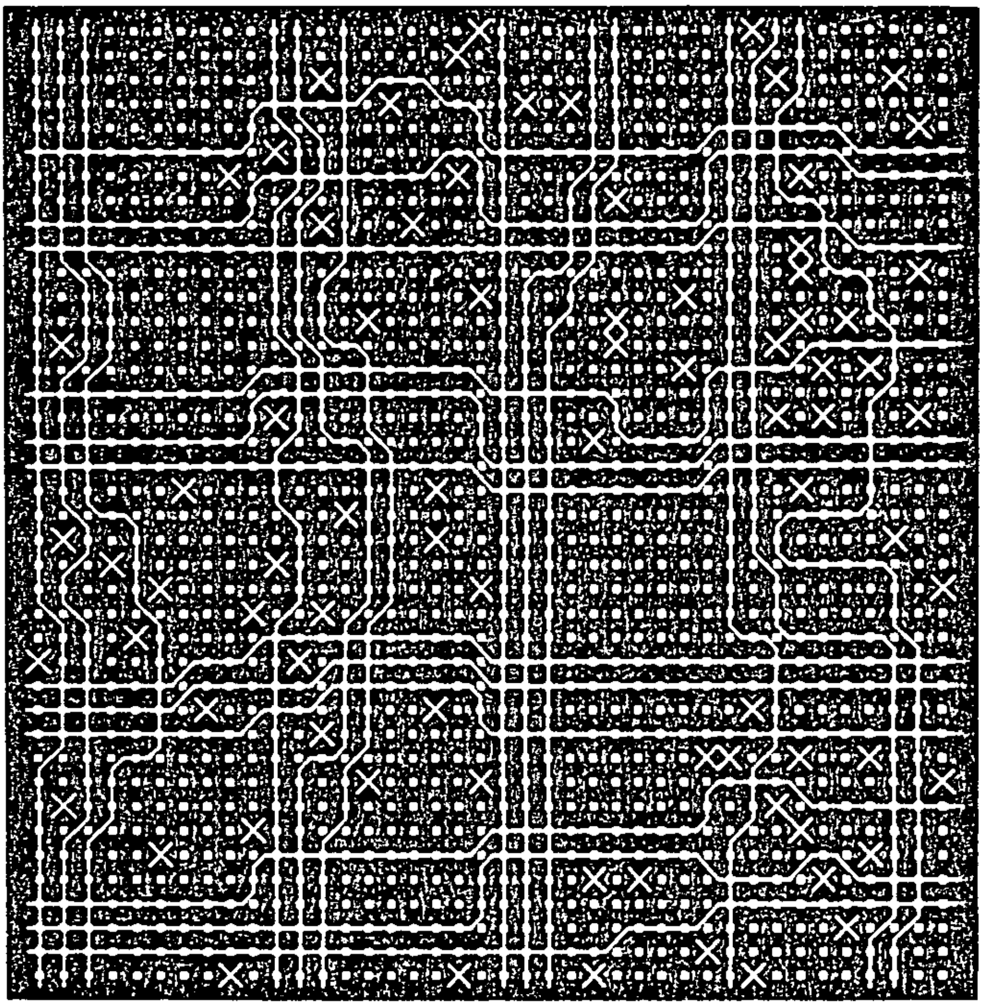
2.7b.



One 8 X 8 grid is embedded in the array.

2.7c.

FIG. 2.7b,c. "Blockoff's" repair of 20 X 20 array with 5% flawed cells. Ref. 34.



"10-connect algorithm" embeds sixteen 3 X 3 interconnected grids. Only 144 of the possible 1,520 locations are utilised.

FIG. 2.8. "Blockoff's" repair of a 40 X 40 array with 5% flawed cells. Ref. 34.

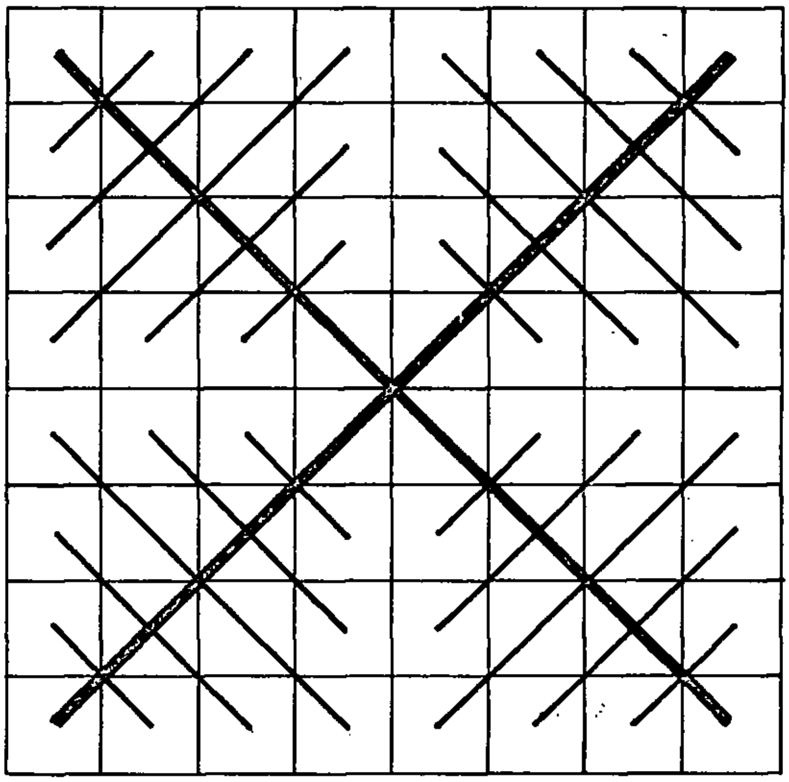


FIG. 2.9. An interpretation of the tree structure proposed in Ref. 34.

3. PHILOSOPHY OF WAFER-SCALE INTEGRATION

This Chapter commences with a brief, critical account of the procedures adopted in the conventional integrated circuit technology and continues with a discussion of the advantages to be gained by wafer-scale integration and a comparison of the various approaches to this goal. A detailed summary of the approach adopted in this work concludes the chapter.

3.1 THE CONVENTIONAL APPROACH TO IC MANUFACTURE.

In the conventional technology, the processed integrated circuit wafer, having the vast majority of its individual components (often > 99.99% of transistors, resistors, diodes, interconnections etc) in good working order and a reasonable proportion of its chips (each possibly containing over 10,000 such components) within functional specification, is subjected to a testing routine which selects those circuits which appear to function satisfactorily as a result of these limited tests and rejects those which do not. The "good" chips are then carefully nurtured through the scribing and separating process and passed for individual assembly after further close visual scrutiny. Each chip is treated individually, the alloying (or other attachment procedure) and bonding being carefully performed by skilled operators and the resulting lead-frame with its bonded device is again carefully inspected. Hermetic sealing or plastic encapsulation then follow and the devices are again tested to eliminate those which have failed to survive the (traumatic) assembly operation.

This committal to functional testing of all devices is continued by the device user who again tests the devices on receipt before assembling them (perhaps) onto a printed circuit board. The entire board is then tested; if it fails it may either be discarded or further time and money may be spent in attempting to locate and replace the faulty component. This procedure continues after the working board is assembled into a system; device failure is now still more costly, both in terms of replacement cost and in terms of system down-time. In one extreme case such failure could result in the loss of data obtained by a space probe. Manning[34] notes the cost of failure at various system development stages as quoted in Table 3.1. (Numbers indicate costs in US dollars).

Market	Incoming	Board Mount	System Test	Field Use
Consumer	2	5	5	50
Industrial	4	25	45	215
Military	7	50	120	1000
Space	15	75	300	200M

Table 3.1

Attempts are made to avoid this catastrophe of failure using built-in redundancy but there is still the underlying acceptance of the philosophy that devices shall be 100% tested at many strategic points during their manufacture and installation. This committal to 100% testing is a major factor in the dramatic cost escalation from the processed chip to the shipped device. The cost of assembled, tested devices exceeds by 2-3 orders of magnitude the cost of the processed untested wafer of equivalent gate complexity. Even with state of the art circuits the chip cost is still a relatively minor part of the device cost so manufacturers must continue to push chip sizes up towards the limit where chip yields tend to zero in an attempt to reduce assembly and testing costs by containing complete system functions on fewer chips. Although this is extremely inefficient in terms of usage of silicon (2-3 good chips from a possible 80 on the wafer is not uncommon for complex microprocessor chips) it does have the advantage of drastically reducing the complexity of printed circuit board interconnections and backplane wiring which are often serious limitations on system operating frequency. Sutherland and Mead[53] note that computer science has evolved from a technology in which wires were cheap and switching elements were expensive to one in which switching elements are virtually free and wires are the expensive component as they "occupy most of the space and consume most of the time". They then suggest that

> "As integrated-circuit technology progresses there will be individual circuits of increasing speed and complexity. No relief is in sight, however, for the costs and delays inherent in communicating information from one circuit to another."

This viewpoint is noted in the context of describing the potential benefits of distributed array processing. However it is now shown that W.S.I technology can offer substantial savings in wiring complexity.

3.2 ADVANTAGES OF WAFER-SCALE INTEGRATION

Consider for a moment the contacts and interconnections required to interface between two integrated circuits on adjacent printed circuit boards. From the output of the driving gate - say an n^+ diffused contact to the epitaxial layer defining the collector of a transistor - to the input of the second gate - perhaps a p-type transistor base region - the current path is tortuous indeed. It proceeds:- silicon - metallisation - bondwire - package lead - solder - printed circuit board - plated through holes - copper to gold at edge connector - gold wire (wrapped joint) and so in reverse to the other integrated circuit. This represents a total of ~ 20contacts - including intermetallics of doubtful reliability - between two gates.

The intermetallics between chip metallisation and package lead have caused notorious problems in the past - for example "purple plague" (an alloy of gold and aluminium) and "black death" (a structurally weak compound of gold aluminium and silicon). While these intermetallics are now largely understood and no longer a serious reliability hazard the major cause of IC failure is still (for a sound process and device layout) the chip to package interface. It is therefore advantageous to pack as much circuitry onto each IC as possible. Hodges[25] notes that

> ".... in semiconductor memories failure rates in field service are better correlated with the number of separately packaged chips in the system than with the total number of bits in the system stores."

While this point also operates in favour of increased chip complexity the advantages offered by W.S.I technology are far greater.

It is assumed that those aspects of device screening and environmental test (e.g. 20,000g test, leak test) in the conventional technology which are relevant to W.S.I will be carried over tofull-slice structures to ensure a product of equal integrity of construction. Techniques currently applied to conventional devices are noted in (61).

It should, however, be noted that one aspect of device screening prior to conventional assembly could not be readily carried through to W.S.I technology. Chips are normally visually inspected prior to attachment to headers and, again, after bonding. While the main purpose of this inspection is to eliminate poorly scribed and improperly bonded chips - neither of which is relevant to W.S.I - the operator has, in principle, the opportunity of a full visual inspection of each chip and the ability to reject any which do not meet preset requirements (e.g. minimum metal track width $\not< \frac{2}{3}$ of intended width).

The inclusion of such chips on W.S.I devices would cause some reduction in reliability. This, however, is a matter of degree only; to fully check a complex VLSI circuit for faults down to $\sim 2 \mu m$ requires several hours of inspection per chip. It is, in any case, possible in principle to eliminate such chips from automatically configured W.S.I memories by ensuring that they are inoperable - for example by scratching with metal probe (or, more reliably, using laser beam) to open-circuit the metallisation.

There are other reasons for increasing IC complexity to the practicable limits. Apart from the above reliability consideration other benefits accrue from (i) cost reduction arising from fewer packages and reduced assembly effort, (ii) increased packing density (gates per cubic inch of controlled environment) and (iii) improved performance - for example emitter coupled logic (ECL) benefits by having to drive fewer inter-package transmission lines with attendant propagation delay and power dissipation in line driver stages.

The limit of complexity may be defined (for example in subnanosecond ECL) by thermal dissipation problems but is most often determined by defect density. As ICs cannot be made at 100% yield there is an optimum chip size which depends on defect density. This, as noted earlier in Section 2.2.3, has increased as defect densities reduced with advances in technology. The logical conclusion for existing technology is for chip complexity to increase to the level where chip costs become comparable with assembly plus all other post-slice costs.

However if imperfect devices could be eliminated mechanically or electrically the chip size could become as large as a wafer, yield no longer being a dominant limitation.

3.3 APPROACHES TO WAFER-SCALE INTEGRATION.

The philosophy of W.S.I is to reduce chip complexity (and hence chip costs by a large factor) while simultaneously virtually eliminating assembly and specialised testing operations. This in itself will have little impact on the overall system cost - even if all the integrated circuits in a large computer were available at zero cost the total system cost would only fall by a few percent. However the major savings offered by reduced wiring complexity (and hence improved performance) and increased reliability - in addition to the graceful degradation attribute discussed in Section 6.5 - must represent a massive reduction in system assembly, installation and operating costs.

The potential cost reduction and improvements in packing density and reliability of full-slice technology have long been envisaged. R Petritz[41] stated in 1967:

> "We at Texas Instruments feel that the full potential of semiconductor technology for integrated electronics will be realized only when the entire semiconductor slice constitutes the packaged product."

This claim for the potential advantages of a full-slice technology,made at the height of the investigations into large array technology by the discretionary wiring of interchip connections (as surveyed in Section 2.2.1) remains equally true today. Developments in wafer-processing technology during the ensuing ten years and the continuing high costs of chip assembly have only served to increase the attractiveness of a full-slice technology.

"Discretionary wiring", however, proved not to be an economically viable way of achieving this aim. The basic idea of interconnecting the required number of chips on the wafer into a complex array using a second level, custom designed metallisation pattern suffered from several disadvantages in its implementation. Firstly, each wafer required a tailor-made mask, the layout of which could only be specified after probe testing the full wafer; this added a substantial overhead to the processed slice cost. Secondly, the requirement to test the wafer before completion of processing is undesirable as it must introduce some degree of contamination and damage to the first aluminium layer in addition to requiring each chip to carry a set of bonding pads (for probe test purposes) which will be redundant in the final discretionary-wired array. The major problem, however, was the implicit assumption that the processing of the second level, full wafer metallisation would be 100% perfect and that no chips tested as good would fail during the additional processing schedules. The problems of ensuring that all vias make adequate contact while no cross-overs short to the underlying metal are well-known to the industry. A vast amount of research effort was expended on discretionary wiring prior to 1969 when the technique appears to have been virtually abandoned in favour of the conventional discrete chip approach for which yields had substantially improved with advances in process technology.

The various approaches to wafer-scale integration are compared in Fig. 3.1. Post-programming of the final interconnection pattern (as described by Elmer[15] for example and summarised in Section 2.2.2) is certainly preferable to the pre-programming procedure of discretionary wiring in that it permits further levels of surgery on the wafer if the first attempt fails to achieve the desired result. It still suffers from the drawback, however, that each wafer must be treated individually in conjunction with elaborate and expensive testing procedures.

An analysis of the problems associated with these approaches led to the realisation that to minimise manufacturing costs in a production environment a full slice technology must postpone the decision as to which chips are to be included in the array until <u>all</u> the wafer processing is finally completed. Further, since all slices are best treated identically on a production line (customised treatment of each slice being a costly operation) it should not require a detailed knowledge of the actual fault distribution on any slice. The chosen technique may then use the same interconnection pattern to photoengrave the metal layer on all slices.

As shown in Fig. 3.1, such a Fixed Interconnection approach has only two possible routes; either the slice must be 100% perfect - an impossible target with today's process yields limiting chip dimensions to ~ 6 mm square - or the procedure must be fault tolerant. The concept described in (10) has this property; it is the first in a family of Fixed-Interconnection, Fault-Tolerant procedures for Full-Slice Technology.

The term "Wafer-Scale Integration" has been introduced for two reasons; firstly the term "full-slice technology" has come to be associated with the discretionary wiring approach - generally rejected by the industry as being uneconomic in implementation. Secondly, it is felt that the concept might be more readily accepted if given an acronym (W.S.I) alongside the generally accepted terms for levels of integration:-

Acronym	Title	Level of Complexity (Gates/monolithic chip)
S.S.I.	Small-scale integration	1-9
M.S.I.	Medium-scale integration	10-99
L.S.I.	Large-scale integration	100-999
V.L.S.I.	Very large-scale integration	1000-9999

Table 3.1(a)

Assuming that the superlatives continue by analogy with the frequency spectrum there will be no conflicting usage of "W.S.I" as we may expect

Acronym	Title	Level of Complexity (Gates/monolithic chip)
U.L.S.I.	Ultra large-scale integration	10,000-99,999
S.L.S.I.	Super large-scale integration	100,000-999,999
E.L.S.I.	Extreme large-scale integration	1M - ~ 10M

Table 3.1(b)

As current technology is capable of penetrating through the S.L.S.I band and well into the E.L.S.I range with a full slice, fault tolerant, fixed interconnection procedure it is logical to coin a generic name - hence Wafer-scale integration, or W.S.I.

Techniques which require configuration before manufacturing is complete cannot tolerate additional failures - e.g. discretionary wiring. Those which are configured after completion of processing may or may not be reconfigurable to bypass chips which develop faults. Fusible links and mechanically-scribed or laser-burned interconnection would provide very limited capability for reconfiguration. Techniques based on EAROM or RePROM structures (discussed in Section 2.2.2) would be considerably more versatile in this respect; it is arguable whether these techniques really belong in the second or fourth branch of the tree in Fig. 3.1. Those approaches which have a truly fixed interconnection pattern and select only those chips which are found to be operational at the time the device is configured offer the greatest degree of tolerance to failures occurring during use. It is a member of this family on which Chapters 4 and 5 are based; the concept is described in the next section.

3.4 THE W.S.I APPROACH OF THIS THESIS.

This section is devoted to a description of the particular fault tolerant fixed interconnection scheme of British Patent No. 1,377,859[10] which is central to this thesis.

3.4.1 Design Concept

Like the approaches previously discussed (references 18 and 34) this design uses the capability of linking together the good chips on a wafer without requiring any additional mask or even a prior knowledge of which chips on the wafer are good and which are faulty. To achieve this, each chip is given, under external control, the capability of linking to any one or two of its four nearest neighbours. Connections are made to the input and output of one chip on the wafer and to the power supply/clock grids supplying all chips on the wafer.

The creation of a chain of chips is now discussed in relation to Fig. 3.2. Let us suppose that the chip is a 1K bit shift register. A known bit pattern of 1K bits is fed into the chip and the output pattern is compared with the input to ensure that the memory is functioning correctly. This chip is then instructed to access the adjacent chip due East and 2K bits of data are fed into this two chip serial memory. If there is no corruption of the returned data then the second chip is known to be good also and is instructed to access the chip to the South. 3K bits of data are fed in and we will now suppose that an error is detected in the returned data, thus indicating that the third chip is faulty. The second chip is now instructed to close its links with the existing third chip and to access the chip to the East, which now becomes the third chip in the chain. In this way it is possible to build up long chains of good chips, thus producing very large serial memories.

It is advantageous to encourage a spiral structure in the developing chain so the process of constructing a chain of good chips is termed "forming the spiral".

Logic Design.

A design for the on-chip logic is included in the patent[10]. This, reproduced in Fig. 3.3, provided the basis for the design of Chapter 4.

Many different possibilities exist for the control of spiral formation. Some switching logic is obviously essential on each array chip (called "Chip A" by Catt). This has been kept to a minimum in order to minimise the area of silicon wasted (i.e. not available for memory) on each wafer. This minimisation of the on-chip data processing capability leads to an increase in the complexity of the external control electronics which Catt has called "Chip-Z". However since this is capable of controlling many wafers the increased complexity and cost of Chip-Z required by this approach is justified.

The detailed logic design of Chip-Z, as used in the hardware simulation of Chapter 4 is peripheral to the main theme of this work. It will not therefore be described in detail here; a block diagram is given in Fig.3.4, the logic design in Appendix I. Its function is implicit in the description of operation of the on-chip control logic which is now summarised.

All chips in the array are connected to a master grid or grids; these supply the power and signals from which the detailed clock waveforms required by the particular technology may be derived. In the absence of a spiral, all good chips on the wafer are "looking" in the same direction at any given time; that is, their interchip address register will cause only one of the four input gates - one each on the North, East, South and West edges of the chip - to be enabled to receive data at that instant. Similarly all good chips will be "open" to one direction; that is, their address registers will allow only one of the four output gates - one each on the North, East, South and West edges of the chip - to be enabled to output data at that instant. OPEN and LOOK directions on all chips will sequence through the cycle, N, E, S, W under the control of t_o, a timing signal derived from the input waveforms.

In order to feed data into the input/output chip it is essential that this chip is "looking" in the direction from which the data stream will appear. A reset signal, called Master Clear, is arranged to reset both OPEN and LOOK address registers on all chips to "OPEN N" and "LOOK N" so that the OPEN/LOOK direction of all free-running chips on the wafer is known at any time from the number of t_o pulses since the last Master Clear.

The principle of the design is to use part of the shift-register storage element itself to detect and decode commands by tapping off certain key bits near the beginning of the shift register. Only two commands are required in this implementation; the first, called "FREEZE", causes the chip addressed to cease following the cyclic N, E, S, W directions of all free-running chips and remain looking in the current direction until either power failure or a further Master Clear pulse. This will cause it to lock onto the upstream chip; FREEZE also causes the OPEN direction to be held in its current state. The second command called "STEP 90", causes the OPEN address register to disable the current OPEN direction (frozen during the previous FREEZE command) and access the next one in the

N,E,S,W, cycle. With reference to Fig. 3.3(b), both commands are represented by an all zeros word (to ensure that no spurious "1" bits are present in the shift register) followed by a "1" tben a "0". These two bits are redundant in data words, all of which must commence with "00". FREEZE contains only zeros in the rest of this new word while STEP 90 contains an extra "1" at a key point in the word. The appearance of this "1" at a predetermined point in the shift register during a command word triggers the STEP 90 routine.

The command must, in general, pass down the spiral through other chips to reach the intended chip at the end of the spiral; a further requirement of the on-chip control logic is therefore to ensure that only the intended chip implements the STEP 90 command. To achieve this, part of the command word is allocated to a field which contains a number equal to the number of chips currently in the spiral up to the chip to be addressed. Each chip subtracts one from this number so that by the time the word reaches the addressed chip it has been reduced to zero. The last chip attempts to subtract one from zero, causing an unrequited borrow to propagate up this field. The control logic detects this unrequited borrow and enables a gate which would otherwise inhibit the execution of the STEP 90 command.

3.4.2 Extension to Pseudo-Core Memory

As described so far and as investigated in this work the memory is essentially a large shift register from which data may only be read serially. As the commercial viability discussions of Section 6.8 assume a substantial demand for W.S.I devices it is important to note that this concept is not restricted to serial memory. The extension of this structure to pseudo-core memory as proposed in (10) is now outlined.

By running a parallel access line along the developing spiral as shown in Fig. 3.5 the memory becomes content addressable; as described in the patent it is then possible to undertake data processing within the memory. These ideas, which Catt believes will have far-reaching effects on the design of computers, have been presented verbally[11]. The general principle of this type of data processor is now summarised.

The spiral, when first configured, may be represented schematically as in Fig. 3.5(a), where each bulge on the slow line represents a shift register of perhaps 1K bit length and each junction J_1-J_N represents control logic capable of comparing an address code on the fast line with a tag on the data in that particular chip. When a match is found the word is switched onto the fast line and read from the memory, ensuring that neither it, nor the address, can cause any further transfers onto the fast line. If part of the data itself becomes the tag then listing operations are readily achieved by increasing each successive address code - for example the restructuring of a telephone directory in ascending telephone number order. This configuration of the shift register is termed "barrelling".

The inclusion of a limited processing capability at each junction $J_1 \ldots J_N$ (e.g. compare fields and add contents on match) considerably improves the computing power of the system, but at the expense of increased chip complexity.

We can arrange, using the control logic at $J_1 \ldots J_N$, to isolate each individual chip from the spiral so that the main data stream takes a short cut across such chips while the data in the individual shift registers at the time of this "looping" operation continues to recirculate until the loops are brought back into the main shift register path. This permits a second fast line through the array, as shown in Fig. 3.5(b).

If the last chip is instructed to barrell while the rest loop then the contents of this chip will pass all the other words in the shift register, enabling comparison of selected fields and possible logical operation with matching words. Conversely if the first chip in the shift register is instructed to loop while the rest barrell then all other words will pass the looping data. This mixed or "precessing" mode is illustrated in Fig. 3.5(c).

The above extension of the serial shift register to a pseudo-core memory has not been studied in depth in this work.

3.4.3 Choice of Slice Fault Distribution Model

A realistic assessment of technical feasibility and commercial viability of such a serial memory structure requires a detailed knowledge of the performance to be expected for any particular chain-generating algorithm on chip arrays of known yield. Ideally one would investigate algorithm performance for real cases; that is the fault distributions for slices manufactured on a particular process line would be used for the arrays in algorithm investigation. There are, unfortunately, two serious drawbacks to this ideal situation. Firstly, manufacturers tend to be rather secretive about their yield levels and showed reluctance to release this information. Secondly, the fault distribution is likely to depend strongly on the process technology. For example the balance between crystal fault density (which tends to increase towards the edges of the slice) and photolithography induced faults (pseudorandomly distributed across the wafer and far more serious on a 10 mask ECL process with the additional complication of buried layer, epitaxy and extra diffusions than on a 4 mask MOS process) must depend on the process technology. An analysis based on one particular process would therefore not apply in general to other processes.

The theoretical treatment of integrated circuit device yields is complex. Various yield models e.g. Murphy[38] and Price[43] have been proposed; the more widely accepted ones are critically compared in a recent review[53]. The choice of model depends on what assumptions are made concerning the nature of the defects. Factors to consider include (i) the randomness of distribution and interdependence of defects, (ii) whether or not they are distinguishable and (iii) the size and effect of different defects.

If it is assumed that all defects are very much smaller than individual chips and local increases of defect density (e.g. towards the wafer edge) are ignored, the siting of good chips on the wafer may be modelled by a random distribution. Such a random case with zero clustering of bad chips (as distinct from statistical grouping which must occur even in random arrays) provides effectively a pessimistic situation in that if the defective chips are clustered they provide less of an obstruction to the developing spiral than if they are not. Since it is difficult to envisage any procedure in slice processing whereby the random grouping will be inhibited - whereas there are many obvious causes of clustering[37] (e.g. tweezer scratches, large area photolithographic faults, crystal defects) then spiral lengths on actual slices should be greater than those predicted by the studies of Chapter 5. The typical case of defect clustering towards the slice rim, for example, means that the actual yield is considerably higher in the central region of the slice (for perhaps three quarters of the total area) than the overall yield for the slice as a whole. It must, however, be conceded that exceptional cases exist where clustering is detrimental - for example a long scratch extending across the slice diameter would restrict the spiral to one-half of the slice area. Certain ordered arrays (e.g. chess board pattern) also preclude the possibility of long chains but there is no reasonable mechanism whereby such arrays could be created accidentally in practice.

A random fault distribution model was therefore selected for the simulations, reconciling the fact that it would be extremely difficult to ensure that any quasi-random arrays which might be generated were, in fact, truly random with the thought that slice fault distributions are also not truly random. It is noted that Manning[34] also selected a random distribution array model.

It is further noted that Manning's spiral[33] is launched at a corner of the array. This has the advantage that the spiral cannot become trapped between itself and the array edge; however this is not generally a serious problem unless grossly non-square arrays are considered. It will also be seen (in Chapter 5) that, on reaching the array edge, spirals launched near the array centre are forced into retrograde motion. This can be avoided by launching the spiral at the array edge.

However, actual slices have higher defect densities in their edge region so that if the spiral is launched near the central (higher yield) area of the array it will have less chance of being extinguished during its critical early stages of growth. This procedure was adopted in the computer simulations of Chapter 5.

3.4.4 Number of Nearest Neighbours

It had been suggested[10] that an array having six nearest neighbours would permit the development of long spirals at lower yield levels than one with only four choices of adjacent chip. With conventional processing the requirement to scribe the wafer restricts close-packed chip arrays to rectangular or triangular shapes but with wafer-scale integration there is no requirement to scribe the slice and any chip shape is acceptable. The hexagonal and "brick-wall" arrays of chips in Fig. 3.6 both allow edge access to six nearest neighbours. Such arrays are compatible with current step-and-repeat procedures either by interlacing two orthogonal steppings or by a single stepping of the two-chip units shown.

It will be seen in Section 6.2 that a chip's "neighbours" do not have to be physically adjacent to that chip and the limit on the number of neighbours is therefore set by practical rather than topological considerations.

The increased complexity of chip Z required to implement such algorithms is relatively unimportant as this will be shown to contribute only a minor item to W.S.I system costs. A study of the effects of such algorithms on chip complexity is, however of importance in optimising the design of the serial memory.

By way of example, a move from four nearest neighbours to the six of the hexagonal array requires a 50% increase in that part of the control logic concerned with the gating of data across chip interfaces. This increases the chip area (both gate area and metallisation) required to implement the algorithm, resulting in larger chips (and hence lower yield) for a given memory capacity. If the 6-way algorithm is to offer any advantage it must at least compensate for this induced yield reduction by an improved tolerance of lower yields. Such an assessment requires the detailed studies of algorithm performance presented in Chapter 5.

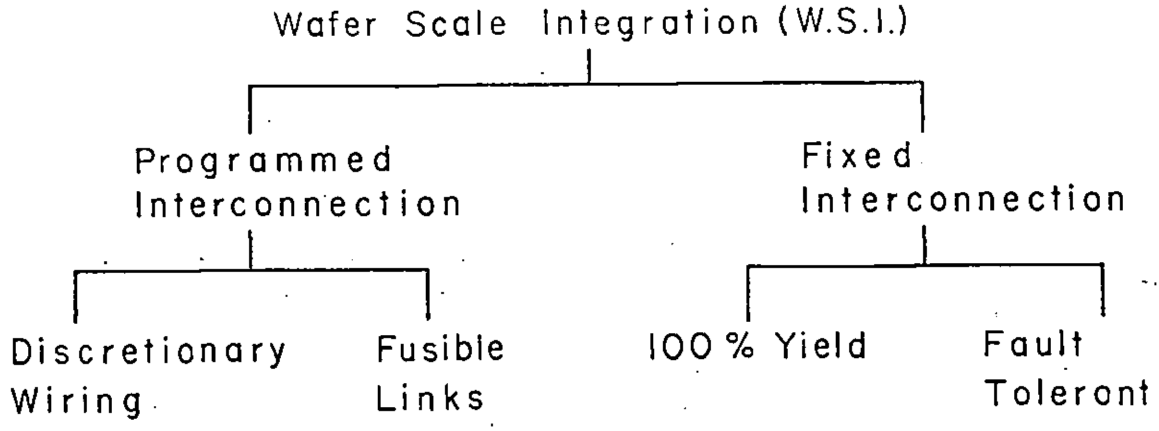


FIG. 3.1. Wafer-scale integration techniques.

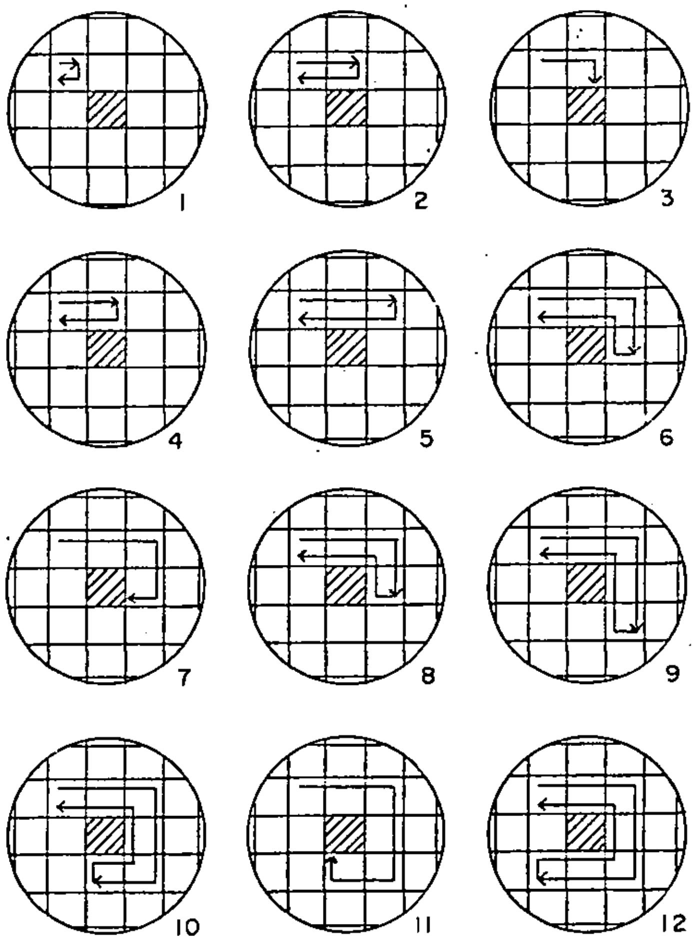


FIG. 3.2. Twelve stages in the early formation of a spiral.

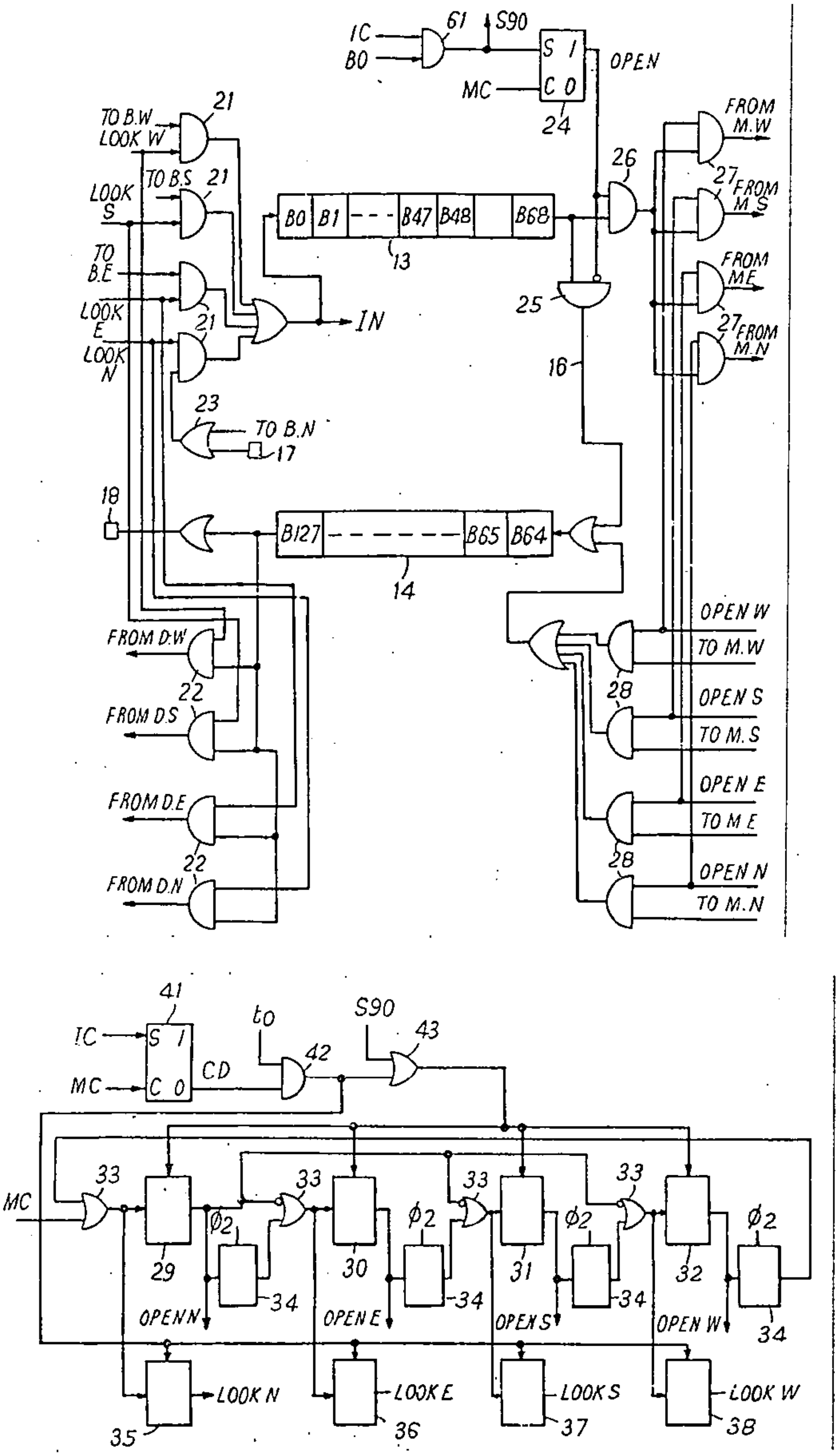
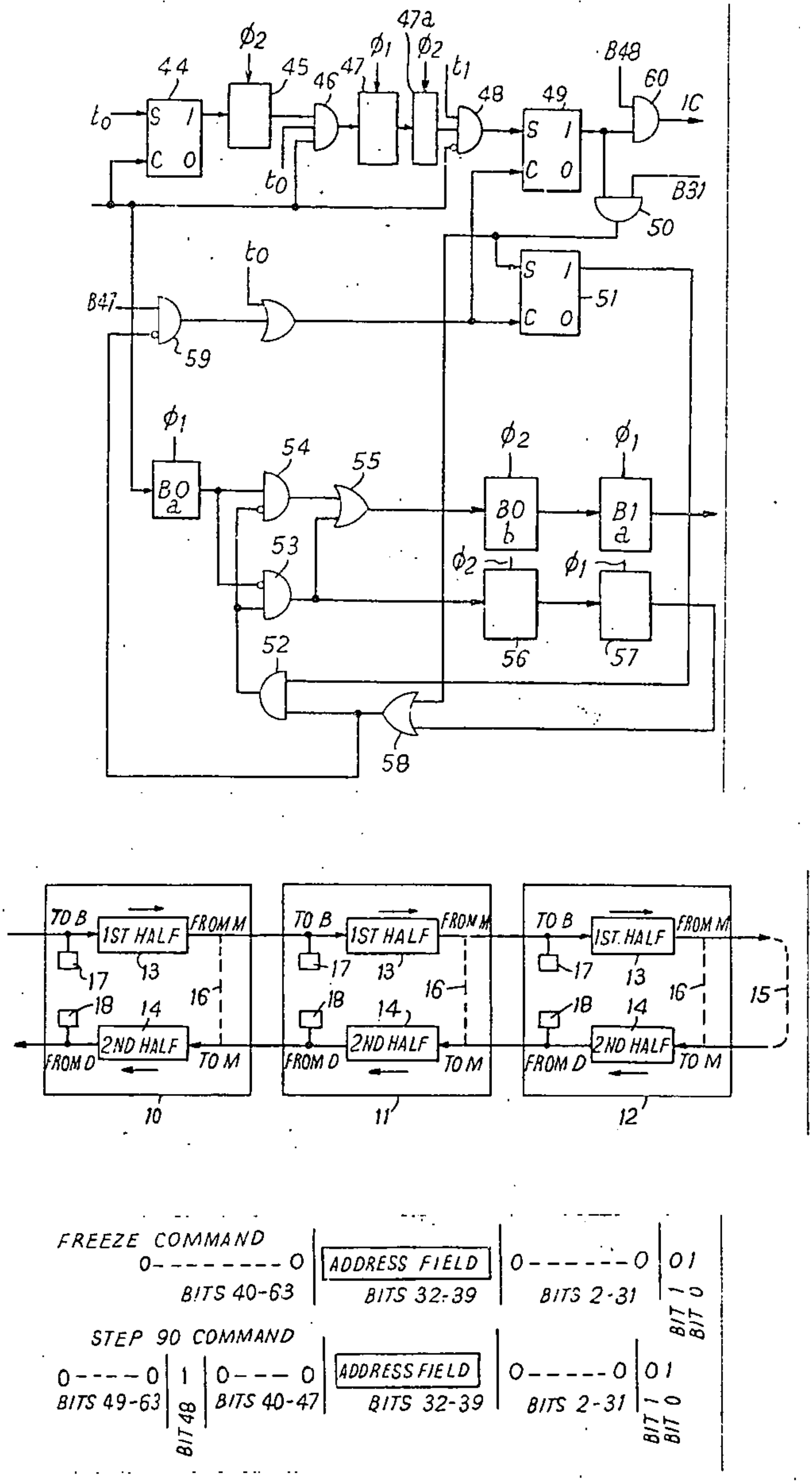


FIG. 3.3a. Logic design from British Patent 1,377859.



FREEZE COMMAND

0--------0 | ADDRESS FIELD | 0-------0 | 0 1

BITS 40-63 | BITS 32-39 | BITS 2-31 | BIT 1 BIT 0

STEP 90 COMMAND

0----0 | 1 | 0---0 | ADDRESS FIELD | 0-----0 | 0 1

BITS 49-63 | BIT 48 | BITS 40-47 | BITS 32-39 | BITS 2-31 | BIT 1 BIT 0

FIG. 3.3b. Logic design from British Patent 1,377,859.

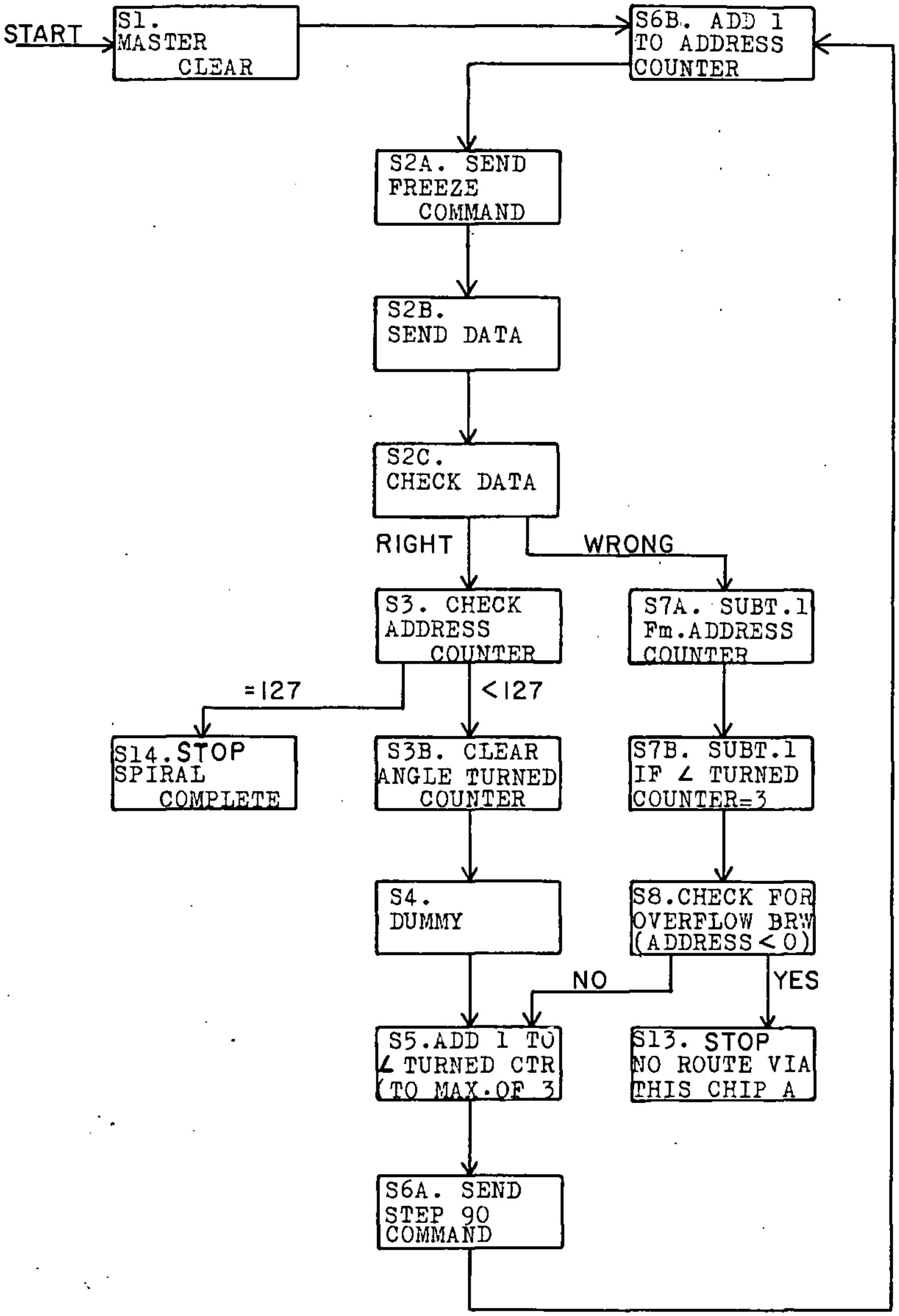

START
S1. MASTER CLEAR
S6B. ADD 1 TO ADDRESS COUNTER
S2A. SEND FREEZE COMMAND
S2B. SEND DATA
S2C. CHECK DATA
RIGHT
WRONG
S3. CHECK ADDRESS COUNTER
S7A. SUBT.1 Fm.ADDRESS COUNTER
=127
<127
S14.STOP SPIRAL COMPLETE
S3B. CLEAR ANGLE TURNED COUNTER
S7B. SUBT.1 IF ∠ TURNED COUNTER=3
S4. DUMMY
S8.CHECK FOR OVERFLOW BRW (ADDRESS < 0)
NO
YES
S5.ADD 1 TO ∠ TURNED CTR (TO MAX.OF 3
S13. STOP NO ROUTE VIA THIS CHIP A
S6A. SEND STEP 90 COMMAND

FIG. 3.4. Block diagram of CHIP Z function.

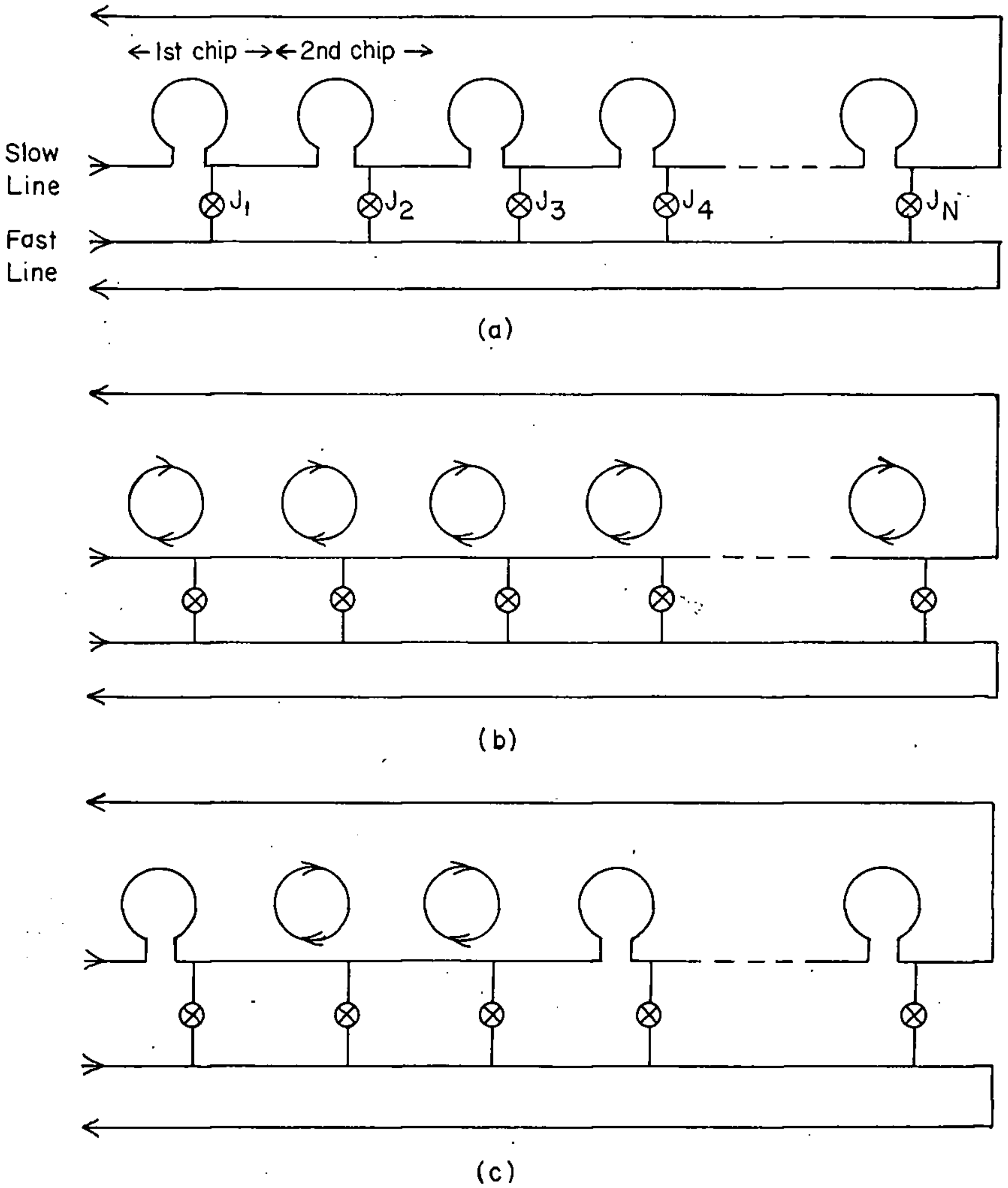


FIG.3.5. Shift register in (a) Barrelling Mode, (b) Looping Mode and (c) Precessing Mode.

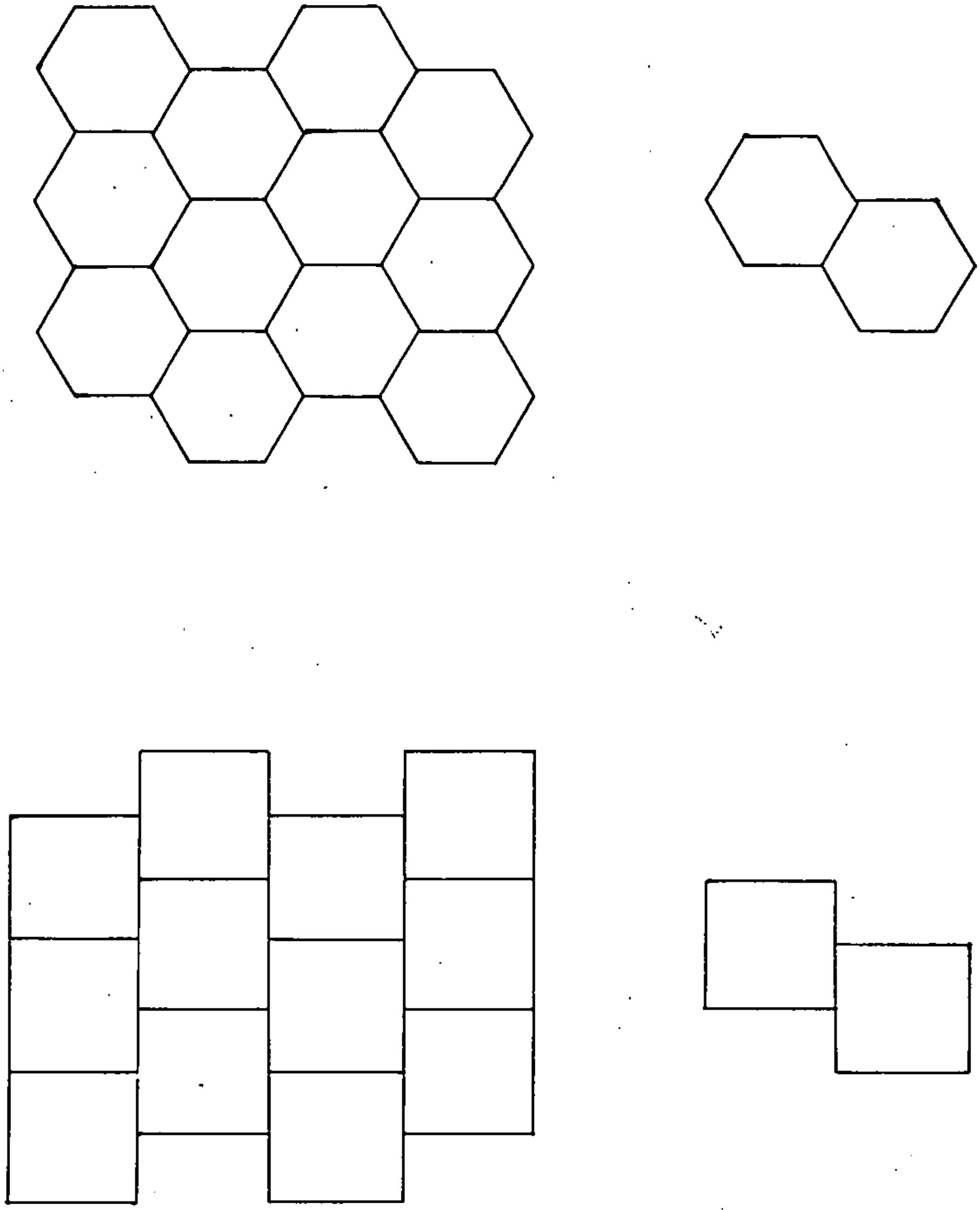


FIG. 3.6. Two forms of hexagonal array with two-chip blocks suitable for conventional step-and-repeat (orthogonal) techniques.

4. DESIGN AND FABRICATION OF THE SHIFT REGISTER ARRAY

This chapter describes the verification of the logic design by TTL simulation and the design of an IC chip to investigate the feasibility of laying-out the on-chip control circuitry (chip A) in a reasonable area of silicon.

The fabrication of functional chips has not yet been achieved; this objective is part of an on-going programme of work.

4.1 DESIGN VERIFICATION

The logic design described in Section 3.4.1 had not been proved prior to the commencement of the project. Two minor errors were noticed and, although these were easily rectified, it was considered advisable to verify the complete design of the on-chip control logic by hardware simulation. The decision had already been taken to use 4ϕ dynamic P-channel enhancement mode MOS technology for reasons to be stated in Section 4.2.1 so,at the expense of a considerable increase in circuit complexity, it was decided to simulate as far as reasonably possible some of the essential features of 4-phase circuitry. In particular, the concepts of "delay-free" and half-bit delay gates and the use of two clocks to drive alternate gates (equivalent to the ϕ_1 and ϕ_2 waveforms of 4ϕ dynamic MOS) were simulated. This last point was considered particularly important as a phasing error (which would have caused a race condition) was noticed on the original circuit design.

4.1.1 Shift Register

Fundamental to the recognition of commands is the shift register for data storage on each chip. To evaluate the logic it is sensible to use the shortest practicable length of register. Examination of the requirements of the shift register suggests a minimum length of 22 bits for the outgoing data path and zero for the return path. These figures are based on the following assumptions.

a) The address field, from which a chip decides whether or not it is to execute a command requires 8 bits to allow for a 128 chip memory.

b) Three additional bits are required at the end of a word to establish whether the command is FREEZE or STEP 90.

c) The first half of a command word (i.e. before the address field commences)must be all zeroes with the exception of the first bit.

The minimum length is therefore (8 + 3) x 2; the next orthodox (2^n) shift register size above this is 32 bits so this was selected for the outgoing path.

The next point to note is that access is required to bits other than the first and last of each register so at least part of the shift register must have parallel outputs. Rather than complicate the design of the simulation by using combinations of SISO and SIPO registers, four SN 74164 (serial input, parallel output) devices were selected to represent the outgoing shift register; this would also have permitted access to other bits had this proved to be desirable. In principle the return register could have been of zero length but, as this would not permit as full a check on the design as one of non-zero length, a 32 bit register was selected for this also. As 32 bit SISO registers are not readily available in TTL four 8 bit SIPO devices were used instead so that flow and return data paths were identical and the facility for tapping the return shift register also existed should this have been required.

4.1.2 Random Logic

The rest of the on-chip control logic proved to require 39 TTL devices in this simulation. These were suitably partitioned between three boards as illustrated in Figs. 4.1, 4.2, 4.3. The design was checked on a logic tutor panel prior to hardwiring the printed circuit boards. The section controlling the cyclic enabling of the four LOOK and OPEN ports in the free-running condition was easily verified; the other two boards, however, had to wait for the commissioning of the external control electronics (known as chip Z) to be fully proved. Light emitting diodes were wired on to the eight OPEN/LOOK terminals to indicate the state of each chip.

4.1.3 Pulse Generator

The clock waveforms required for the particular variant of 4ϕ PMOS technology selected are very simply related; those shown in Fig. 4.4 are quite adequate for the purpose. Such waveforms are readily available at 5V by NAND gating the relevant outputs of a counter as illustrated in Fig. 4.5 The other regular timing signal, t_o, is readily obtained by frequency dividing the appropriate clock pulse (ϕ_3) as shown. Master Clear in normal operation would only be required prior to the initial setting up of the spiral and in the event of a chip failure necessitating collapse and regrowth of the spiral. However, for evaluation purposes it is convenient to periodically master clear and reconstruct the spiral; a suitable waveform for this purpose is obtained by dividing down t_o as shown.

The master clock input to this pulse generator was initially supplied by a multivibrator, adequate waveshaping of this far-from-ideal waveform occurring in the early stages of the pulse generator. This was made frequency variable by capacitor selection but was eventually replaced by a purpose-built oscillator.

A photograph of the completed module and the detail of the circuit boards for the pulse generator and control logic is given in Fig. 4.6.

4.1.4 16-Chip Array Simulation

This single model of chip A, described above, provided a check on much of the on-chip control logic as well as of chip Z (the external control electronics required to govern the development of the spiral). It was, however, quite incapable of evaluating the performance of chip Z in bypassing faulty chips and backtracking out of blind alleys during spiral generation. It was therefore decided to build a rack representing a matrix of sixteen chips in an orthogonal 4 x 4 array, on which chip Z could attempt to build a spiral through a quasi-random distribution of 10 good chips. Each site provided three edge connectors to accommodate the set of printed circuit boards carrying the TTL devices. To indicate the directions to which and from which each particular chip was sending and receiving data, eight light emitting diodes were wired into all 16 chip sites. A further three LED's were provided to indicate the state of other key points in the logic, to assist in any debugging of the rack which might prove to be necessary.

A model of chip Z, designed and built by the inventor, was installed into the rack, together with the purpose-built clock generator described earlier. This, however, required substantial expansion of its fanout capability in order to cope with the 390 gates loading ϕ_3 and the 310 loading ϕ_1 in addition to the load imposed by chip Z. This could have been achieved using power transistor stages but was, in fact, implemented using eleven 74128 line drivers (taking care to rely on matching of switching parameters only within a device rather than among devices) as sketched in Fig. 4.7.

The rack was now completely self-contained, requiring only a 5V dc input - although the high current requirement (13A in total) demanded care in the selection and routing of the wiring.

The completed rack is illustrated in the photograph of Fig. 4.8. The experiments performed with this are summarised in Section 4.3.

4.2 DESIGN INTEGRATION

In evaluating this approach to wafer-scale integration it is important to establish that the on-chip control logic can be laid out in a reasonable area of silicon. It is well known that random logic generally requires a substantially larger area for interconnection than ordered structures (e.g. shift register) and it was further anticipated that the requirement to route four signals to all four edges of the chip would involve a substantial overhead in chip area. The proportion of chip area required for the control logic is expected to depend to some extent on the choice of circuit technology while the actual area depends strongly on the design rules (i.e. minimum permissible dimensions) acceptable to each particular production line. Rather than attempt to compare the parameters for different types of process it was decided to construct a detailed layout for one particular process. It was further decided that the design should be fully engineered for the in-house processing of working devices with the possibility of later adaptation to a full-slice array.

4.2.1 Choice of Process Technology

Both this particular serial memory structure and wafer-scale integration generally are technology independent so any semiconductor process technology may, in principle be selected for their implementation.

The initial selection of technology to be applied presented little problem, for while the Polytechnic Microelectronics Centre had both a proven p-channel MOS capability and a bipolar process line, the latter was based on RTL technology and was at that time lacking in-house buried layer and epitaxy facilities. It was felt that possible delays in setting up these processes or awaiting processing at an external facility were best avoided so the p-channel MOS process was considered the obvious choice.

While a CMOS process would have offered the advantages of reduced power dissipation and increased circuit design versatility the further development of the processing capability to cater for CMOS would have been a major undertaking and so was not considered for the project.

Having decided to utilise the p-channel enhancement mode process the next choice was static or dynamic logic. Again there appeared to be an obvious choice - in favour of dynamic logic as static circuitry seemed to offer no real advantage in this application to offset its substantially increased chip area and power dissipation.

The more difficult choice arose when the many forms of dynamic logic were compared. Many variants of both two and four phase circuits existed, requiring either ratioed or ratioless transistor sizes. The types requiring ratioed devices for the active and load transistors have the advantage of reduced noise and fan-out sensitivity and low clock line capacitive loading. These circuits have, however, the big disadvantages of considerably increased power dissipation and gate area, as ratioless circuits may use minimum area transistors - i.e. unity aspect ratio ($^W/L = 1$) of smallest possible W and L - set by processing limitations and typically ~ 10μm.

While both two phase and four phase dynamic ratioless circuits may use minimum area transistors it is possible to design a substantially smaller shift register stage with four phase circuitry than with two phase. Power consumption is similar for the two types.

In choosing the circuit configuration to be adopted it was considered better to utilise a type known to have been in common use by the device manufacturers rather than to experiment with the (perhaps) theoretically superior designs quoted in the literature. It appeared from examination of data sheets and catalogues from the major MOS manufacturers that there had been a general move away from four-phase circuitry by the industry. This was due primarily to the problem of noise generated by the clocks when driving the device at high speed ($\gtrsim$ 5 MHz). However, in this particular application there was no requirement for high speed operation, the dynamic mode having been selected for its other advantages of minimum bit size and low power dissipation, so the clock edges could be reduced in slew rate if necessary to avoid this problem.

A comparison of the power/clock requirements of 4ϕ dynamic MOS technology with those of the more popular 2ϕ circuitry (see Fig. 4.9) showed that both types require four lines to be distributed around the chip. The 2ϕ dynamic configuration requires a separate drain voltage supply and a wired earth connection to connect the P^+ diffused MOST source regions. The 4ϕ dynamic inverter requires no dc supply and the only earth connection is that implicit in the parasitic capacitances to the substrate, for which a back face contact (also required for the 2ϕ circuit) is adequate. As the ϕ_2 and ϕ_4 clocks are readily generated at each chip from the ϕ_1 and ϕ_3 signals, only two power/ clock lines need supply each chip in 4ϕ dynamic operation.

4.2.2 Circuit Design

Any logic design may be implemented using only one logical function - either the NAND or the NOR may be used. The NAND function (defined using negative logic for P-channel MOS) requires the input transistors to be in series; this in turn requires the spacing between clock lines to be increased. The NOR function, in requiring a parallel configuration for the input transistors, wastes an appreciable area of silicon in the vicinity of that gate but does not require any increase in spacing of the clock lines. It therefore depends strongly on the relative incidence of single input gates (inverters) and multiple input gates whether the NAND or the NOR configuration will offer a more compact layout for any particular circuit. The NOR element was selected as this particular design contained many basic inverter stages; this enabled all the clock lines to be run at minimum spacing (see,for example, Fig. 4.13)

Control Logic

Prior to conversion of the (random) logic design into a circuit schematic, the circuit design phase commenced with a study of the literature. Various design features were noted from (21), (B1),(B4),(B8) and (B12) and more sophisticated points arose from discussion with workers in the industry. Examples of pitfalls for the unwary are now summarised in relation to Figs. 4.10 and 4.11.

a) Delay-free inverter.

Delay-free inverters (i.e. gate types 2 and 4) must not be fed directly from a stage which drives another gate type; for example, the circuit illustrated in Fig. 4.10(a) is not acceptable, but must be replaced by that of Fig. 4.10(b). This may be seen as follows.

If input is at logic 1 (i.e. -ve) C_1 charges during ϕ_1 time and discharges through TR2 and TR3 when ϕ_1 goes to Earth. C_2 is also charged by ϕ_1. On the negative-going edge of ϕ_3, C_3 will cause the voltage at node N to go negative; this -ve excursion can exceed the threshold voltage of device TR8 which will turn on at the beginning of ϕ_3 time, thus destroying the logic 1 held on C_3 by permitting charge to pass to ground via TR8 and TR9. This problem is eliminated by deriving the inputs to the type 2 and type 3 gates from separate output circuits of the driving (type 1) stage as shown in Fig. 4.10(b). Several instances in the design required this treatment.

b) Charge injection into the substrate.

Without considering the detailed waveforms obtained at the circuit nodes it will be readily observed that the voltage at a node can, under certain conditions, make a positive excursion with regard to the clock zero voltage. This is shown qualitatively in the sketch of Fig. 4.11(b).

The critical case is for logic 1 at the input, so that node A decays towards zero during the ϕ_2 period (after ϕ_1 has returned to Earth) and is driven through zero to a positive voltage by the effect of the rising ϕ_2 edge being coupled onto node A by parasitic capacitance (mainly gate-to-drain capacitance of TR2). This positive excursion of node A allows the PN junction forming the source of TR1 and the drain of TR2 to go into forward bias and inject charge into the substrate. The long carrier lifetimes of MOS material allow this charge to travel to adjacent, or even distant, regions carrying a favourable (negative) bias. As this will cause unpredictable changes in logic level at these collecting nodes, such forward injection cannot be tolerated. The easiest method of preventing this effect is to work with a pseudo earth on the clock lines (or to bias the substrate positively which amounts to the same situation). A positive bias of 2 V with regard to the clock zero should be adequate; this could be increased if necessary but it should be noted that the resultant increase in source-substrate bias will increase the threshold voltage of all the MOS transistors on the chip.

A complete circuit design was produced, taking into consideration such details as the above; this is presented in schematic form in Fig. 4.12.

Gate Protection

General principles of MOS integrated circuit gate protection are discussed in Section 6.8.3.

It was expected that the proposed MOS process, having a single dielectric, would provide a greater ruggedness than structures employing a composite dielectric (e.g. MNOS) where there is a less favourable electric field distribution within the insulator.

However, there was no information concerning the ruggedness of the gate dielectric in the current process and it is well known that submicroscopic inhomogeneities in the dielectric can cause premature breakdown (and therefore lack of ruggedness). It was therefore decided to include full gate protection networks on all terminals (other than substrate earth) of the discrete chip to avoid any evaluations being dogged by failures due to static electricity.

4.2.3 Circuit Layout

The layout of 4ϕ dynamic MOS circuits is best undertaken using a six-clock-line format. The extra space required by the additional ϕ_1 and ϕ_3 lines is adequately compensated by the close-packing of gates achievable on this system. At least two variants on this six-clock-line scheme result in an optimised layout; the one selected is illustrated in Fig. 4.13, which is an extract from the complete layout of Fig. 4.14.

Design Rules.

In deciding minimum dimensions for linewidth and spacings it is necessary to take full account of both the limitations of the maskmaking/photoengraving equipment and the constraints imposed by the particular process invoked (4ϕ dynamic MOS). Included in this latter category are sideways diffusion, punch-through breakdown voltage (dependent on spacing and resistivity) and parasitic capacitance effects. The dimensions quoted in Table 4.1 were drawn up for this project as a reasonable compromise for the facilities available within the Polytechnic Microelectronics Centre.

Feature	Mask	Minimum Dimension (µm)
P^+ track width	1	10
Source/Drain separation	1	10
Unrelated P^+ - P^+ separation	1	15
P^+ region overlap of gate oxide on channel width	1,2	2.5
Unrelated P^+ - gate separation	1,2	15
Gate overlap of Source/Drain	1,2	2.5
Unrelated gate-gate separation	2	15
Gate linewidth	2	10
Overlap of contact hole by gate	2,3	2.5
Contact hole linewidth	3	10
P^+ edge-contact hole separation	1,3	5
Aluminium track width	4	10
Aluminium track spacing	4	10
Aluminium overlap of channel and contact holes	2,3,4	5

Table 4.1

It must be emphasised that these dimensions are by no means minimal. For example many workers in the industry would claim that the minimum separation for unrelated P^+

tracks and minimum clearance between any gate area and unrelated P^+ tracks could be reduced to 10µm for the intended material specification of 3-5 Ω cm. However, spurious effects have been detected in MIS integrated circuits of various manufacture in the past due to both of these dimensions being reduced to dangerously low values. These effects are probably less serious for MOS than for MNOS structures, are believed to be confined to higher resistivity silicon ($\gtrsim$ 12 Ω cm) and are not expected on the 3-5 Ω cm material to be used in this application. Nevertheless it was thought desirable to increase these dimensions to the 15µm quoted to ensure an adequate safety margin.

In order to retain a high degree of flexibility in the layout, no attempt was made to minimise the chip area by packing the gates together into the smallest possible space. As will be observed in the completed layout of Fig. 4.14 there are many areas of unused silicon; the control logic area - unlike the shift register which is tightly packed - could therefore be substantially reduced, thus improving the ratio of storage: control logic areas of the chip.

Test Components.

Several test structures were included in the design; these were directed primarily towards assessing the integrity of the P-channel MOS process. The more important parameters monitored were, aluminium step continuity, sheet resistivity and contact resistance, P^+ sheet resistivity, mask and operator misalignment, avalanche, punch-through and field-aided breakdown voltages and parasitic (field) threshold voltage. A discrete MOST was also included in the layout.

To assist in debugging of the design, three key points of the control logic were brought out to additional bonding pads; this would assist in fault diagnosis by permitting access to sections of the logic via a low capacitance probe or, if necessary, via additional bond wires.

Wafer-Scale Integration Compatibility.

To meet the possible requirement for future stepping as a W.S.I array the North, South, East and West input/output lines were positioned so that, on removing scribe lines and bonding pads etc., the P^+ tracks from any chip would link directly into the corresponding tracks of its four neighbours. For example, the "OPEN W" P^+ track was positioned on the left-hand side of the chip, directly opposite the "LOOK E" track on the right-hand side.

To avoid the requirement for bonding pads on any (and therefore all) of the array chips it was proposed that a special input/output interface chip would be placed in the array. The advantages of this are discussed in Section 6.8.4.

Artwork and Mask Preparation.

Having prepared a complete line drawing of the layout at the smallest convenient scale (500x) several possible routes for completion of artwork were available. Hand cutting of rubyliths was impracticable for reasons of both complexity and scale. Computer assisted mask generation routines had been partially developed[17] on the Polytechnic computer but the core storage capacity of 32k characters (2^{18} bits) and plotting facility (10" max. along axis of drum plotter) then available were inadequate to cope with a layout of this size. Sample plots of basic gates are presented in Fig. 4.15.

The most promising approach at that time seemed to be the CAMP system at RSRE Malvern. This accepts both Fortran and digitised data as input; both were investigated. The longhand program input is ideal for highly repetitive arrays but not for random logic; for example the basic type 1 inverter stage of Fig. 4.16(a) requires the 22 line program quoted in Fig. 4.16(b). The full layout would have required a very lengthy program with, almost inevitably, extensive debugging which cannot be done efficiently by postal communication.

The digitiser input possibility was therefore investigated; dimensional instability of the cartridge paper with changes of temperature and relative humidity proved to require redrawing of the layout at 1000x. Part of the layout was, in fact, redrawn at this scale and successfully digitised before the highly stable melinex grids (which would have permitted a scale of 500X) were obtained - three months behind schedule.

At this point an offer was made by the Allen Clark Research Centre of the Plessey Company for free use of their CALMA Interactive Graphics system. This was gratefully accepted and within a further 50 hours the layout had been generated, plotted and checked.

It was not practicable to produce the masks on the Polytechnic step and repeat camera; the maximum image field is ~ 90 thou.. square so the stepping of the 153 x 124 thou. chip would have required the interlacing of four separate patterns for each mask. This would have required very accurate registration across the boundaries of the image field - the region where distortion is at a maximum. The masks were therefore produced by Plessey Limited, Swindon, via pattern-generator plotted reticle plates.

4.2.4 Wafer Processing

Several batches of slices were fabricated; this involved modification to the p-MOS process to achieve compatibility with current industrial practice. Major changes included reducing the P^+ diffused junction depth to ~ 2μ, modifying the boron drive-in/oxidation schedule to reduce parasitic capacitances to the overlying clock lines and developing a phosphosilicate glass passivation layer process to improve device stability and resistance to contamination.

Problems arose mainly from the lack of any previous requirement to process LSI devices requiring both small geometries and high complexity patterns simultaneously. These were overcome by careful attention (and modification) to the photoresist exposure and aluminium etching stages. One problem, however, still remains. While batches have been made which demonstrate low and stable values of threshold voltage (- 3.5 to 4.0v) there is a recurrent problem of high and variable threshold voltage. This has now been tracked down to the metallisation stage - in particular to the tungsten filaments from which the aluminium is evaporated. Tungsten filaments generally contain sodium - a necessary additive to increase the ductility of the metal but a serious contaminant causing grave instability in MOS transistors. Although purchased as "sodium free" these filaments are introducing serious contamination to the aluminium/gate-oxide interface. Control slices from the Polytechnic process line, metallised elsewhere by electron beam evaporation have produced consistently low threshold voltages while in-house metallisation (on others from the same batch) produces high and variable threshold voltage. The general consensus of opinion in the industry is that filament evaporation of aluminium cannot be used as a consistently reliable procedure for MOS devices.

In view of this fact and the current lack of alternative in-house facilities it was decided not to spend further time attempting to process working devices until in-house electron beam deposition is available.

4.3 CONCLUSIONS ON CHIP A DESIGN FEASIBILITY.

The TTL simulation has verified the design of both the on-chip control circuitry (chip A) and the external control electronics (chip Z). This work illustrated the necessity for retaining information on the route taken by the spiral in order to backtrack out of blind alleys. Chip Z now stores the "OPEN" directions of the last 8 chips; to cope with longer back-tracks would require further storage capacity.

In addition to creating single spirals the rack has illustrated the ability of this logic design to permit the simultaneous generation of two spirals without interference. Two models of chip Z connected to two different chip A sites controlled the growth of two adjacent spirals in the array without any interaction.

Chip Z is currently being rebuilt to a more fully engineered design. This will incorporate output buffers to provide p-MOS compatible waveforms (0v and -25V rather than 0v and + 5V).

The integrated circuit layout of chip A demonstrated the feasibility of other aspects of the design concept. The control logic, although dominating the chip layout, in fact required an area approximately equivalent to a 256 bit shift register - a negligible overhead on (say) a 5K bit device. Although the metallisation is required to route four signals to all four edges of the chip there were no serious topological problems in implementing this. Chip area requirements of control logic and interconnections are discussed further in Section 6.8.5.

The supply of working chips, although of real benefit to algorithm simulation had it been achieved, was not a necessary part of the present study, being more relevant to the subsequent on-going development of W.S.I devices. However it is noted that the lack of adequate metallisation facilities is now believed to be the only problem preventing the manufacture of working chips; certainly several slices have produced many chips on which there are no visible defects likely to prevent correct operation of the device.

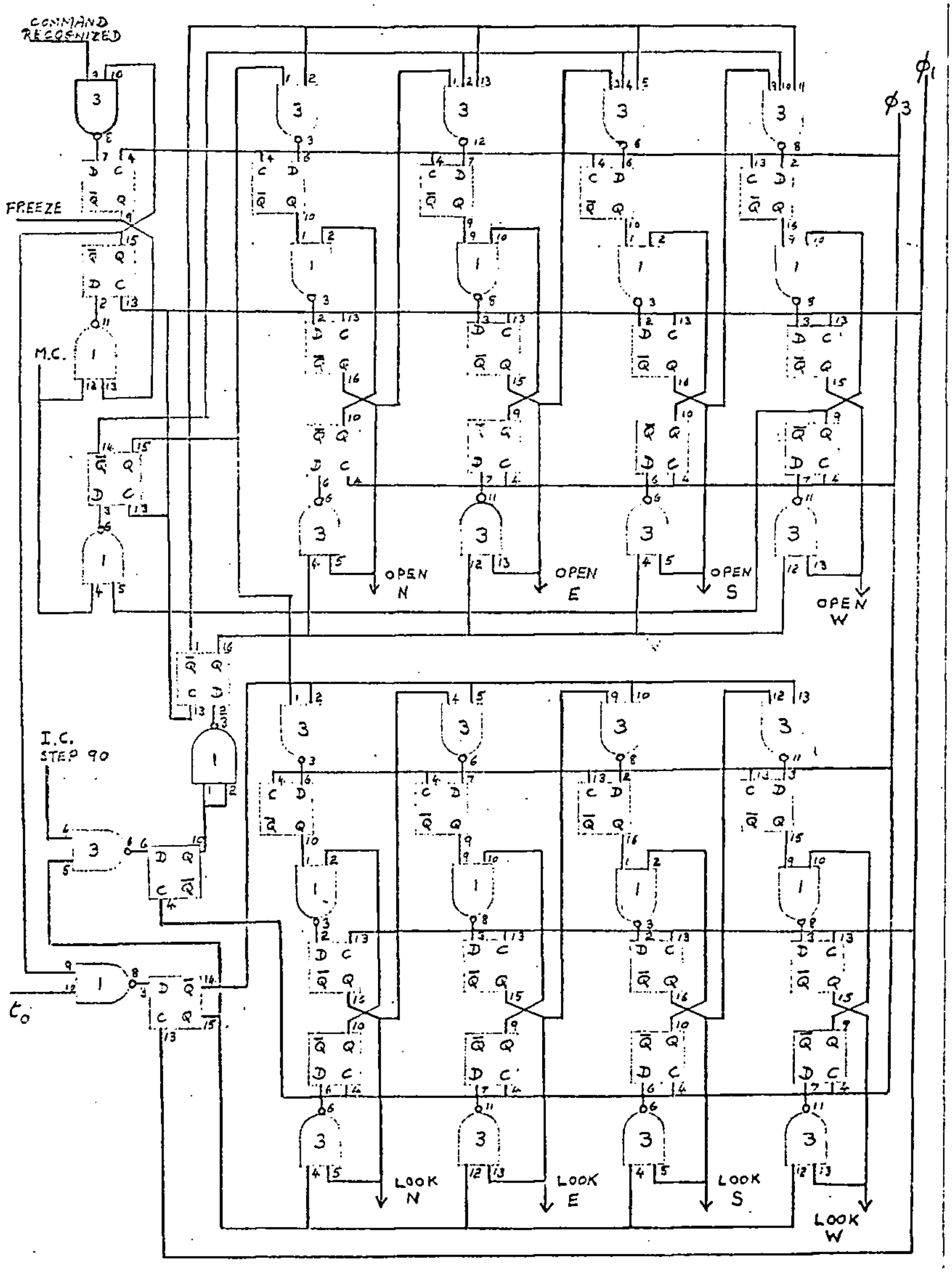


FIG. 4.1. Circuit for TTL simulation of on-chip control logic. (Part 1 of three).

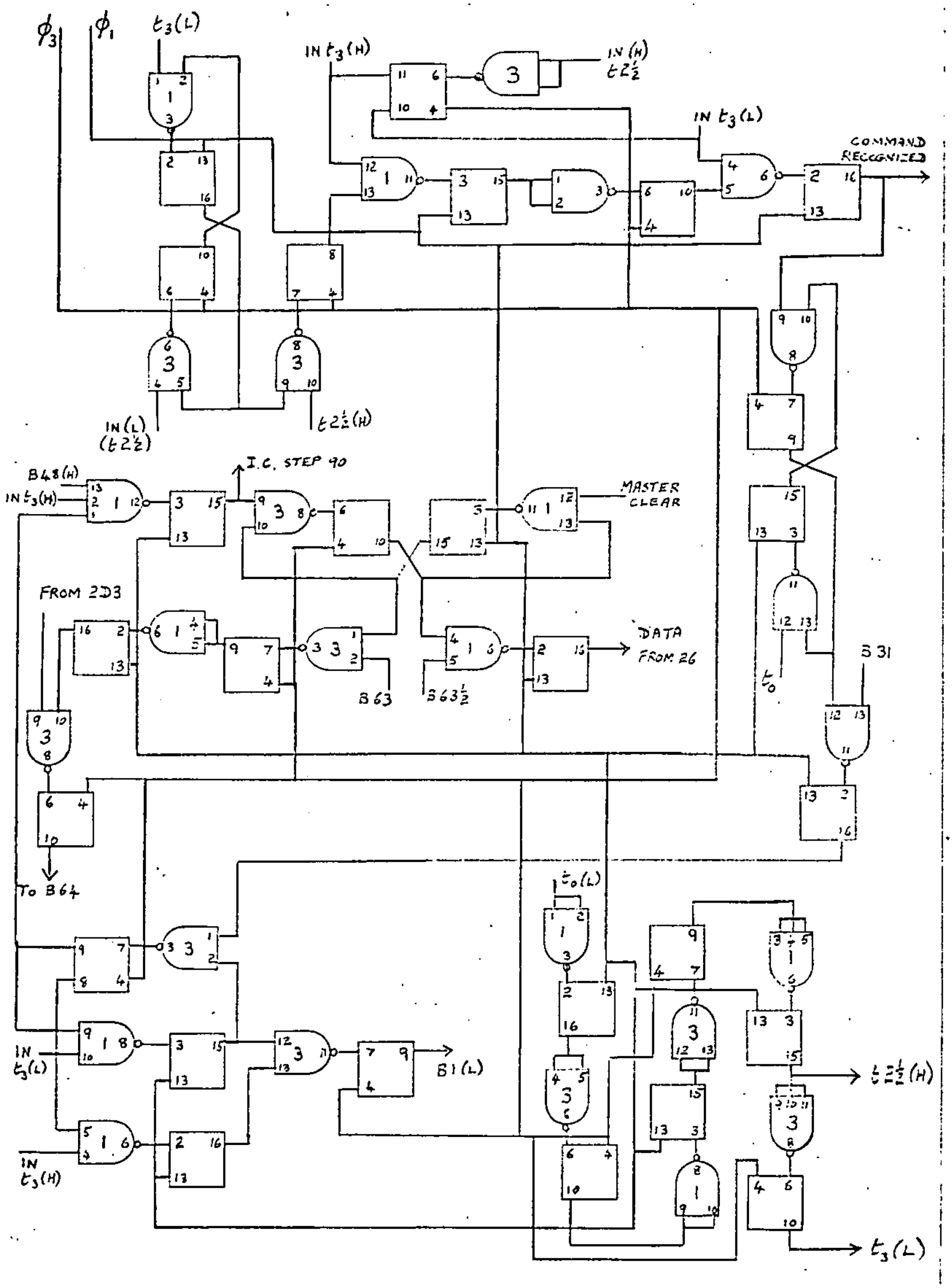


FIG. 4.2. Circuit for TTL simulation of on-chip control logic. (Part 2 of three).

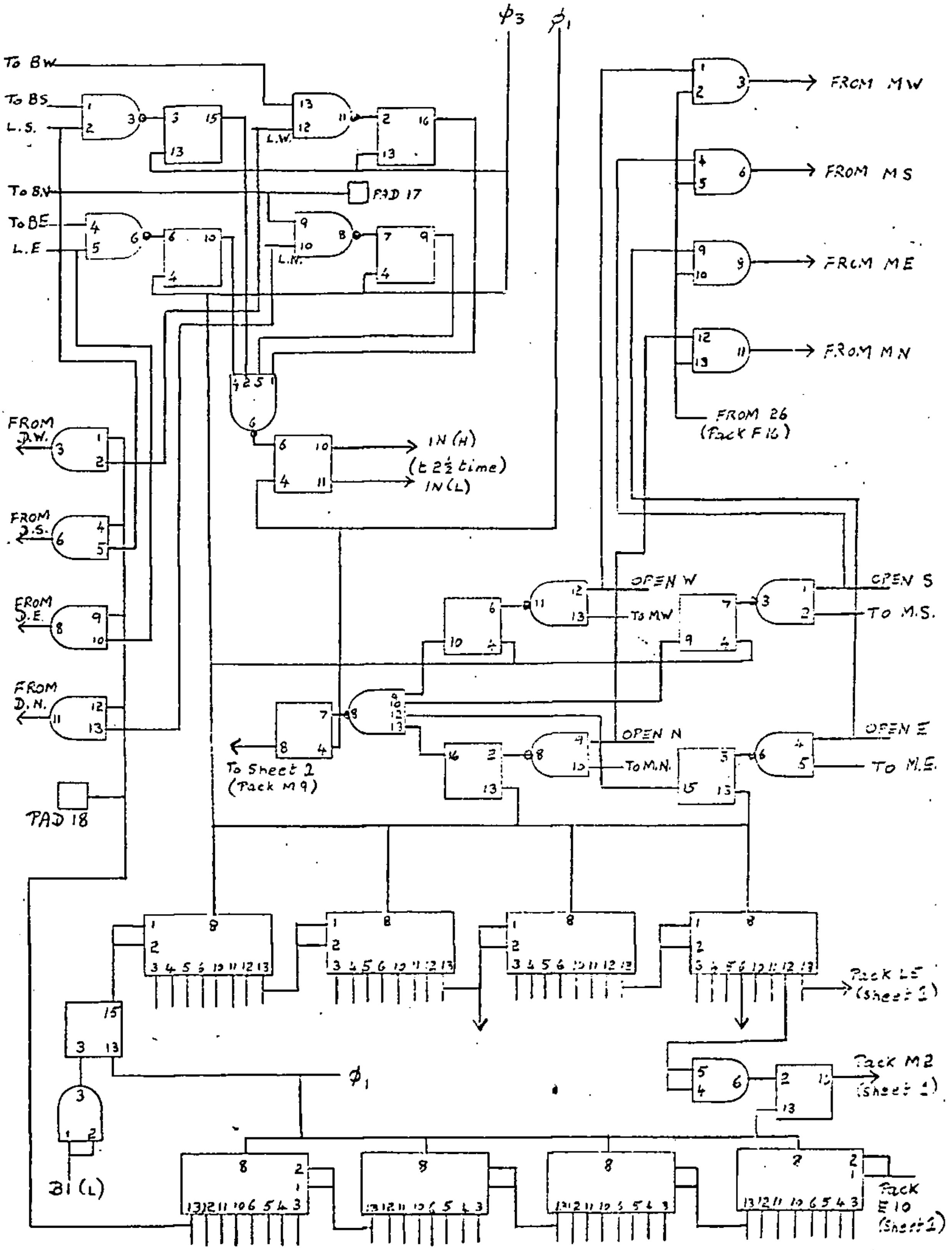


FIG. 4.3. Circuit for TTL simulation of on-chip control logic. (Part 3 of three).

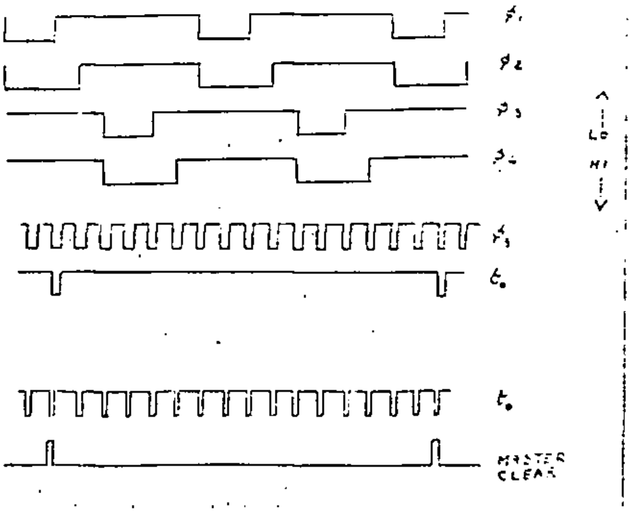


FIG. 4.4. Pulse relationships for four-phase MOS implementation of on-chip control logic (chip A).

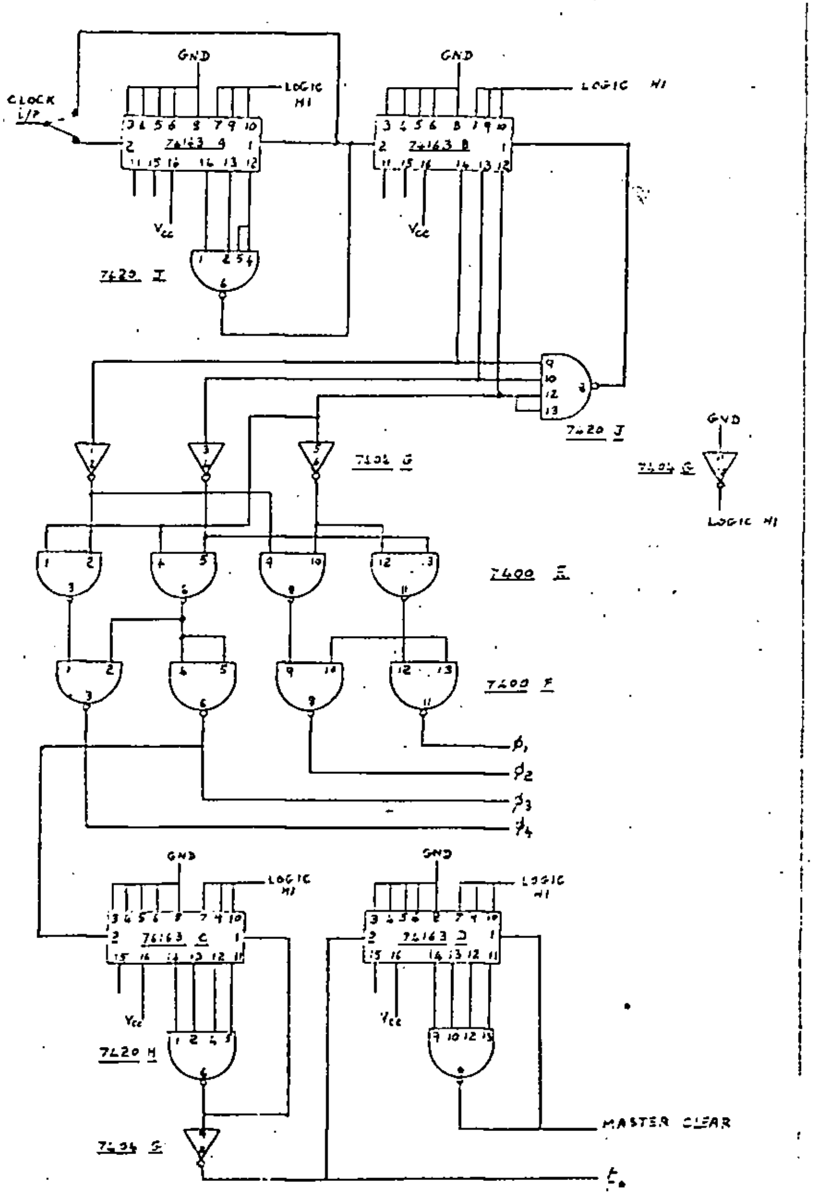


FIG. 4.5. Pulse generator circuit.

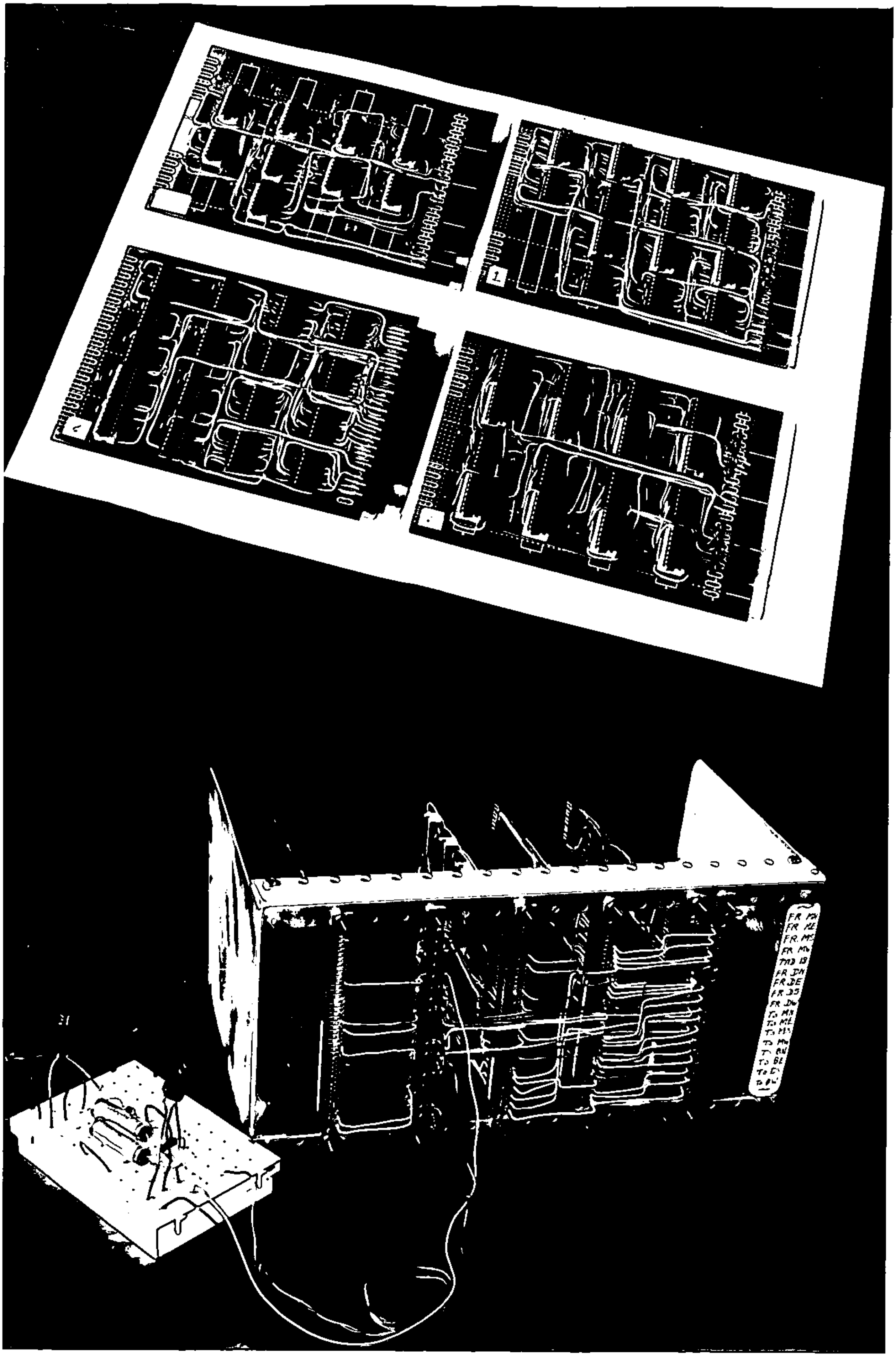


FIG. 4.6. Completed circuit boards and first chip A model.

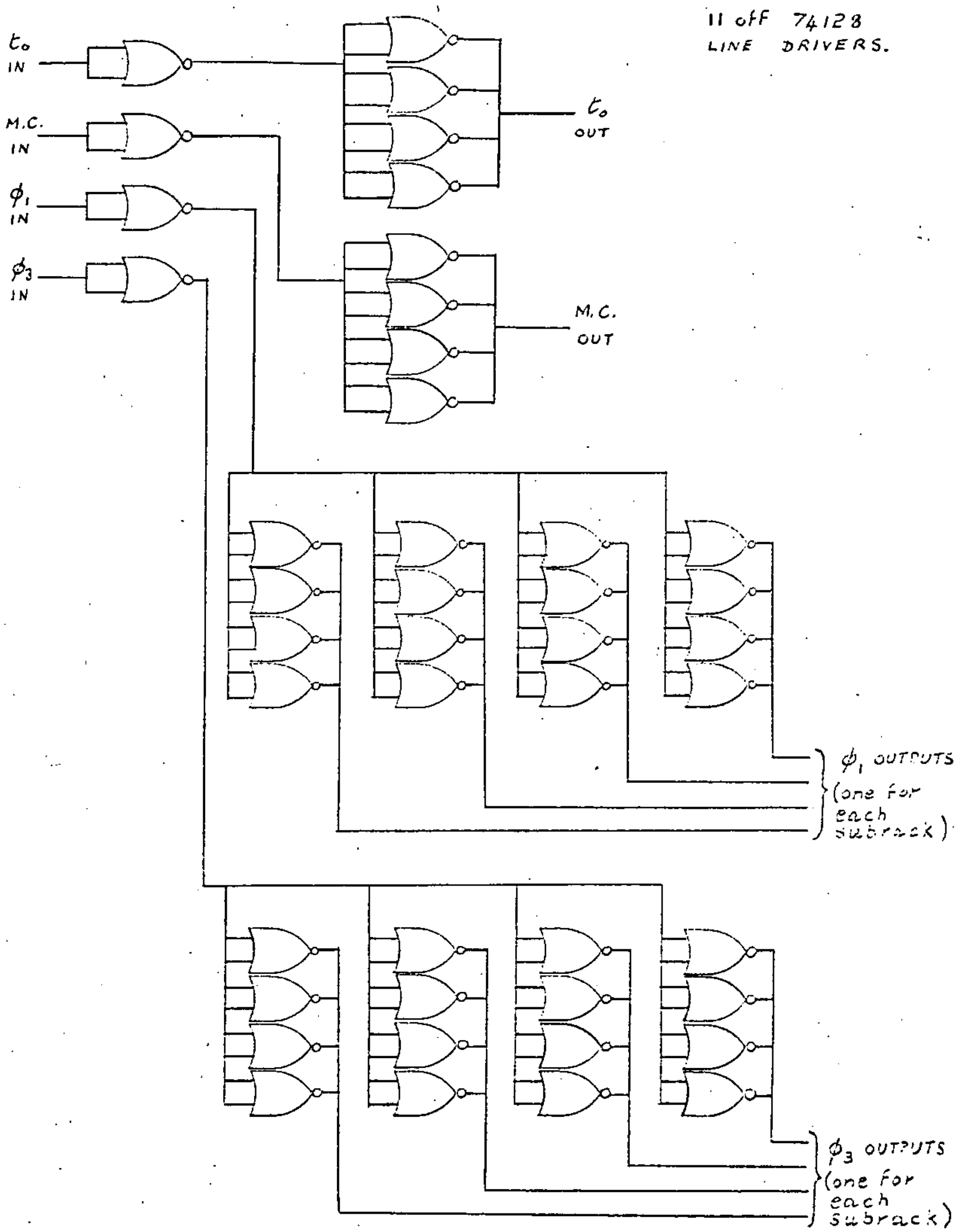


FIG. 4.7. Power driver output stage for pulse generator.

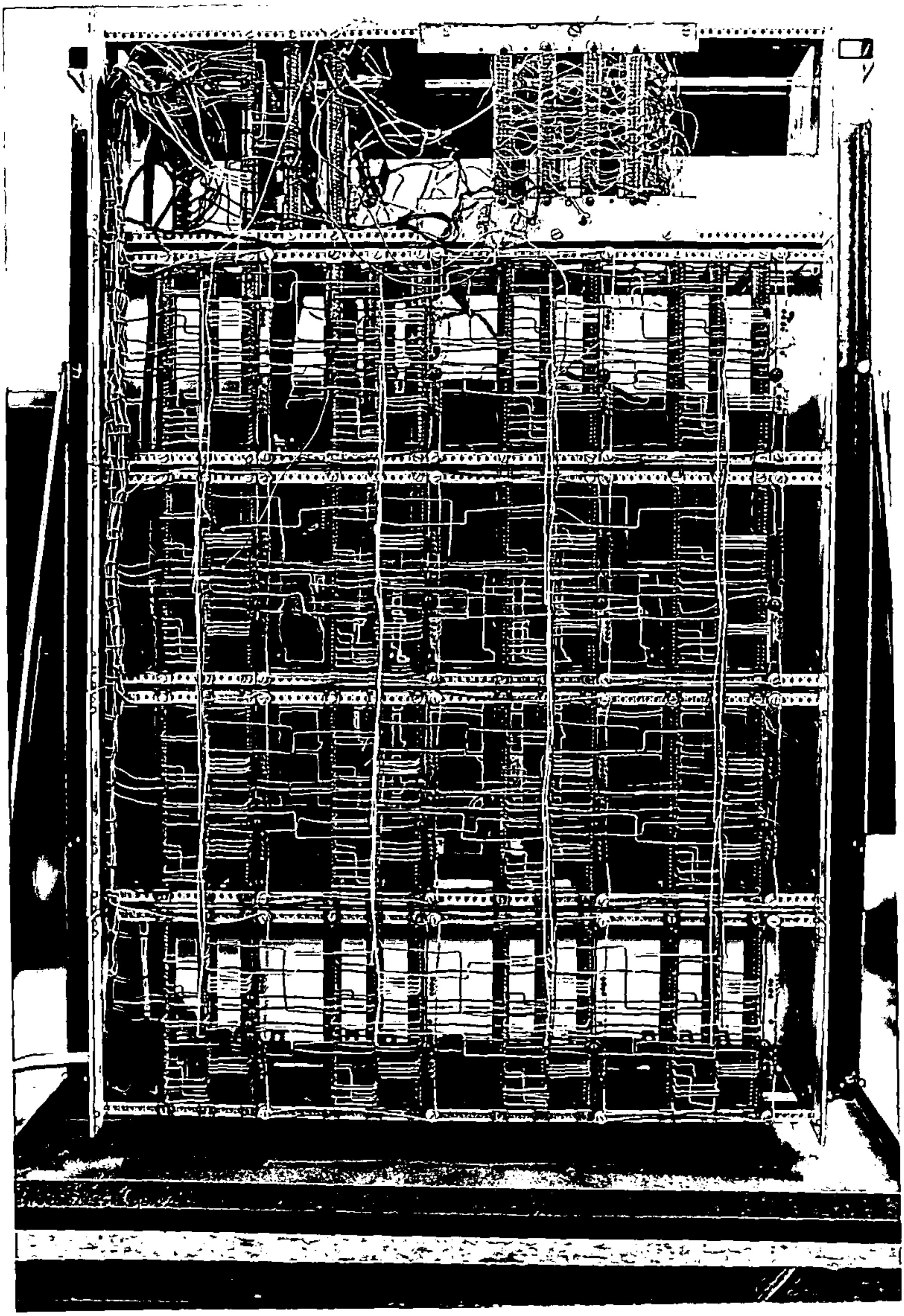
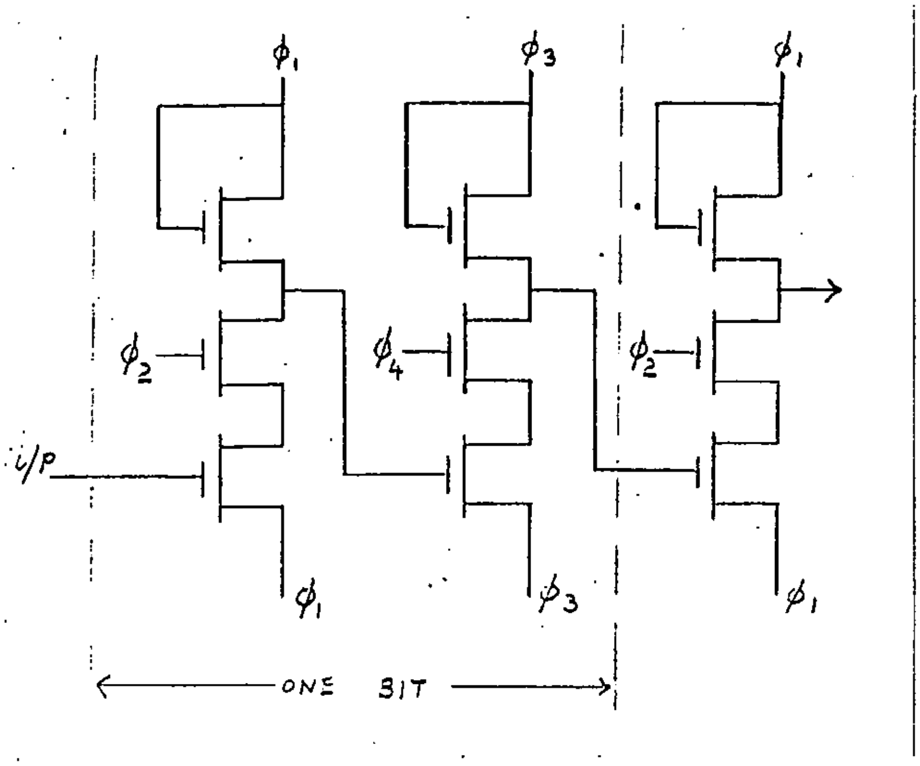
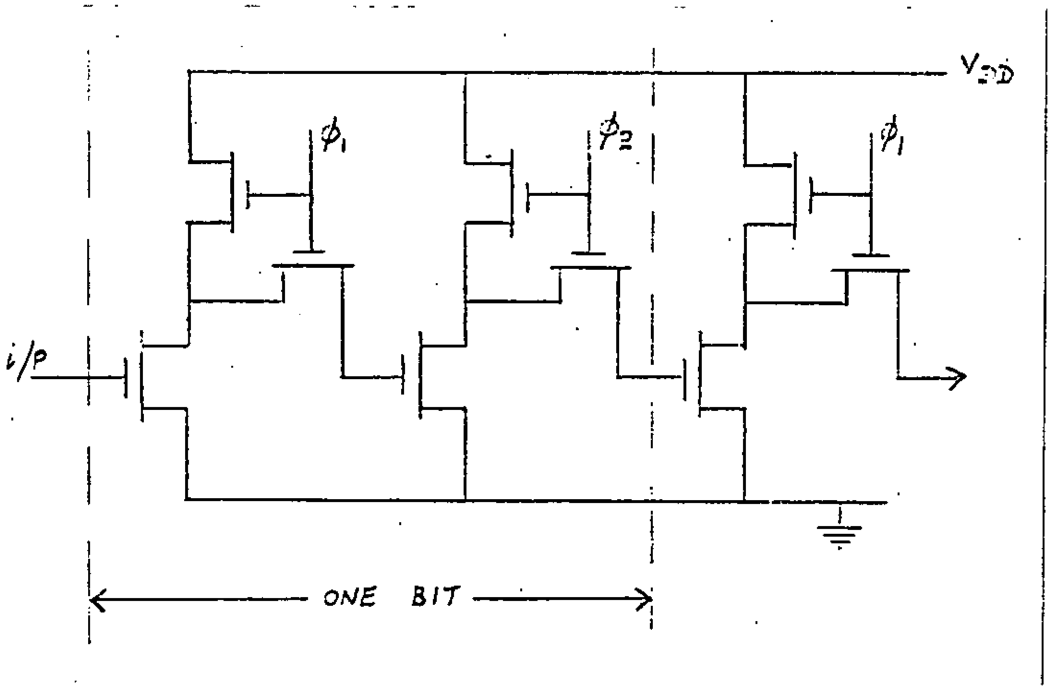


FIG. 4.8. The completed rack for the 4 X 4 chip array simulation.

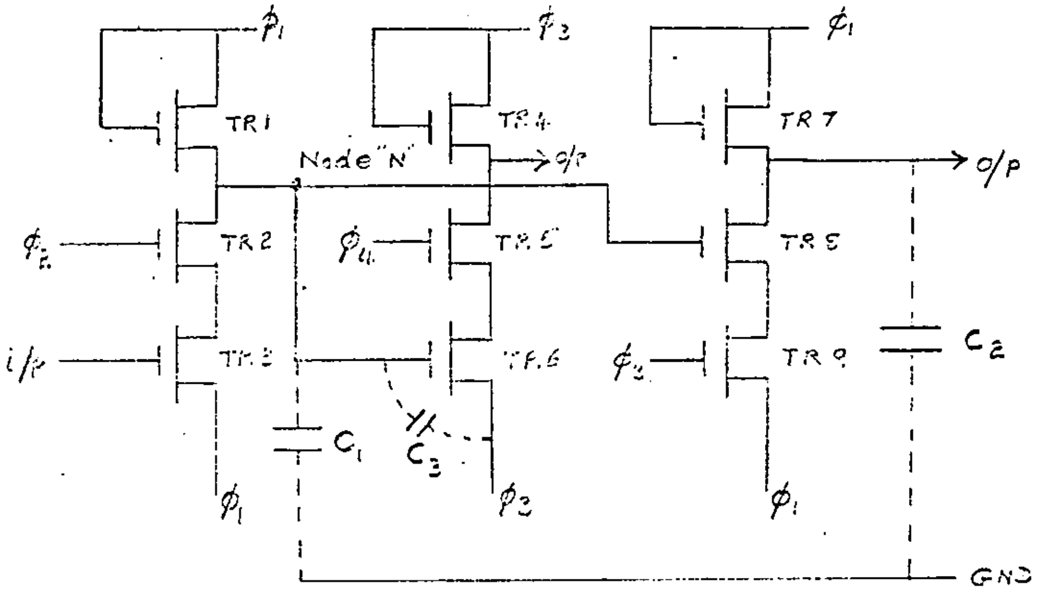


Four-phase dynamic MOS shift register cell.



Two-phase dynamic MOS shift register cell.

FIG. 4.9. Comparison of 4Ø and 2Ø dynamic MOS power/clock rail requirements.



4.10a. Incorrect circuit - causes loss of data.

4.10b. Acceptable circuit for type-1 gate driving types 2&3.

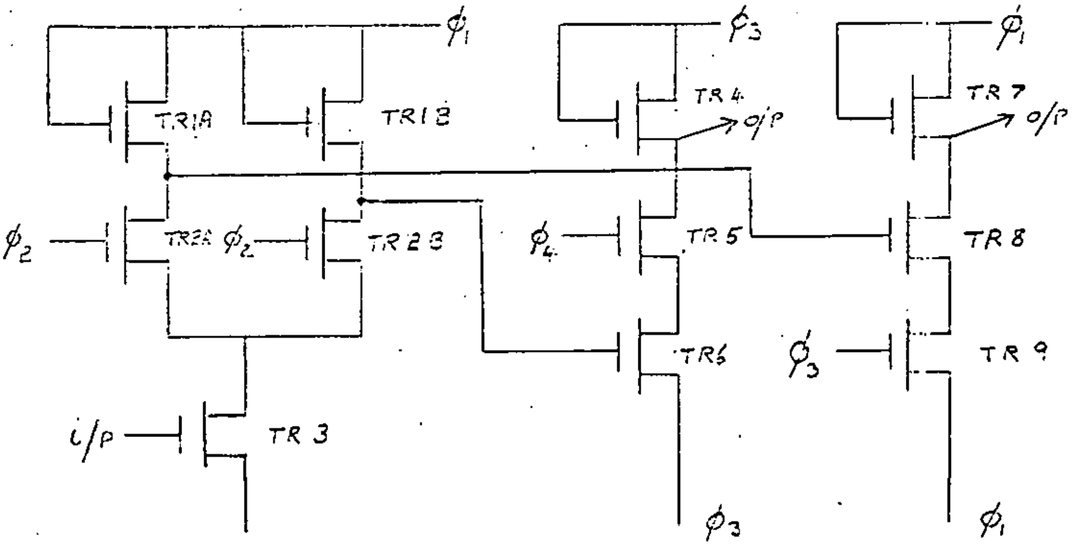


FIG. 4.10. Driver stage for delay-free inverter.

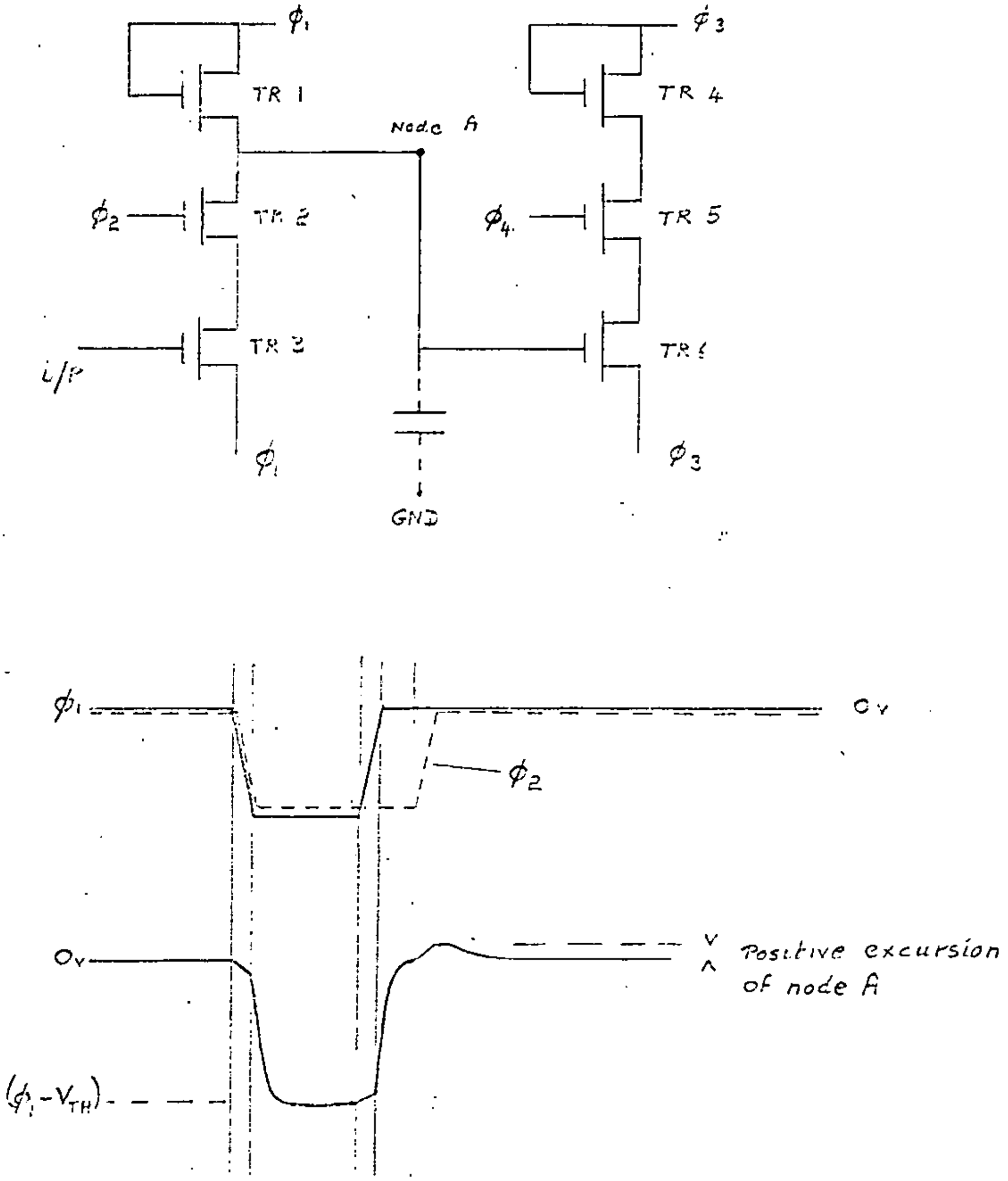


FIG. 4.11. Circuit, clock waveforms and voltage at internal node - showing origin of charge injection into substrate.

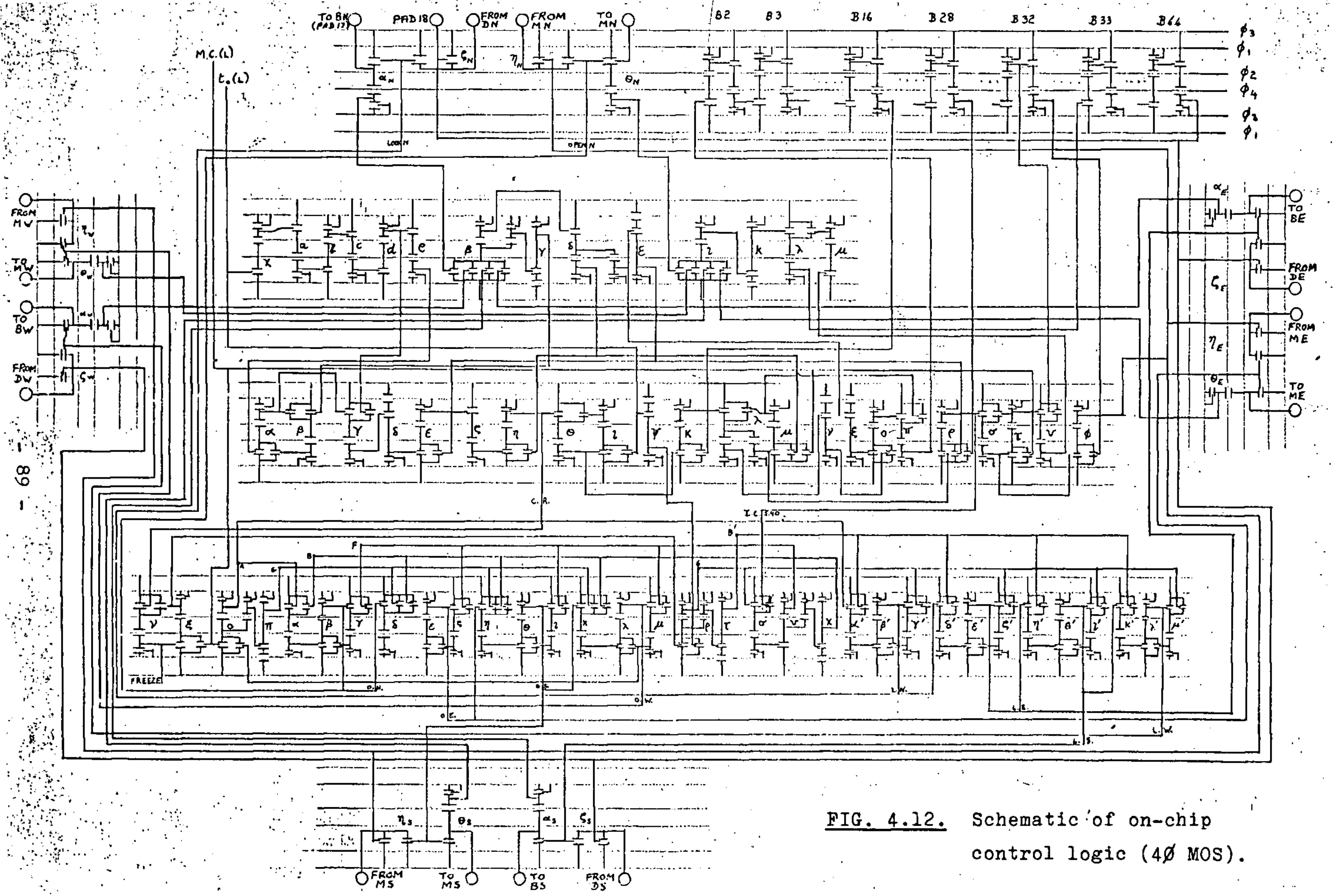


FIG. 4.12. Schematic of on-chip control logic (4Ø MOS).

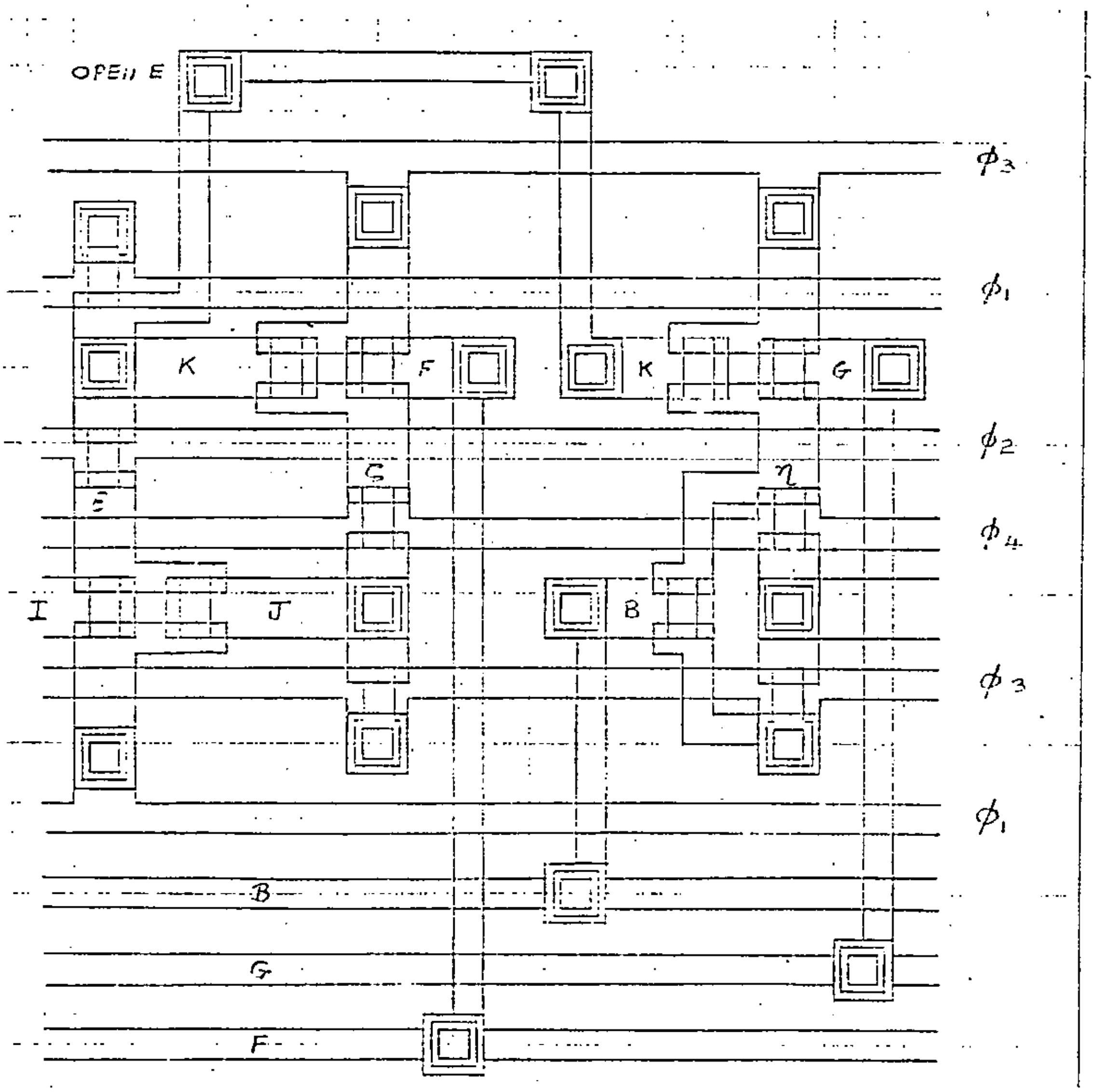


FIG. 4.13. Typical section of layout of on-chip control logic.

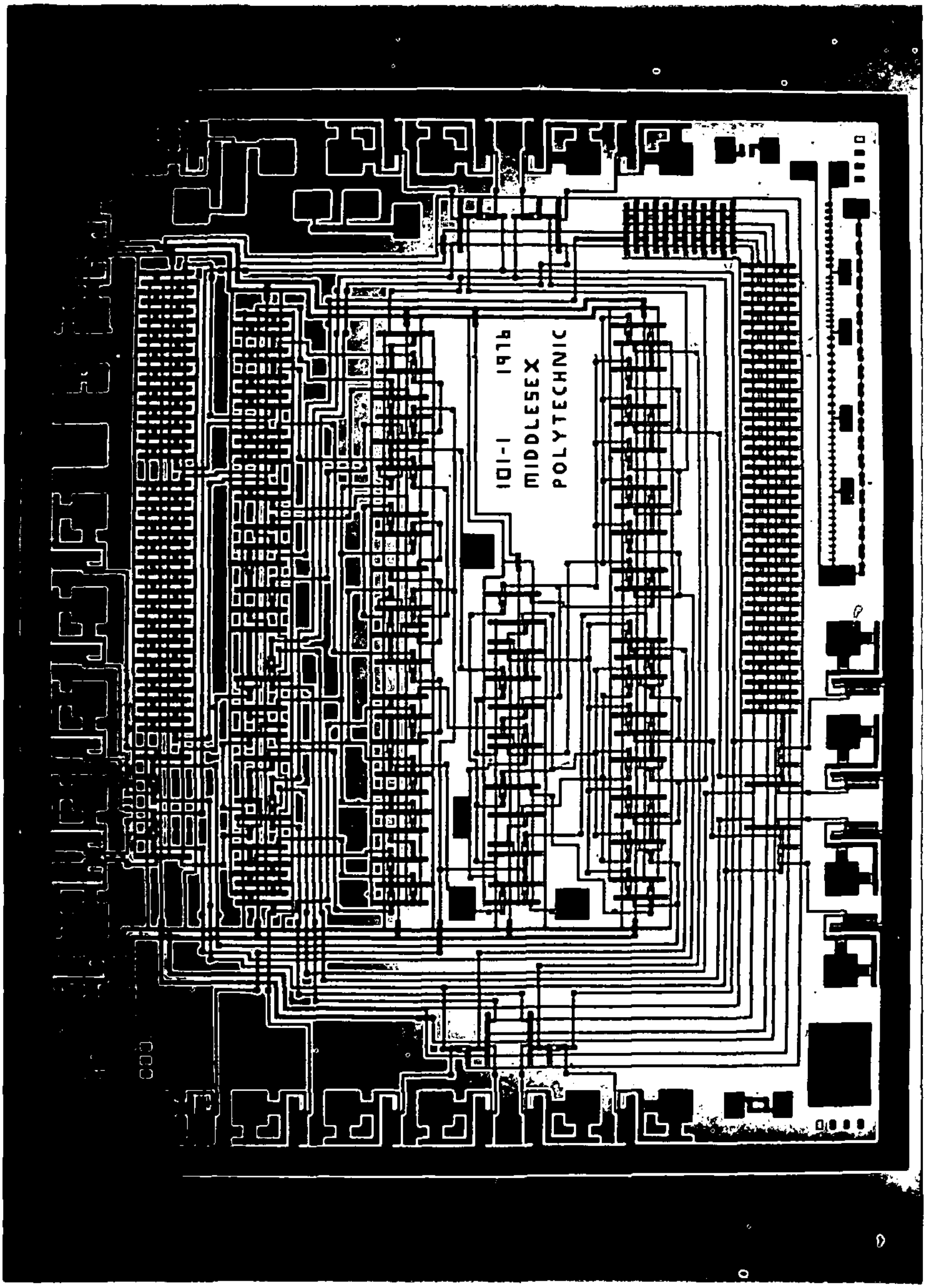


FIG. 4.14. Layout of completed chip.

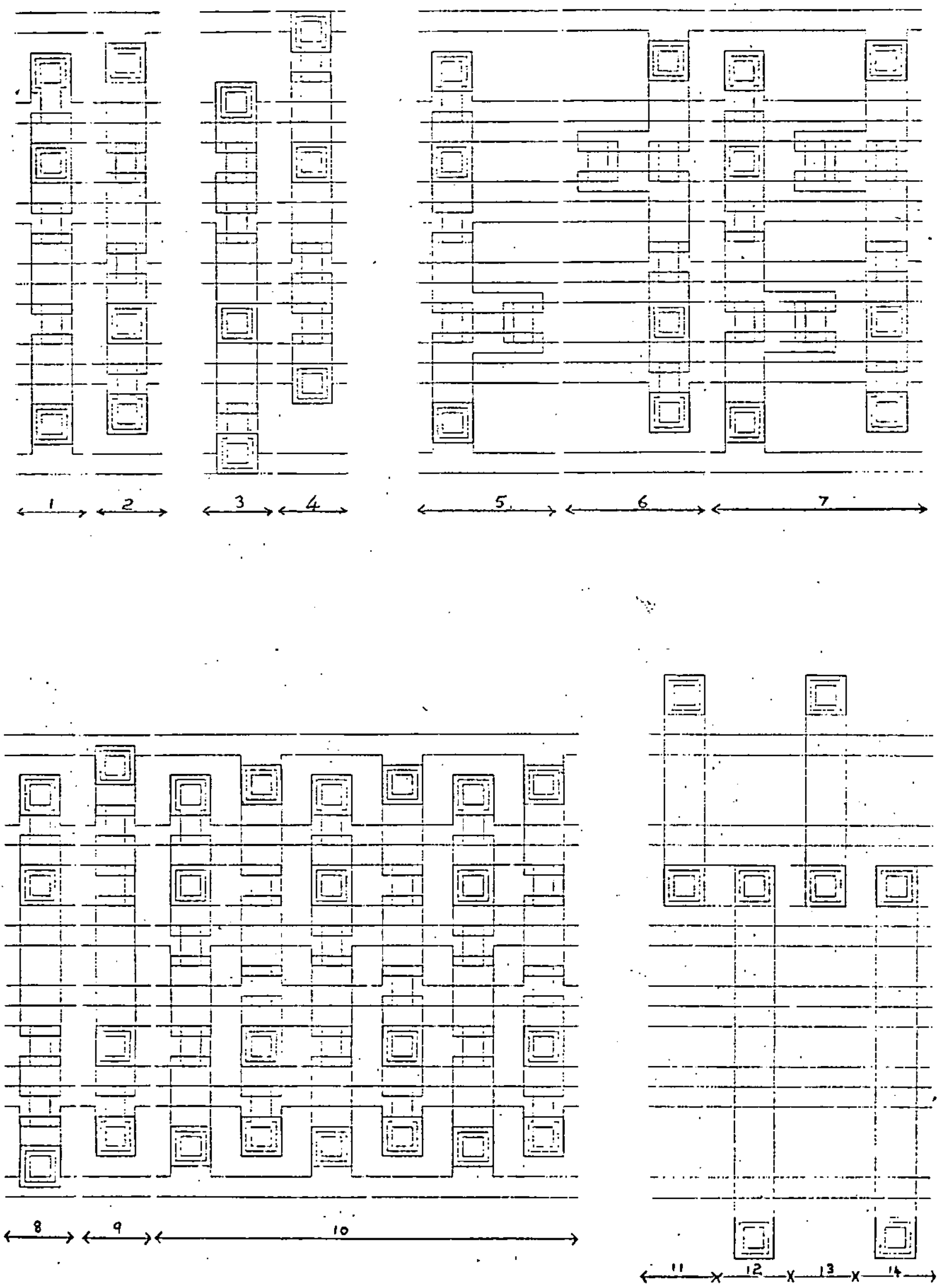
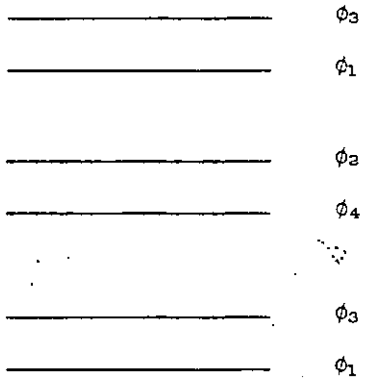


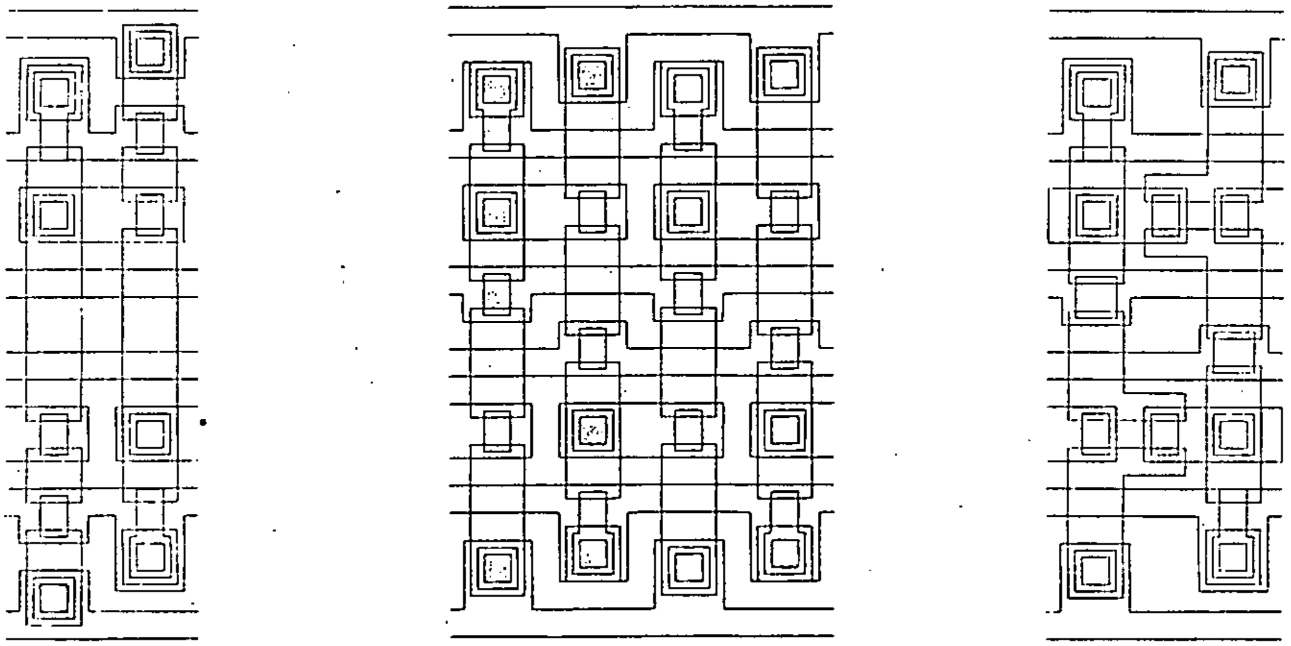
FIG. 4.15a. Examples of in-house computer-assisted layout of 4Ø MOS gates. (Key overleaf)

1. Half bit delay inverter type 1
2. Half bit delay inverter type 3
3. Alternative layout for type 1
4. Alternative layout for type 3
5. 2-input NOR gate type 1
6. 2-input NOR gate type 3
7. Flip Flop (type 1 - type 3)
8. Delay free inverter type 2
9. Delay free inverter type 4
10. Shift register (3 stages)

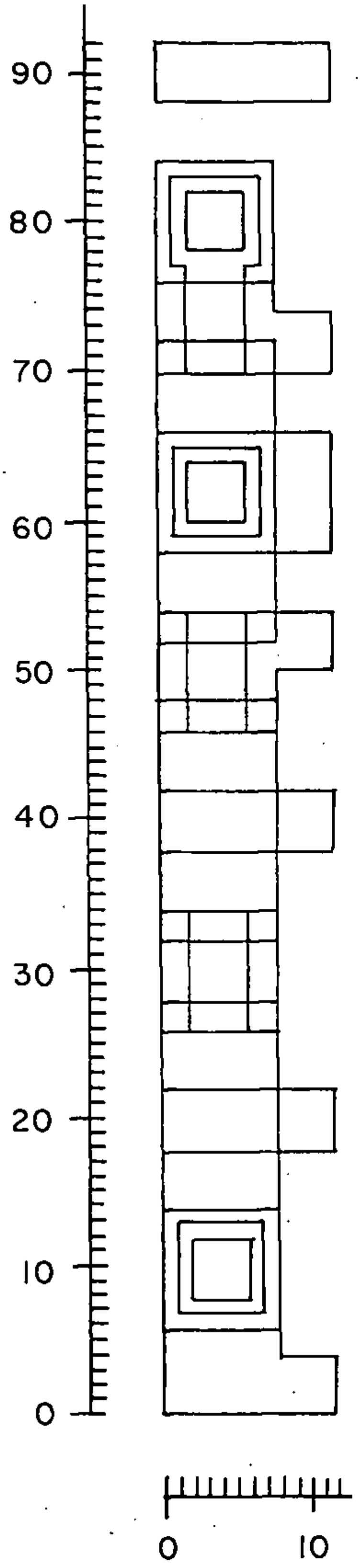
11-13 Examples of the (many) tag types required.

<u>Layout of clock lines</u>





<u>FIG. 4.15b</u>. Further in-house computer-assisted layouts.



```
"JOB" MPRCA;
"TITLE" M.O.S. TYPE 1;

 1    "NEWGROUP" INVTR 1;
 2    "RECT" (1) 0,6:8,22;
 3    "RECT" (1) 0,32:8,16;
 4    "RECT" (1) 0,52:8,20;
 5    "RECT" (1) 0,76:8,8;
 6    "RECT" (2) 1,7:6,6;
 7    "RECT" (2) 2,26:4,8;
 8    "RECT" (2) 2,46:4,8;
 9    "RECT" (2) 1,59:6,6;
10    "POLY" (2) 2,70:4,7,1,6,-6,
                 -6,1,-7;
11    "RECT" (3) 2,8:4,4;
12    "RECT" (3) 2,60:4,4;
13    "RECT" (3) 2,78:4,4;
14    "POLY" (4) 0,0:12,4,-4,10,
                 -8,-14;
15    "RECT" (4) 0,18:12,4;
16    "RECT" (4) 0,26:8,8;
17    "RECT" (4) 0,38:12,4;
18    "POLY" (4) 0,46:8,4,4,4,-12,
                 -8;
19    "RECT" (4) 0,58:12,8;
20    "POLY" (4) 0,70:12,4,-4,10,
                 -8,-14;
21    "RECT" (4) 0,88:12,4;
22    "ENDGROUP";
```

FIG. 4.16. Layout and CAMP program for 4∅ MOS gate, type 1.

5. SIMULATION OF SPIRAL GENERATION IN RANDOM ARRAYS

This chapter describes the computer simulations of the building of chains of good chips in flawed arrays. It is seen how a consideration of the limitations of algorithms for spiral generation led to the proposal of a toroidal array structure. Results of such simulations are presented.

The main features characterising alternative approaches to the creation of serial memories in iterative chip arrays are:-

i) the spatial configuration of devices which may input signals to or receive output signals from a given device, called its neighbours and

ii) the sequence of selected input and output connections as an assembly of devices is configured.

A schematic of the algorithm is given in Fig. 5.1; the growth of a spiral in each array type studied in this section is indicated in Fig. 5.2.

The algorithm names adopted in these studies are historical (rather than logical) but are mnemonic.

C1 is Catt's original algorithm[10] on the square array.

M2 is the second algorithm investigated on the same square array (M ≡ Middlesex). This is an 8 nearest-neighbour algorithm.

T1 is the C1 algorithm on a toroidal (square) array.

M2T8 is the M2 algorithm (8-way) on the toroidal (square) array.

Q2 is the M2 algorithm on a quincunx array.

HX is an extension of the C1 algorithm to a hexagonal (six nearest neighbours) array.

5.1 SQUARE ARRAYS.

In order to compare various algorithms it is necessary to set some form of benchmark for comparison. A target of a 128-chip spiral in an array of 400 chips was selected (somewhat arbitrarily) as such a standard for assessment. Initial studies in this area were made by laboriously plotting random chequer-board patterns as shown in Fig. 5.3, using random number tables (e.g. 0-2 for black, 3-9 for white generates, on average, a 70% yield array). The sizes of groups of "good chips" were also noted, realising that this represented a maximum possible spiral length as the chain might well not be able to utilise all the good chips within the largest group. A computer program was later developed to generate and plot such arrays by look-up of random number tables. This was then extended to produce the required random numbers using the computer and to generate a chain of "good circuits" in the manner required by the particular algorithm. This procedure, detailed in the computer program of Appendix II, is now summarised with reference to Fig. 5.4.

A random array of good and bad chips is generated to a target yield by assigning to each site on the 20 x 20 matrix a random number in the range 00-99. For a target yield of, say, 73 percent the numbers 00-72 represent good chips and 73-99 represent bad ones. A check is kept on the frequency distribution of the numbers so generated to ensure that no detectable order is built into the array. The nature of this check is as follows.

It may be expected, on average, that four numbers will occur either less than once (i.e. not at all!) or greater than 8 times when selecting 400 numbers at random from the range 00-99, the general form of the distribution for each number being as shown in Fig. 5.5. The frequency of occurrence of each number is plotted at the top of Fig. 5.4 (direction of increasing frequency is downwards), the horizontal lines representing frequencies of one-half and $8\frac{1}{2}$ so that, on average, four peaks or troughs may be expected outside these limits. It is accepted that this is not a rigorous test of randomness of the array so plotted but - as already noted - the distribution of bad chips on actual slices is not truly random either.

Having generated the required 20 x 20 pseudo-random array the program then attempts to generate a spiral of good chips on the array, bypassing faulty ones and back-tracking out of blind alleys as necessary until either the required 128-chip spiral has been generated or the spiral has failed to reach the 128-chip target in the preset limit on the number of trials. The maximum and final lengths of unsuccessful spirals are also noted, as illustrated in (the earlier plots of) Fig. 5.6(b), for which a key is presented in Fig. 5.6(a).

It will be observed on these plots that the spiral starts at the centre of the array; if the 11, 11 point is a bad chip (co-ordinates related to origin at bottom left hand corner of the array) the simulation is wasted. The program was later developed to permit the starting point to wander on the array until a good chip is found. If this first chip happens to be walled in by faulty ones, however, the run will terminate. The spiral progresses in a clockwise direction until it reaches the array edge when, because it is always attempting to turn clockwise (and hence off the array), it proceeds in retrograde direction, hugging the array edge. Note that the computer has no prior knowledge of either the fault distribution on the array or the position of the spiral in relation to the array boundaries - thus replicating the problem observed by chip Z in constructing a spiral. Every new chip is added only after testing for a faulty chip and an array boundary, as would be the case when generating a spiral on an actual wafer; the implications of this point in limiting algorithm performance are discussed in Section 6.1.

These studies were then extended to arrays on which the spirals were permitted to attain their maximum possible length and a range of target lengths.

Having verified the spiral simulation procedure by analysis of plotted output these array plots were then abandoned in favour of a listing of the salient data summarising each spiral's performance as listed in Table 5.1. Six thousand simulations of slices with pseudo-randomly distributed faulty devices were carried out at yields ranging from 40% to 100%.

Results of these trials are presented and analysed in Section 5.4.

Type of algorithm (e.g. C1)

Starting number of random number generator

Dice yield on array

Percentage of good chips used

Percentage of total array chips used in spiral

Final spiral length

Maximum spiral length

Total number of chip tests to final length

Total number of chip tests to 128 chip length

Total number of chip tests to maximum length

Total number of data times to final length

Total number of data times to 128 chip length

Total number of data times to maximum length.

Table 5.1

5.2 HEXAGONAL ARRAYS

The 4-way algorithm computer program was extended to cover the hexagonal array; this, together with its flowchart, is detailed in Appendix III. A plot from this program is illustrated in Fig. 5.7.

In this format the good chips are represented by hexagons and the bad ones by six-pointed stars. The spiral is launched near the centre of the array and proceeds in an anticlockwise direction, again bypassing faulty chips, backing out of blind alleys and turning retrograde at the array edge. On this plot are also indicated,as short spurs on the spiral leading from the centre of each good chip, directions which have been accessed and rejected - either (i) because a chip was faulty, or (ii) it had already been included in the spiral, or (iii) it was outside the array boundary.

As for the square array, these trials were initially restricted to a target length of 128 chips and were then extended to various other target lengths (including the maximum possible length obtainable by the spiral). Results from six thousand such simulations are summarised in Section 5.4.

5.3 FINITE UNBOUNDED ARRAYS

During the studies of the rectangular and hexagonal arrays it became increasingly apparent that the serious perturbation imposed on the path of the spiral by the array edge caused a major reduction in algorithm performance. In being forced into retrograde progression along the array edge the spiral tends to trap large areas of good chips so that they are no longer accessible to the developing spiral. They will, of course, eventually be accessed as the (unsuccessful) spiral backtracks (eventually to a single chip) but by this time the damage has been done - the algorithm has failed to construct the target spiral. On realising the importance of this limitation the (effective) elimination of the array boundary was proposed as follows.

The wafer may be converted to a finite but unbounded array merely by ensuring that all edge chips have their outputs fed back into inputs of other edge chips. Nothing would be gained by feeding back into the adjacent chip as they are already connected in the array. Linking to the next but one neighbour might seem advantageous until it is realised that the edge output on a chip at (say) the northernmost point of the slice would expect to feed out to the North and hence into the South ("LOOK") input of the next but one chip. This chip would, of course, already have its South input connected into the array and could never accept the data in via its free (northern) input from a North edge chip as it should be looking in the opposite direction at the phase of t_o when the signal appeared at its North input. One could, in principle if not in practice, feed the output of each edge chip into its diametrically opposite counterpart but a more subtle organisation is to convert the array, topologically, into a torus.

To achieve this toroidal configuration (without processing doughnuts of silicon!) the East output and input of each chip along the right-hand edge of the array are arranged to connect directly into the corresponding West input and output of each chip along the left-hand side of the array. This converts the array (conceptually) into a cylinder. The North output and input of each chip along the top edge of the array are then arranged to connect directly into the South input and output of the corresponding chip along the bottom edge of the array. The array boundaries then become transparent to the developing spiral - as if the array were constructed on the surface of a toroidal crystal of silicon. The procedure is not limited to square arrays;

the hexagonal array, for example, may be similarly connected across the three opposite pairs of edges - without attempting to visualise the geometric form of the figure so constructed! Similarly the connections do not need to be directly across the array; moving to the next row or column when traversing the array, for example converts the array to a helix. A provisional patent has been filed on these structures.

One advantage of such arrays is that the spiral may be launched near the array edge - thus easing the assembly problems - with a less serious effect on its growth potential.

The computer program to construct this toroidal array algorithm - essentially a development of the basic rectangular array program of Appendix II - is not quoted here, but a plot of such a toroidal array is presented in Fig. 5.8. As the interpretation of this plot in the edge region of the array is not immediately obvious it is now explained.

In order that the array edge should appear transparent to the developing spiral, the same cyclic sequence of directions for LOOK/OPEN must continue across the edges of the array; that is, the use of the toroidal facility must not be a "last resort" but must be included in the normal choice of directions. Consider, for example, the situation after the spiral has backtracked to chip number 57 from the large blind alley (bounded by bad chips and the early part of the spiral) in the lower right quarter of the diagram. The next chip incorporated (No. 58) is at the extreme corner of the array. After attempting the normal clockwise turn and finding chip α faulty the spiral then accesses chip β (top right-hand corner of array) to find this also is faulty; chip γ (bottom left-hand corner) is then accessed (again faulty) before the spiral backtracks. At chip 57 the new direction (only one re-try permitted on backtrack) takes the spiral to chip δ on the left-hand edge of the array. Chip δ is marked ◊ , indicating that the spiral has just come onto the edge from the other side of the array. Chip δ , unfortunately, leads only to two more chips so the spiral must backtrack from this blind alley through chips 57, 56 and 55 before the toroidal facility brings the spiral across the array to chip ε from chip 54 and the spiral proceeds normally to chip 99. At chip 99 the spiral, once again, crosses the array edge - chip 100 appearing at the left-hand side. At chip 104 the first attempt accesses the bad chip on the left-hand array edge. The second try attempts to re-enter the spiral running down the right-hand array edge; the third attempt is successful, finding a good chip (due South of chip 104). Note, however, the change in line format of the plot; this reflects the fact that the spiral has attempted to leave the array edge at chip 104 and so chip 105 is included as if the spiral had just come onto this edge of the array. The two lower chips are similarly represented - before being overprinted with the diagonal cross representing chips rejected on backtracking from a blind alley. Chip 106 (due East of 105) is also represented in this way - the program assuming that an array edge has just been crossed - before the spiral continues normally to its completion at chip 128.

As for the C1 and HX algorithms the spirals were initially restricted to a target of 128 chips and then extended to cover various target lengths from 50 to the maximum possible spiral length attainable; the array yield range was, again 40-100%.

Results of six thousand such simulations are presented in Section 5.4.

5.3.1 Extension to Octagonal Arrays

This simulation was not extended to a finite unbounded hexagonal array but proceeded directly to the most powerful practicable algorithm that could be envisaged. This was an eight-way algorithm (with toroidal facility) where each chip may access both orthogonal and diagonal nearest neighbours - as King's move in chess. Such an array is illustrated in the computer simulation of Fig. 5.9; it could be constructed from identical chips, without wasted space, as sketched in Fig. 5.10. Results of 6,000 simulations of this 8-way toroidal algorithm are compared with the more basic algorithms in the following section.

5.4 COMPARISON OF ALGORITHM PERFORMANCE

It has been noted in Section 3.4.3 that calculation of array yield as a function of chip size and defect density is possible if assumptions are made as to the nature of defects. However it is more realistic to obtain this relationship empirically; it is then found to depend on the chosen technology and design rules and is typically of the form quoted in Fig. 5.11. This plot is from data obtained during discussions with the industry and relates to an MOS type of process.

While W.S.I. arrays may eventually be manufactured on rectangular wafers it is likely that the conventional technology will be applied in the first instance. To assist in relating the algorithm simulations to actual slice arrays, Fig.5.12 plots the number of complete square chips on wafers of 2", 3" and 4" diameter. These curves are interpolated between the points plotted and assume accurate alignment of the array such that the slice diameter lies along a grid line in both X and Y directions.

In order to obtain a direct comparison between the various algorithms it is desirable to select arrays of similar size for each algorithm. However, a further consideration of major importance for a fair comparison is that the array structure should fit the algorithm - that is at high yield levels the algorithm must be able to fully utilise the array. Failure to comply with this requirement is illustrated by the poor performance of the M2 algorithm (see later). On a square array its performance (with eight nearest neighbours) is markedly inferior to that of the basic C1 algorithm (with only four nearest neighbours). On transferring it to the quincunx array, however, its performance (Q2) attains the high level to be expected from an 8-way algorithm. (Note also the effect of the elimination of the array boundary (M2T8) in optimising an array for a particular algorithm).

In striking a balance between these two requirements of constant array size and optimised array shape, the hexagonal algorithm is put at a slight disadvantage (only 392 chips on an approximately circular array) while the quincunx algorithm has 421 chips compared with the 400 chips of the C1, T1, M2 and M2T8 algorithms.

Criteria of practical importance in assessing algorithm performance include the area of device required to implement the on-chip control logic, the variation of spiral length as a function of device yield, the yield required to achieve particular target lengths reliably and the time required to carry out the test and spiral generation.

Fig. 5.13 illustrates the sequence in which a device's neighbours are tested during spiral development for each algorithm. The data are received from the device marked "0" (the current penultimate device in the chain) and the central device (currently last in the chain) tests its neighbours in the sequence indicated until a good device is found and added to the chain.

All the algorithms studied were designed to assemble devices into a tight spiral, bypassing faulty devices as described above. Although the performance of an algorithm may be improved by permitting non adjacent chips to be accessed to extricate the spiral from blind alleys (as discussed in Section 3.4.4), it is desirable to reduce interconnection crossover problems by requiring a device's neighbours to be physically adjacent to that device. This restriction was applied to all the algorithms studied.

In assessing the relative performance of different algorithms one may either set an arbitrary target length for the spiral to achieve or, alternatively, require the algorithm to produce the longest possible spiral length on the array. It is probable that practical applications of such memory devices will require a fixed length for the spiral and, with this point in view, the initial results related to a fixed (arbitrary) spiral length of 128 chips. Results of several hundred such simulations of the C1, HX and T1 algorithms are presented in Figs. 5.14, 5.15, 5.16. Each data point represents a minimum of twenty spirals and the error bounds are calculated for 95% confidence limits. The best smooth curve fit to the data is also drawn; these curves are superimposed on Fig. 5.17.

The sharp transition of the C1 algorithm from short spirals at $\lesssim$ 65% good chips to long spirals at $\gtrsim$ 75% yield compares closely with Manning's result[34] quoted in Section 2.2.4.

The mean values of the maximum spiral lengths attained in the 6,000 simulations per algorithm are plotted in Fig. 5.18 as a function of percentage chip yield in the arrays. Fig. 5.19 records the percentage of good array chips used in the spiral as a function of percentage chip yield for each algorithm.

Fig.5.20 illustrates the effect of varying the target length for the spirals by plotting the percentage yield necessary for 95% of the spirals to reach the target length against that target length.

Fig. 5.21 relates the test time required for the assembly of spirals to the percentage yield on the slice for spiral lengths of 50,100 and 200 chips.

5.4.1 Attempts at Curve Fitting of Empirical Data

The general form of the distributions presented in Fig. 5.17 is a monotonically increasing function. At zero yield there is obviously zero chance of any spirals reaching any non-zero target length whereas at 100% yield a good algorithm on an optimised array should be able to incorporate all the chips in the spiral. Such an idealised characteristic of an algorithm is sketched in Fig. 5.22. The ordinate of the line $x' \, x''$ must depend on the particular algorithm (primarily number of nearest neighbours and transparency of array boundary) and on the target length for an array of given size. In principle any continuous distribution can be represented by an infinite polynomial series so the possibility was considered of deriving the characteristic equations for the algorithms in polynomial form. Exact solution was expected to be impossible as this type of distribution almost certainly involves exponential terms - there is, at least superficially, some resemblance of these sigmoid characteristics to the Fermi function; to approximate this by a polynomial series would require considerably more than the maximum of a 10th order polynomial available on the Polytechnic computer program.

Results of these trials were, however, disappointing. The output for a low order polynomial (degree 3) showed considerable overshoot of the 0 and 100% abscissae as shown in Fig. 5.23. Even the addition of heavily weighted data points failed to prevent substantial overshoot. When given a greater degree of freedom (up to 10th order polynomial) the overshoot was slightly improved but the plot tended to follow individual data points rather than the required best smooth curve through the data as shown in Fig. 5.24.

At this point these trials were postponed - pending the availability of a more general curve fitting program - to concentrate on more urgent aspects of the project and have not yet been resumed.

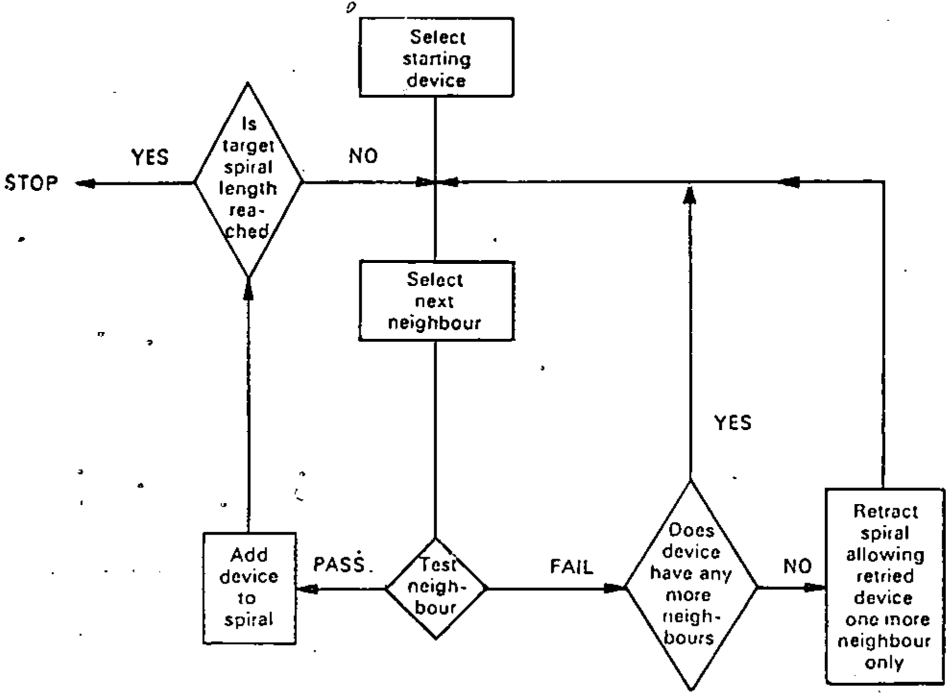
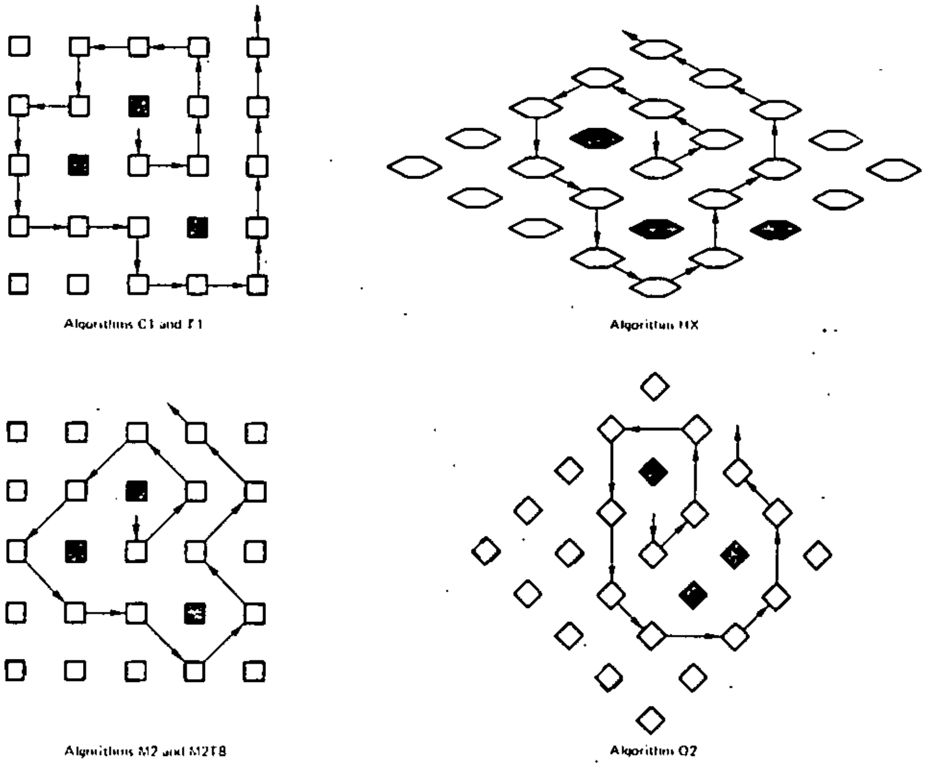

Select starting device
Is target spiral length rea-ched
YES
STOP
NO
Select next neighbour
YES
Add device to spiral
PASS
Test neigh-bour
FAIL
Does device have any more neigh-bours
NO
Retract spiral allowing retried device one more neighbour only

FIG. 5.1. Schematic of the algorithms for spiral generation.

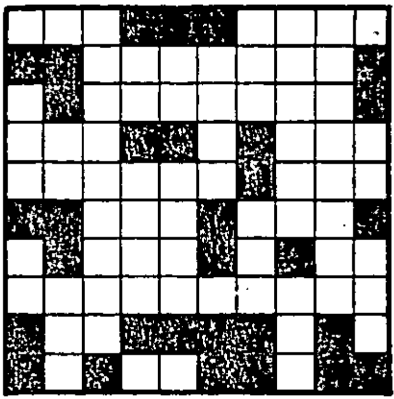

Algorithms C1 and F1

Algorithm HX

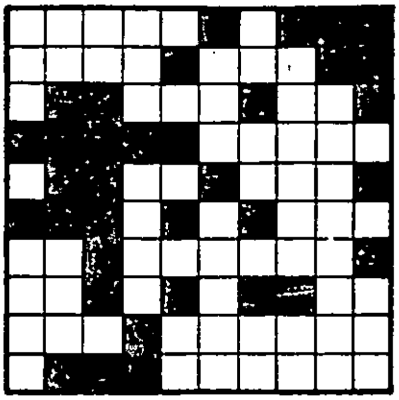
Algorithms M2 and M2T8

Algorithm O2

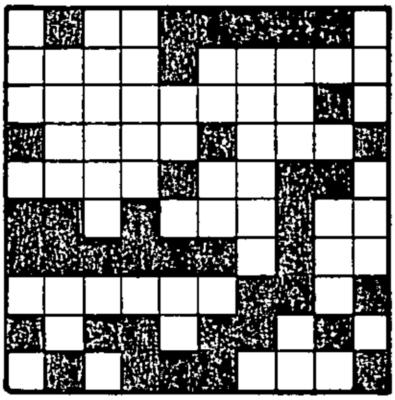
FIG. 5.2. The growth of a spiral, avoiding faulty devices.



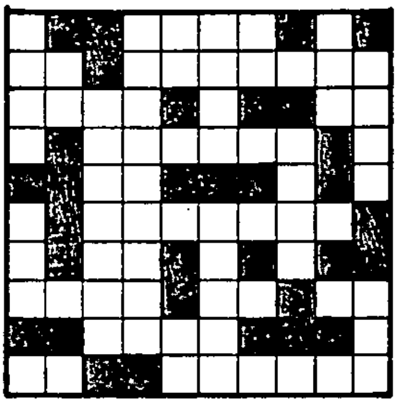
1.



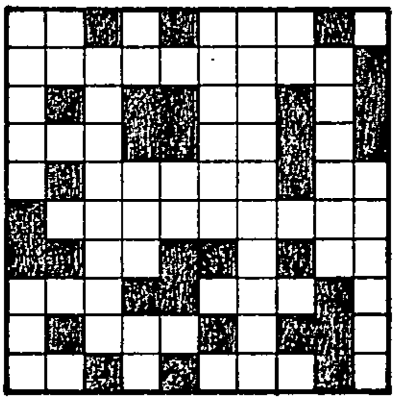
2.



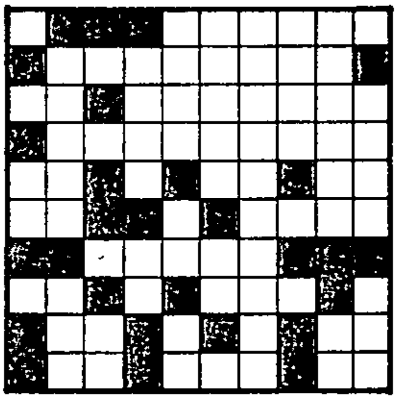
3.



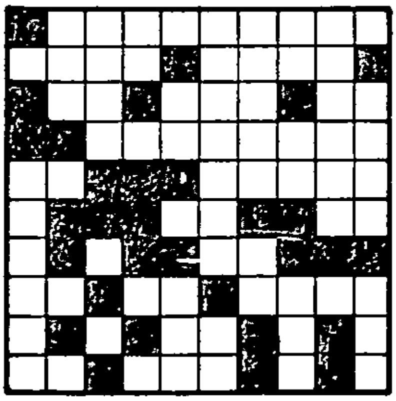
4.



5.



6.



7.

ARRAY No.	% YIELD	LARGEST CLUSTER
1	69	67
2	65	56
3	61	40
4	68	66
5	69	68
6	72	60
7	69	51

FIG. 5.3. Random arrays - 10 X 10 matrix, nominal 70% yield.

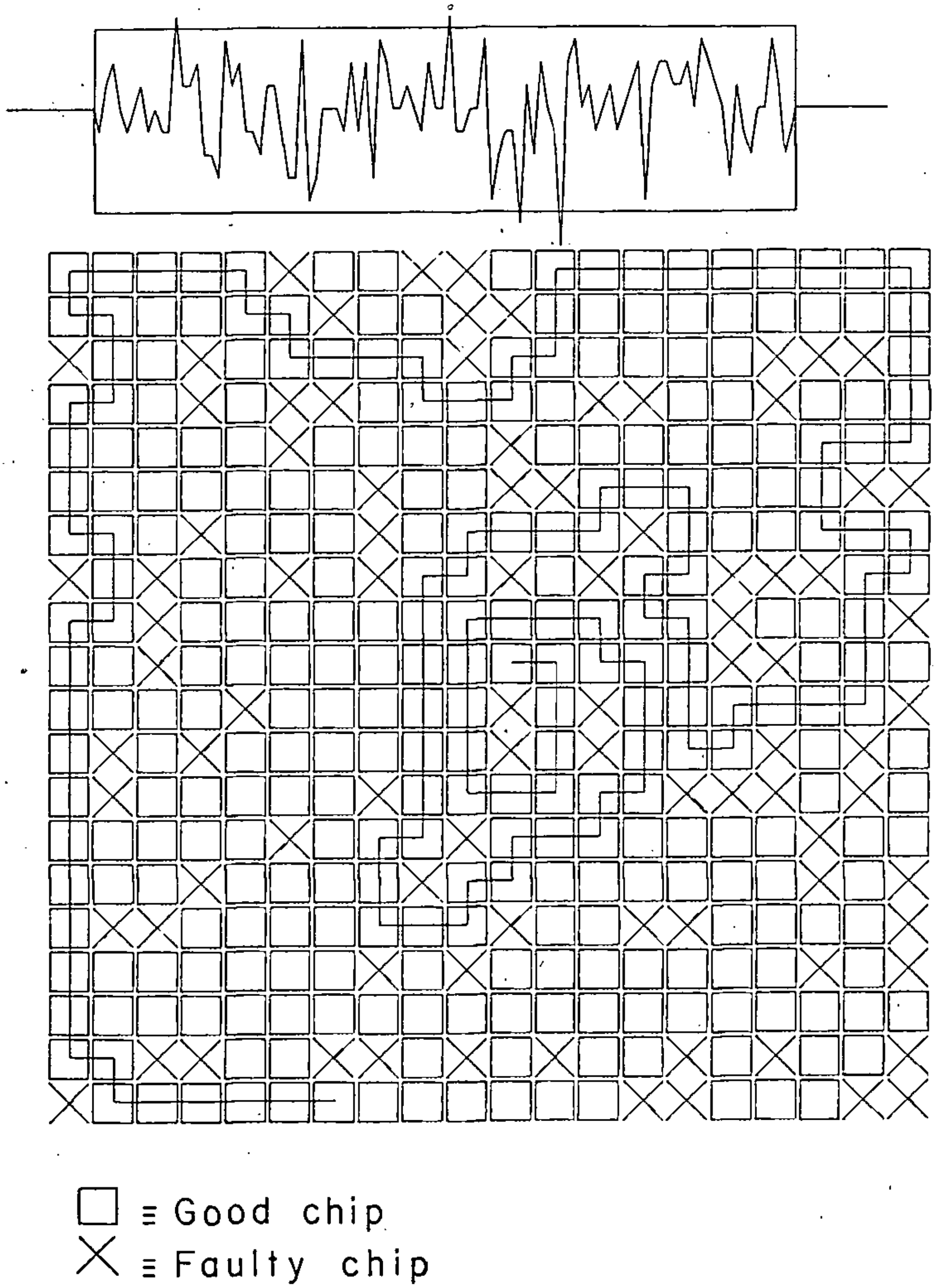


FIG. 5.4. Computer simulation of square array with spiral.

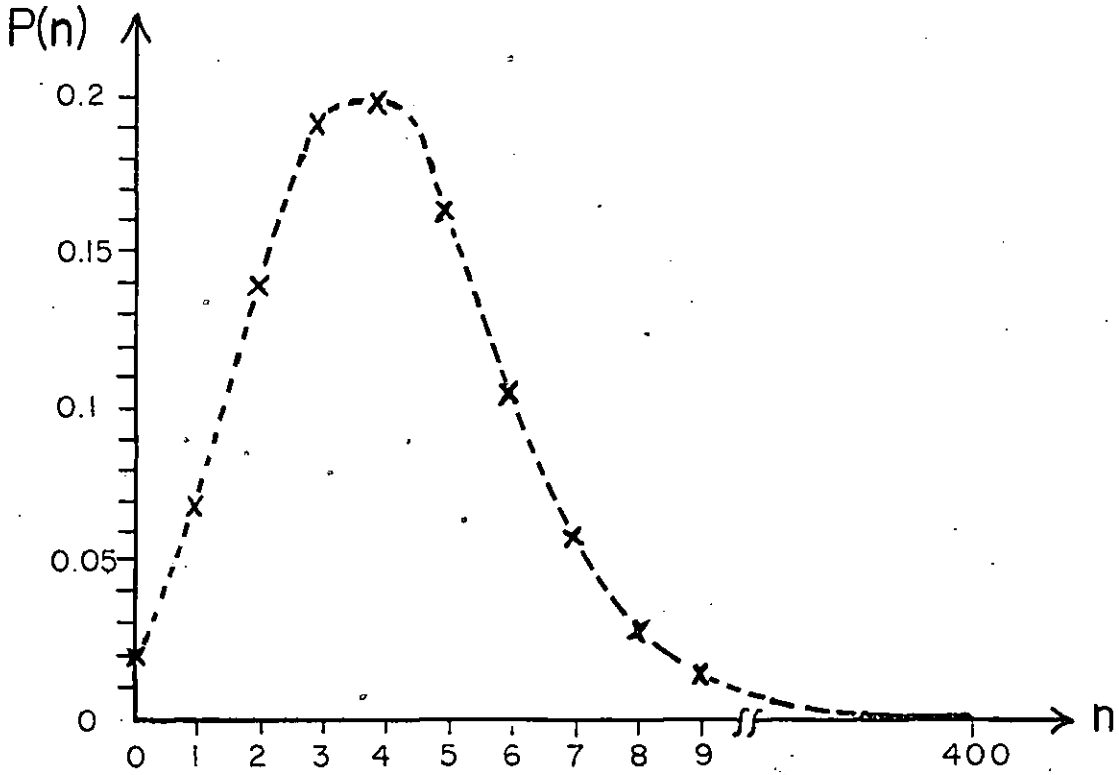
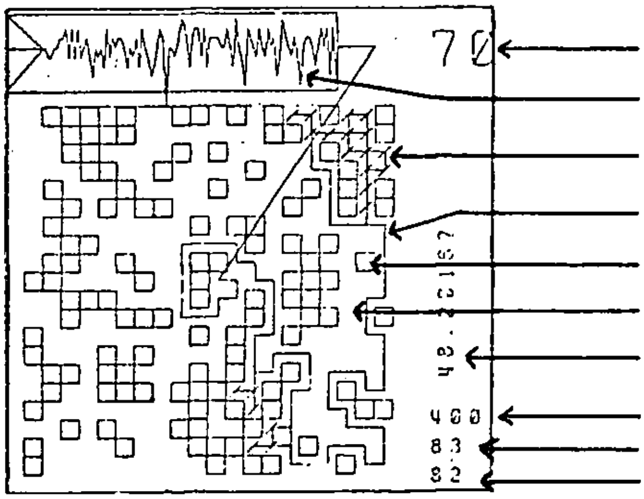


FIG. 5.5. Probability of any particular number in range 00-99 occuring n times in 400 random selections.



target yield (70%).

indication of randomness.

dead-end (good chips rejected).

developing spiral.

bad chip.

good chip.

starting value for random number generator.

number of tests.

maximum length attained.

spiral length at termination of run.

FIG. 5.6. Key to spirals plotted overleaf.

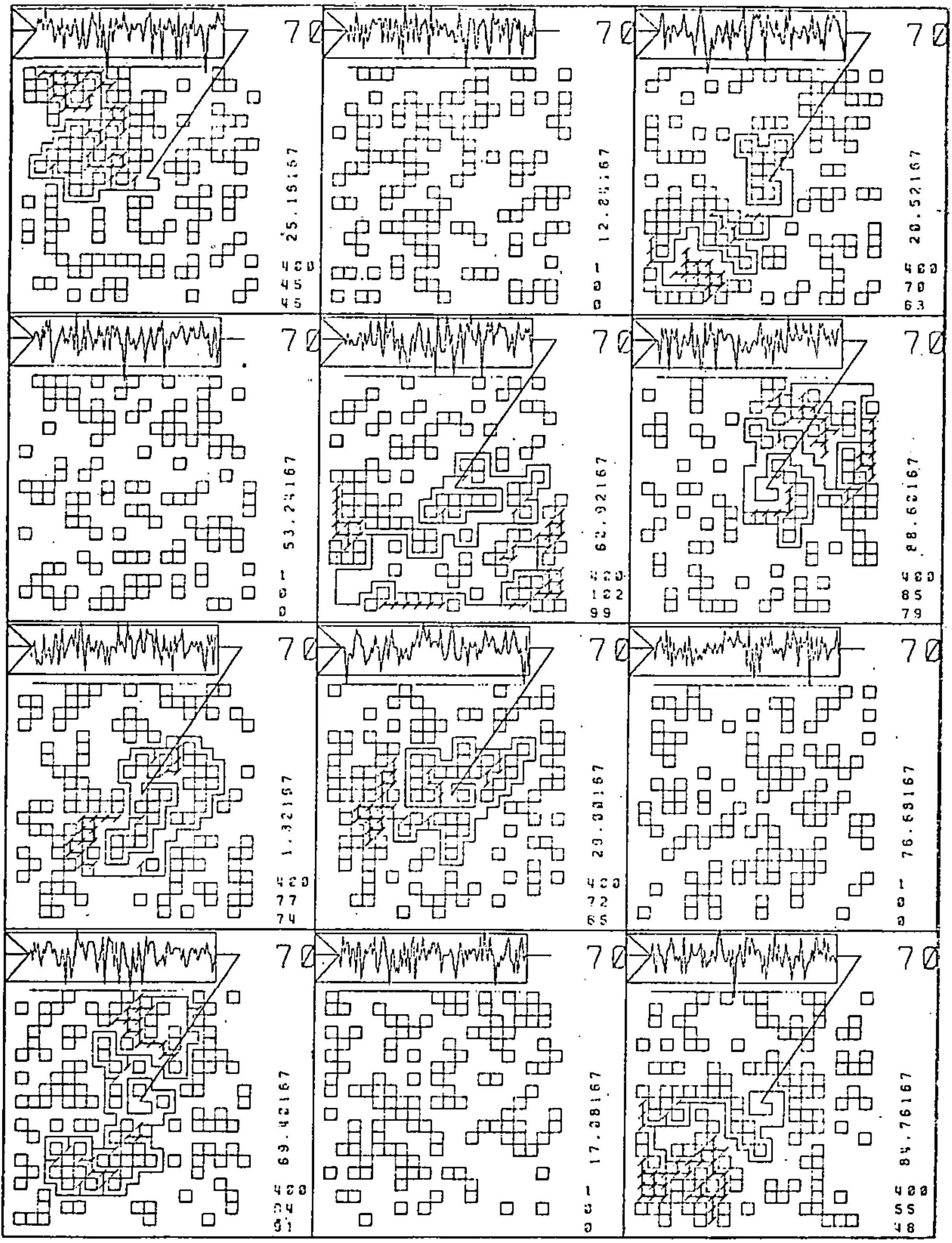


FIG. 5.6. More random arrays with spirals. (Key to interpretation of these is presented on preceding page.)

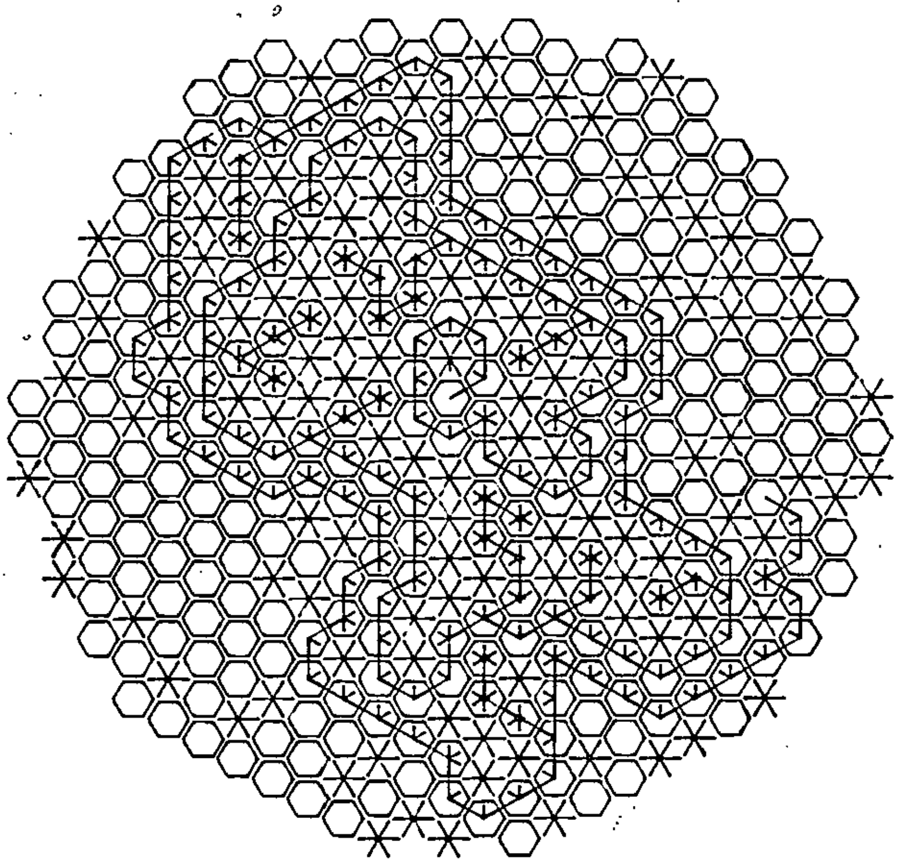
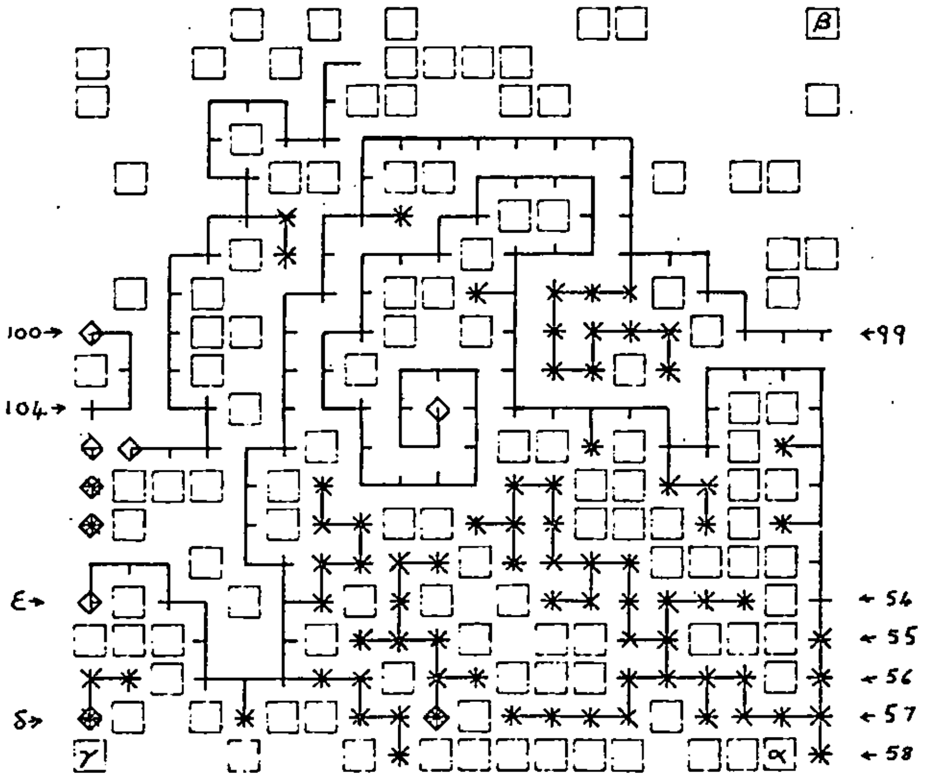
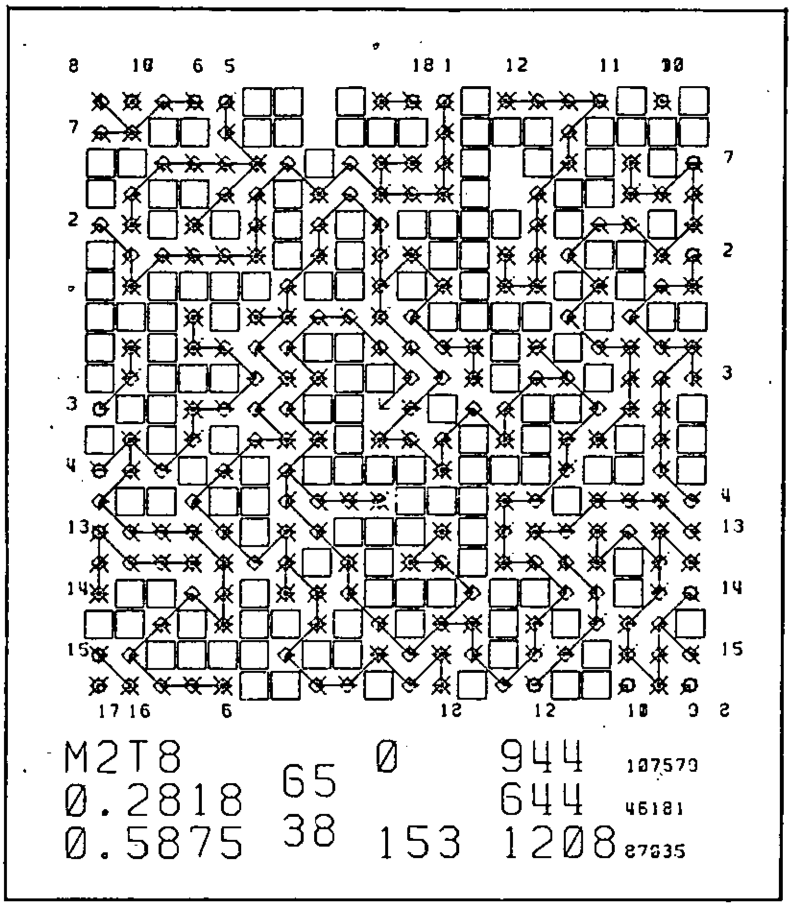


FIG. 5.7. Computer simulation of hexagonal array with spiral.

FIG. 5.8. Annotated plot of toroidal array with spiral.





KEY.		
	M2T8	Algorithm type (toroidal 8-way)
	0.2818	Starting value of random number generator
	0.5875	Dice yield on array (58.75%)
	65	Percentage of good chips used in spiral
	38	Percentage of 20 x 20 chips used in spiral
	0	Final length of spiral (backtracked to input chip)
	153	Maximum length attained by spiral
	944	No. of chip tests to attain 153 - chip spiral
	644	No. of chip tests to attain 128 - chip spiral
	1208	No. of chip tests to attain final length
	107579 46181 87035	Total number of shift register cycle times required to form final, 128- and 153- chip spirals (e.g. 128 chips, 1k bit, 1MHz b.r.f. → 46.2 secs.)

Numbers around array edge indicate the sequence in which spiral leaves the array and rejoins the opposite edge.

FIG. 5.9. Computer simulation of spiral formation with 8-way algorithm on toroidal array.

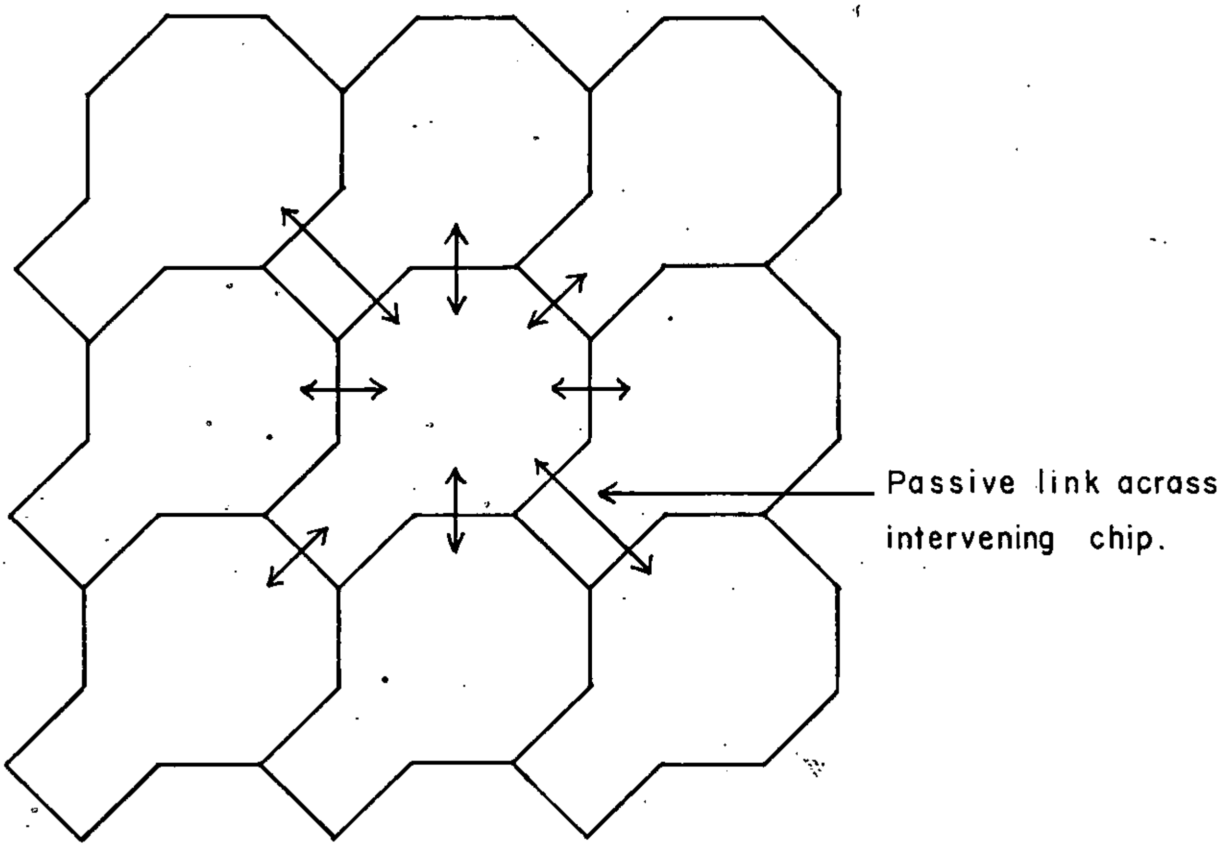


FIG. 5.10. A possible 8-way chip structure.

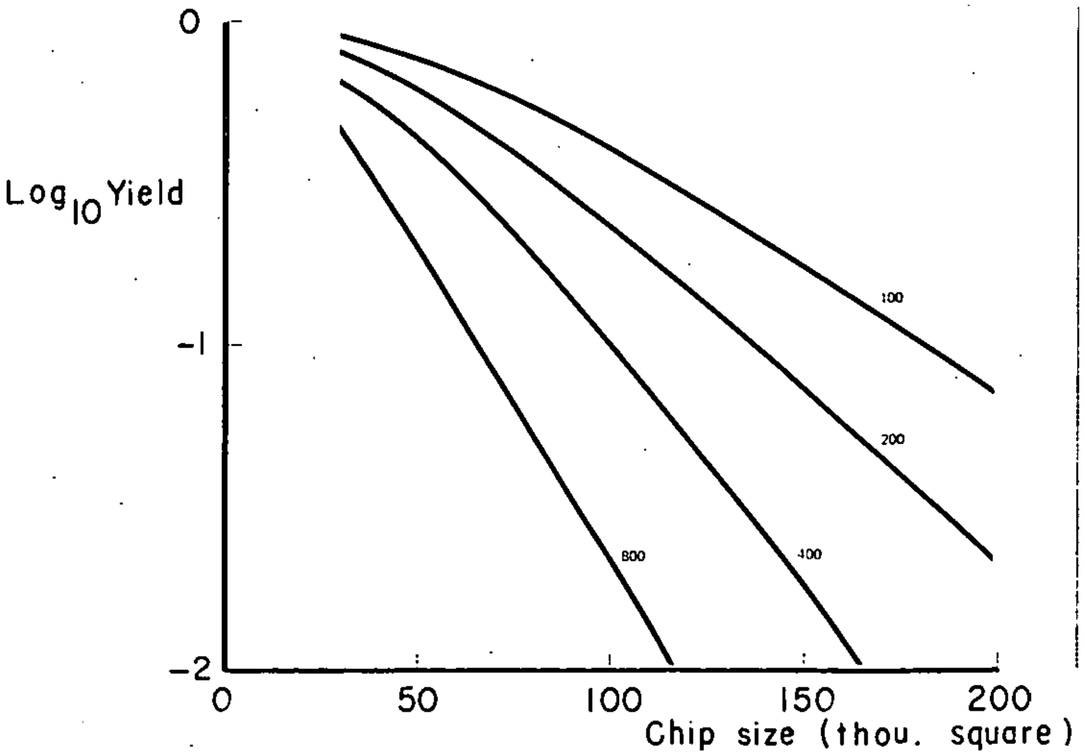


FIG. 5.11. Yield vs. chip area parametrised by defect density (no. per square inch).

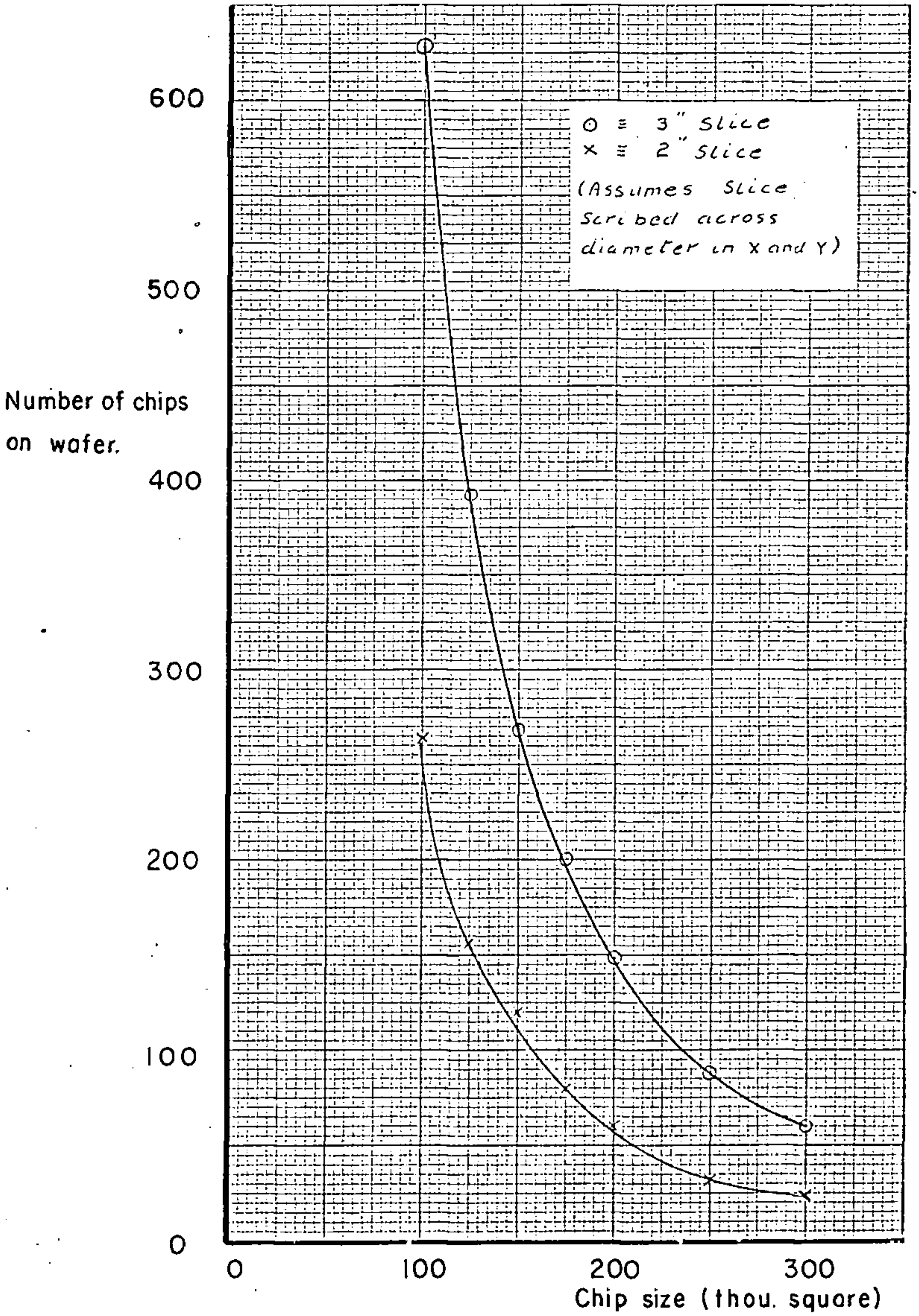


FIG. 5.12. Number of chips on wafer vs. chip size.

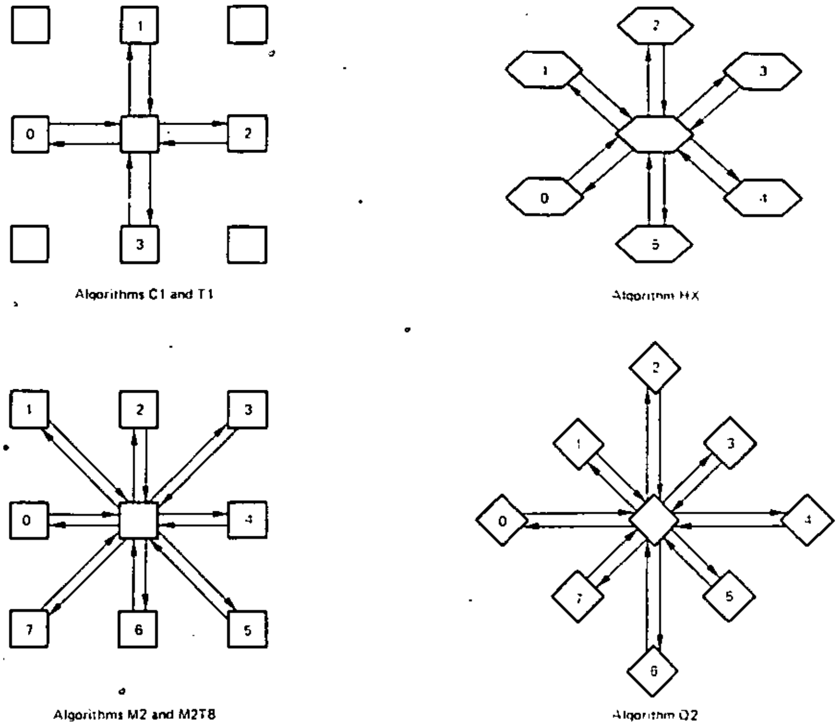


FIG.5.13. The sequence in which neighbours of the central device are tested.

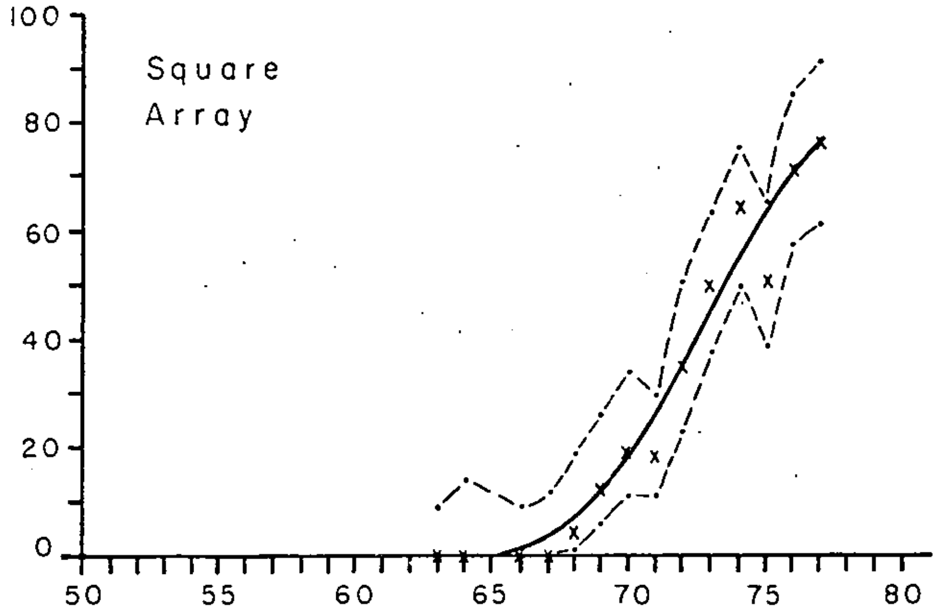


FIG. 5.14. Percentage of spirals reaching target of 128 chips on array of 20 X 20 chips (vertical) vs. chip yield (%) for square array.

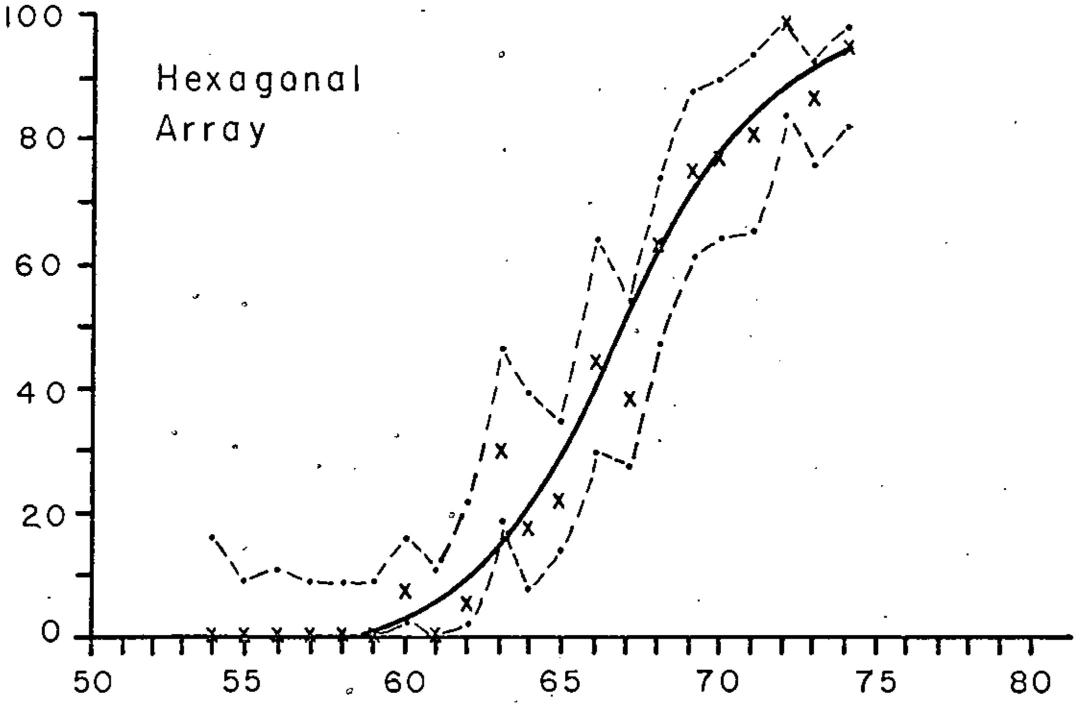


FIG. 5.15. Percentage of spirals reaching target of 128 chips on array of ~400 chips (vertical) vs. chip yield (%) for hexagonal array.

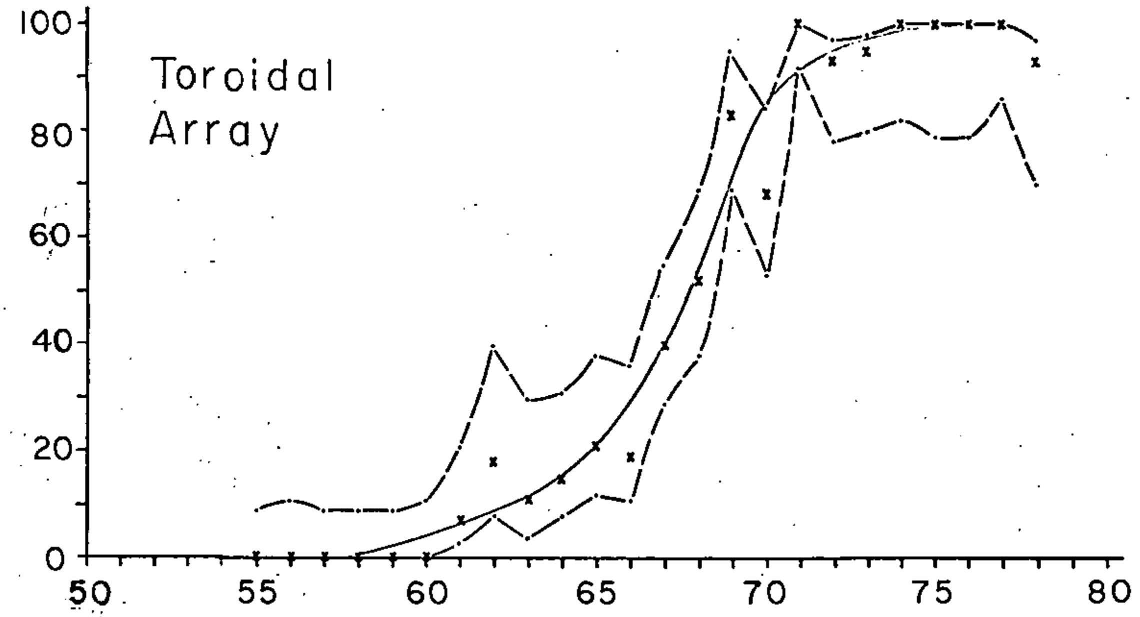


FIG. 5.16. Percentage of spirals reaching target of 128 chips on array of 20 X 20 chips (vertical) vs. chip yield (%) for toroidal array.

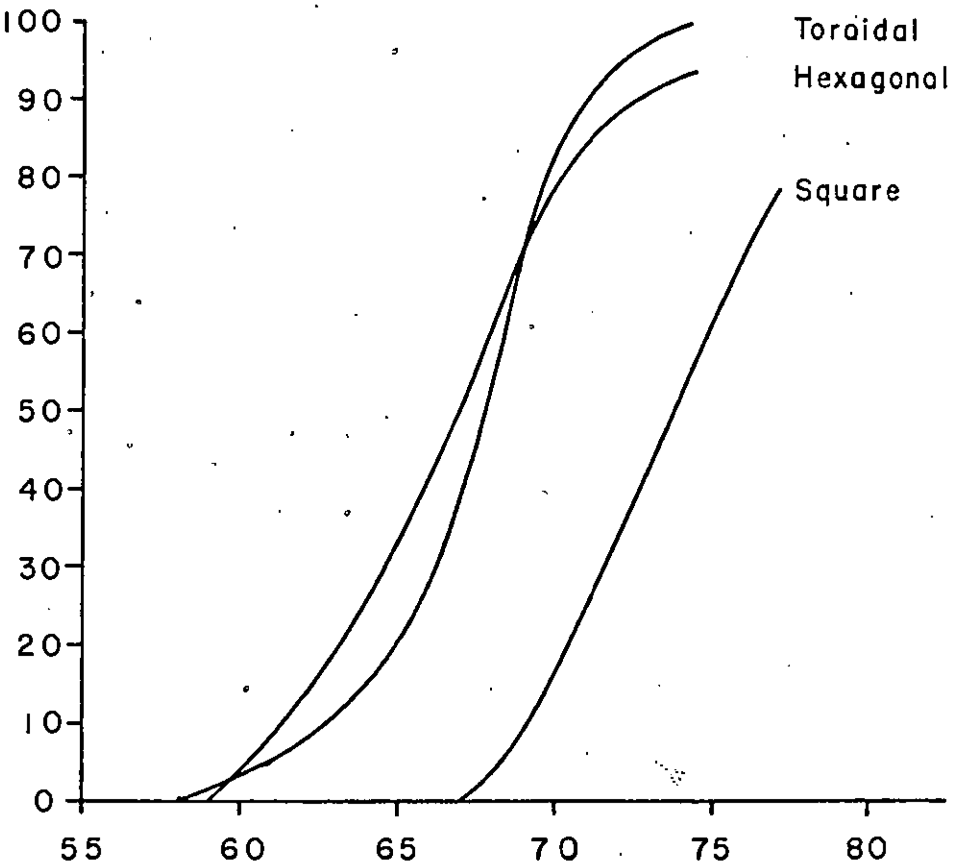


FIG. 5.17. Percentage of spirals reaching 128-chip target (vertical) vs. chip yield (%) for square, hexagonal and toroidal arrays.

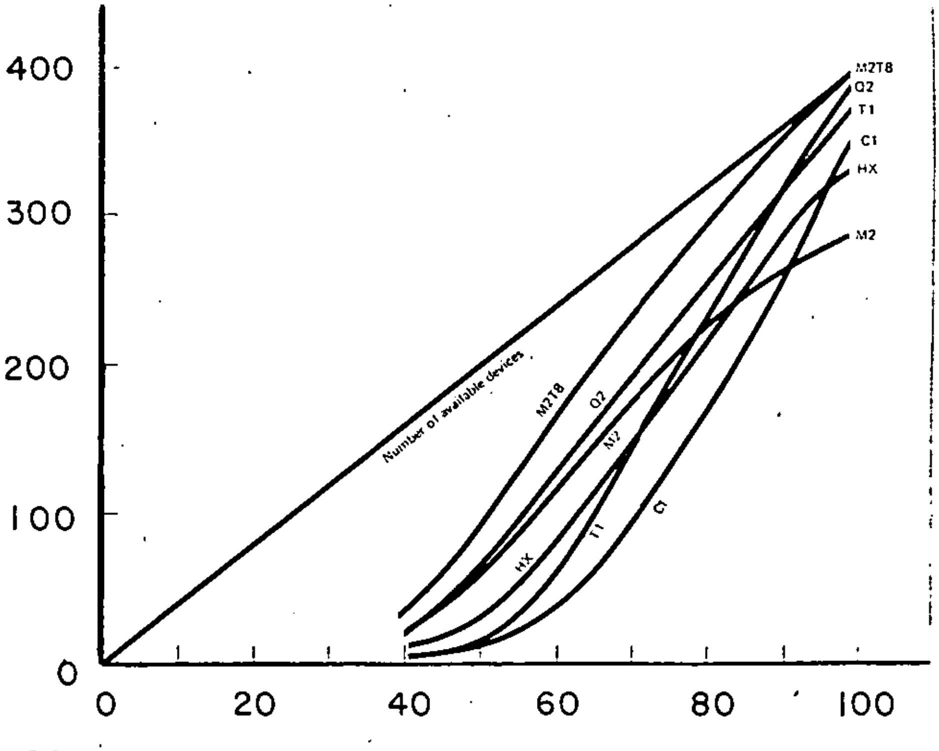


FIG. 5.18. Mean maximum spiral length vs. chip yield (%).

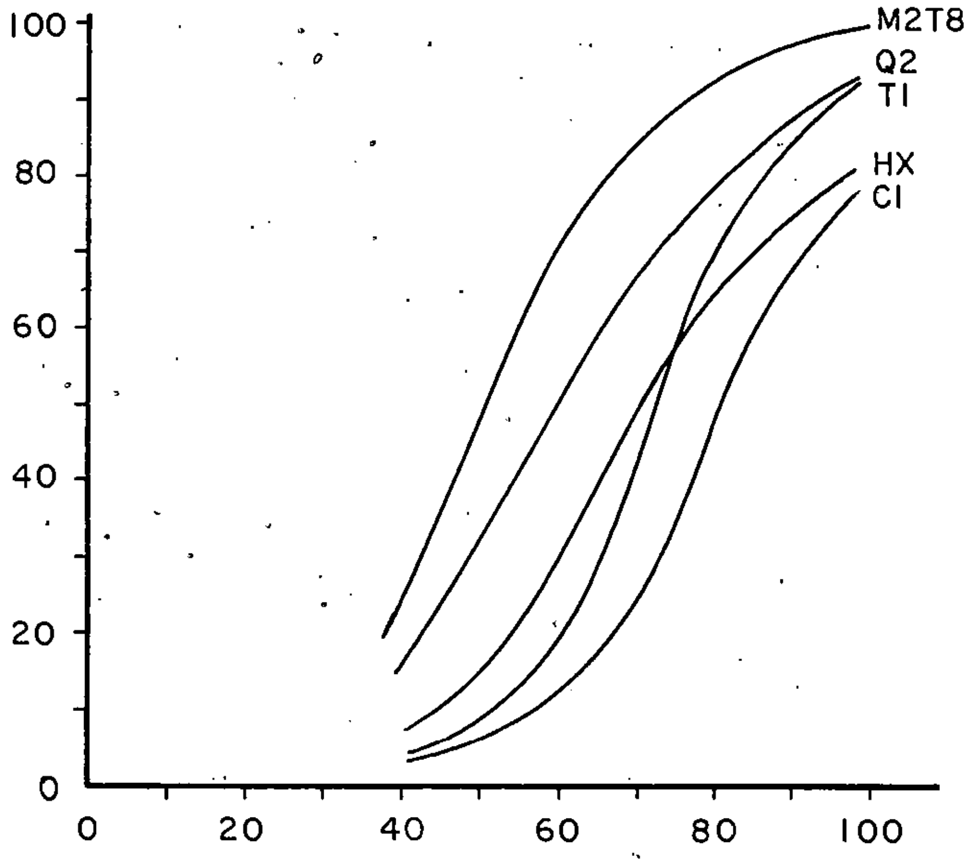


FIG. 5.19. Percentage usage of good chips (vertical) vs. chip yield (%).

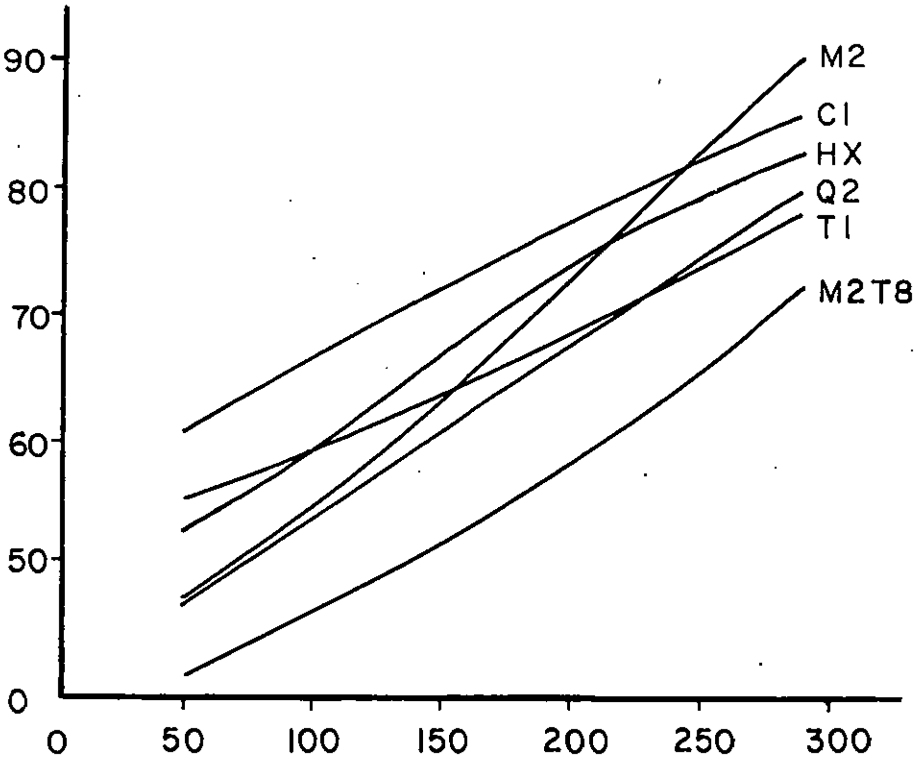
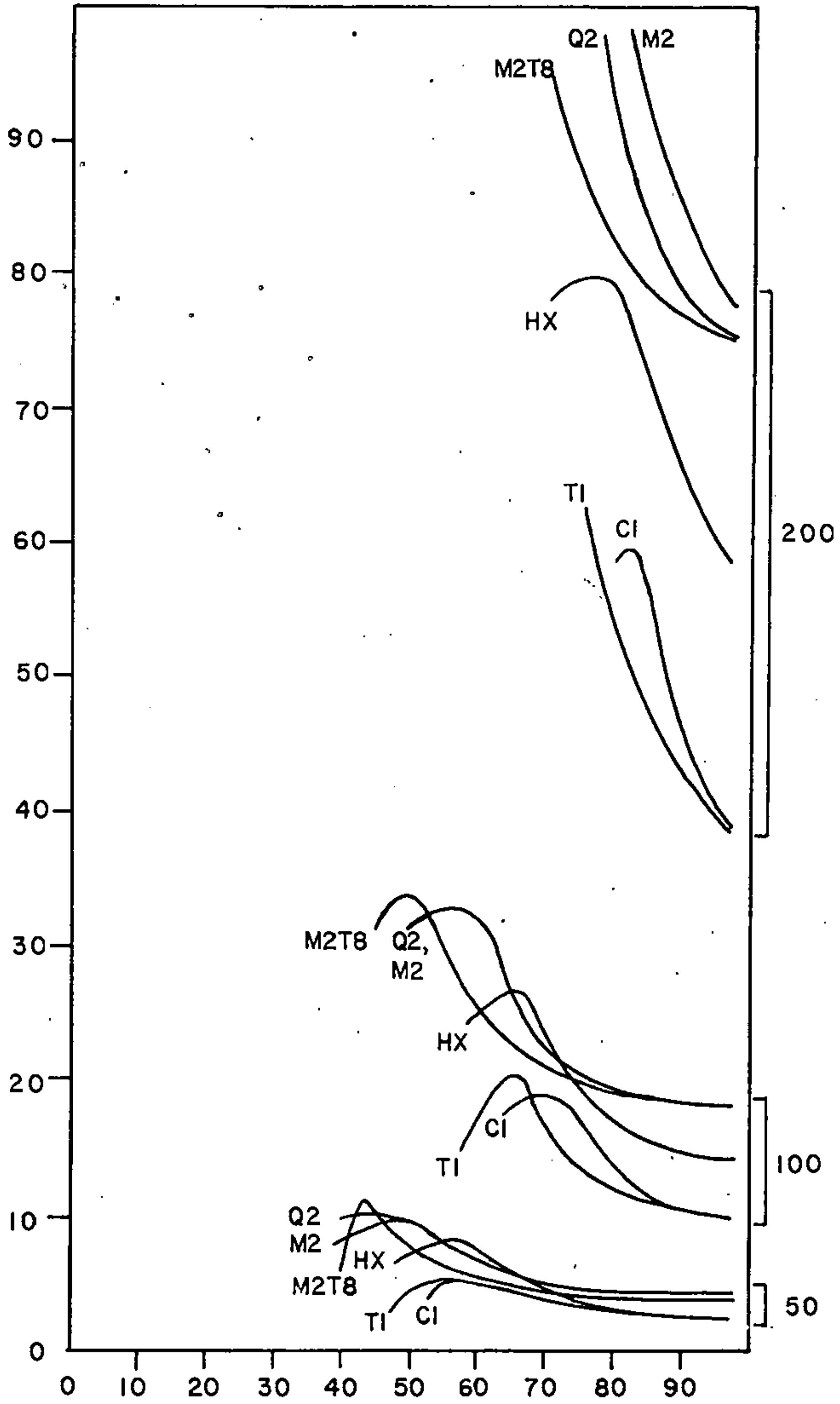


FIG. 5.20. Effect of variation of target length of spiral (horizontal) vs. chip yield (%) for 95% of spirals to reach target.



Mean number of device test times (thousands) to assemble spirals of 50, 100, 200 chips (vertical) vs. chip yield (%).

FIG. 5.21. Speed of spiral formation.

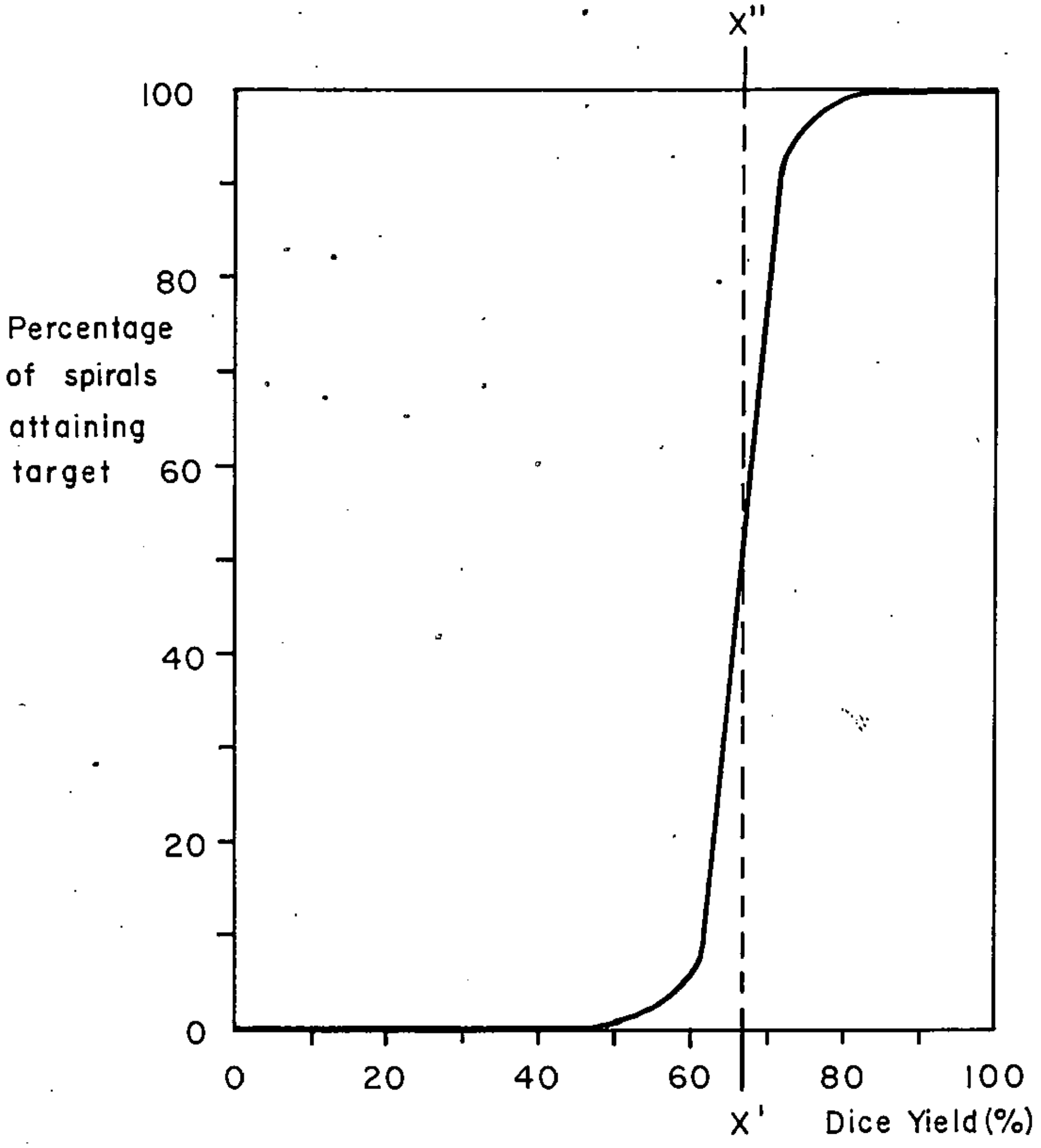
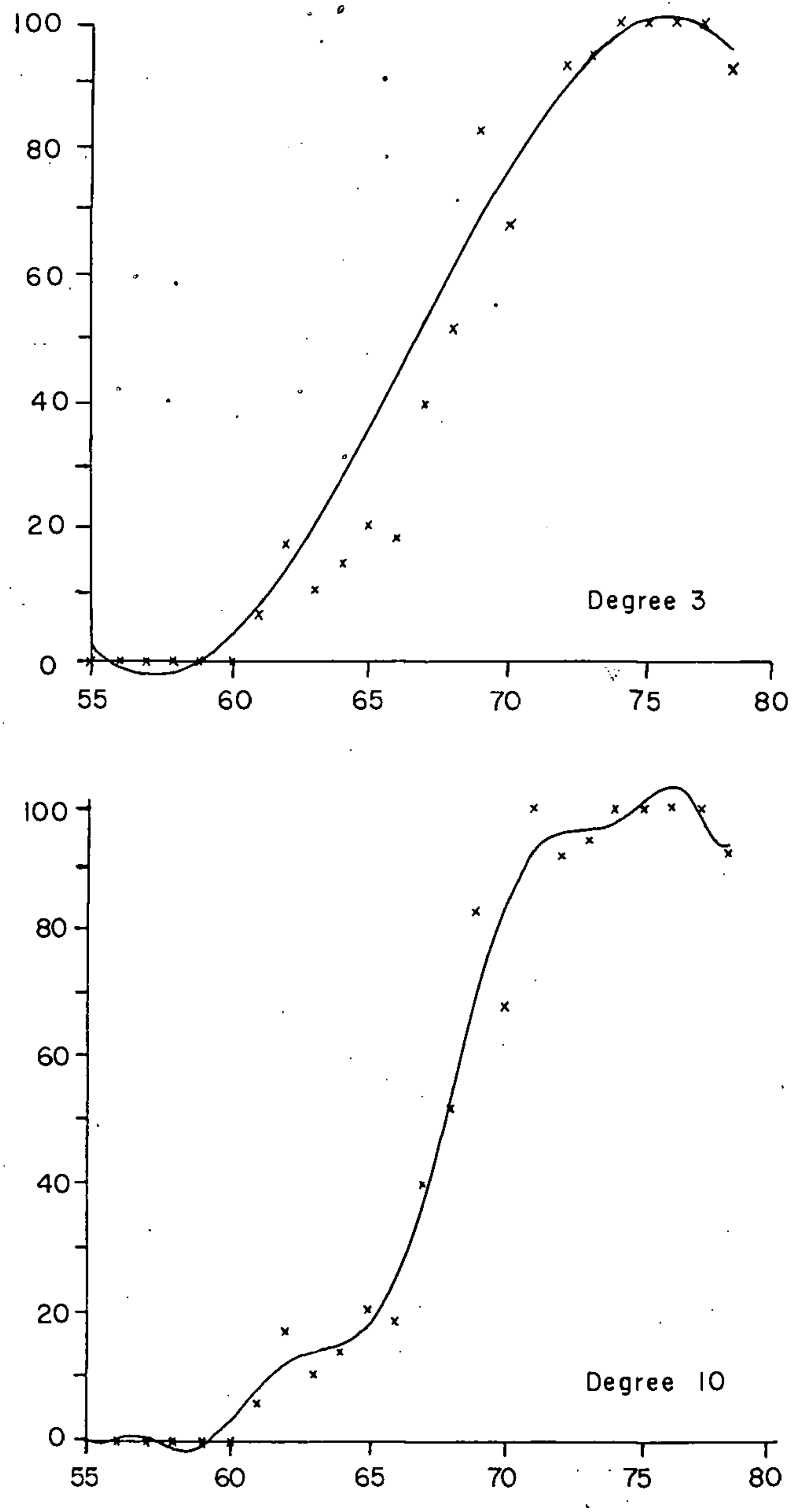


FIG. 5.22. Idealised characteristic of algorithm.



FIGS. 5.23-4. Algorithm approximation by polynomials of degree 3 and 10. (Weighted End Points).

6. DISCUSSION AND REVIEW

Based on an analysis of the results of spiral simulations presented in the previous Chapter the general principles underlying the behaviour of algorithms are considered. This leads on to the presentation of new memory structures. The concepts of fault and failure tolerance are then studied followed by a consideration of potential problem areas in the implementation of W.S.I technology. Topics relevant to the commercial viability of W.S.I memories are then discussed with particular emphasis on assembly technology and a cost comparison of conventional and W.S.I memories.

6.1 DISCUSSION OF SPIRAL SIMULATIONS.

From the studies of these algorithms it would appear that there are two distinct mechanisms which can limit the growth of a spiral. Firstly, the spiral can be extinguished because the yield is inadequate to sustain growth and, secondly, the array boundary may impose an insuperable obstacle to continued spiral development. The first of these effects is analogous to the stifling of a chain reaction (e.g. by increasing the degree of neutron capture in a nuclear reactor) and would, in the absence of edge effects, be expected to show an extremely sharp transition between rapid extinction and indefinite growth as the yield increases. The second mechanism is expected to be a complex function of the array geometry, organisation and algorithm which is attempting to create the spiral. It must also depend on array yield in that more backtracks will occur at lower yield levels and so the spiral will require to make use of a greater proportion of the array and hence be more affected by the array boundary.

A further complicating effect is that the finite array size must cause an abrupt curtailment of the probability of completing the 128 chip spiral at a definite chip yield; e.g. if the yield is < 32% then there must be less than 128 good chips on the 20 x 20 array and so the target cannot be achieved.

These effects must result in a very complex dependence of the percentage of spirals which reach the target on the array yield. In order to attempt to separate out the two effects it would be of interest to compare the results for the rectangular and toroidal arrays with the performance of the normal rectangular algorithm on an infinite array - where the edges could not possibly interfere with the progress of the spiral. A quasi-infinite array is obtained merely by ensuring that the spiral can never reach the array edge; an array of 255 x 255 chips would guarantee this for a 128-chip spiral but a considerbly smaller array would permit the vast majority of spirals to develop without touching the boundary. Any which did so could be rejected from the analysis. However, in view of the limitations of the curve fitting programs, as described in Section 5.4.1, there was little incentive to attempt to separate out these various contributions to the supposed general equation for an algorithm and this proposed study has not yet been undertaken.

The algorithm proposed in (10) aimed at generating a spiral structure - as shown in Fig. 5.4. It seemed intuitively obvious that such a spiral starting at the centre of the wafer where the yield is highest would, by tending to remain in this higher yield region and hugging itself closely to avoid isolating good chips from the rest of the unused array (apart from in blind alleys), achieve longer chains than any without such a tight packing characteristic. The other algorithms were therefore based on this same concept of tight spiral formation.

A casual examination of the path taken through an array by a spiral might lead one to suppose that all the algorithms are of very poor performance - surely the eye should not be able to do, at a glance, better than the computer achieves by a lengthy recursive procedure? It must be remembered, however, that the reader has a tremendous advantage over the computer; he can see instantly where the spiral is in relation to the array boundaries and the remaining areas of good chips. The computer may be likened to a person walking in (say) the Hampton Court maze; a well-planned and carefully maintained maze can be a real challenge to pass through unaided by sun or compass and yet be easily solved when studied from the air or as a map. Similarly the computer has no idea where the spiral is in the array or what is ahead or "around the corner". It has been deliberately restricted in this manner to accurately simulate the procedure of building a spiral on a real slice without first testing all the chips. It therefore selects the first route which achieves the target and does not investigate whether any alternative may be superior - as a person may attempt to negotiate a maze by always turning left whenever possible.

The alternative approach of mapping the fault distribution in each array and computing the optimum track (which may well not be a spiral) would have the advantage of reduced"spiral" assembly time but would lose the graceful degradation feature as the wafer would require to be removed from the system and each chip retested to establish the position of the failed chips before a new spiral could be configured.

These simulations of spiral generation indicate that even the simple four way algorithm will allow long spirals to develop at yields of less than 80% while the more powerful 6-way and toroidal algorithms will construct long spirals at yields below 70%.

A study of Fig. 5.17 shows a very similar performance for the toroidal and hexagonal algorithms. This implies that the elimination of the array boundary achieved with the toroidal configuration is equivalent to an increase of two nearest neighbours, suggesting that one may trade-off some on-chip control logic (i.e. active devices) for passive metallisation in moving from the HX to the T1 array. The toroidal algorithm has the further advantage of reduced spiral assembly time (Fig. 5.21).

To put these spiral assembly times into perspective, it will be noted from Fig. 5.21 that a 200 chip spiral using the T1 algorithm requires a minimum of ~ 40,000 chip test times (N). If the chip is a 5K bit shift register (L) then at a p.r.f of 1MHz (F) the total spiral assembly time is $NL/_F$ second, i.e. 200 seconds. It will be observed that, with the exception of this T1 algorithm, the more powerful algorithms tend to have increased spiral assembly times. This is due to the higher average number of unsuccessful attempts before an unused chip is found and added to the spiral in the cases of 6- and 8-way algorithms as illustrated in Fig. 6.1..

With reference to Fig. 5.21 a 50% increase in minimum assembly time is observed in going from the 4-way to a 6-way algorithm, while the 8-way algorithms take virtually twice as long to assemble a spiral of given length as do the 4-way ones.

It should be noted that the system does not require to be down during this period. On detection of a faulty spiral, EDC (error detection and correction) procedures would either continue to supply the missing bits during the wafer reconfiguration time (if organised to contain one bit of each of 1M words as discussed further in Section 6.5), or arrange to switch in a substitute wafer (already configured and waiting for use).

6.2 OTHER ALGORITHMS

In describing his "arm" generating algorithm, Manning states[34]:-

> "When an arm has grown to a certain tip, it tries to extend itself toward the nearest array edge. Thus an arm spirals towards the centre of an array in a perfect array. If no improvement in the maximum discovered arm is made in one-fourth of the time limit, the program looks at adjacent cells that are not included in this longest arm and are not known to be flawed. The program tries simple jogging of the arm to include these cells".

This procedure of spiralling inwards from the array edge has been compared with the outward spiral in Section 3.4.3. Although it is easy to simulate "arm-jogging" it is difficult to see how this can be realised in practice. Cells not accessed during the development of the spiral could possibly be so included but those "frozen" in fixed states after backtracking from blind alleys - the major group of good unused cells - could not, it is thought readily be reaccessed for addition to the spiral. The technique would, however, work given a detailed map of the spiral and faulty chips in the array when the spiral could be rebuilt to maximum length including such chips.

In the studies of chain-generating algorithms other approaches than the spiral forms discussed have been considered. One such proposal was that the algorithm should always endeavour to minimise the deviation of the spiral tip from the slice centre (i.e. number of steps North minus number of steps South plus number of steps East minus number of steps West). This would have reduced the effects of the array boundary in deviating the spiral but would have required additional hardware on chip Z.

To reduce the complexity (and hence area) of the on-chip control logic it is advantageous to minimise the number of nearest neighbours if this can be done without requiring an unacceptable array yield increase for spiral generation. In looking for the simplest possible algorithm it is important to bear in mind the two fundamental requirements of any spiral algorithm. Firstly, it must be able, in principle, to eventually access all the wafer from any given point; this requires progression in the four directions :-

Forward along X axis
Backward along X axis
Forward along Y axis
Backward along Y axis

These movements do not, of course, require to be independent. The second fundamental requirement is that the algorithm must be capable of bypassing faulty chips without leaving intervening gaps of good chips.

These requirements can be achieved with only three OPEN and three (different) LOOK directions using two chip types, α and β as illustrated in Fig. 6.2.. A further simplification is possible, to a single chip type as illustrated in Fig. 6.3, if it is permitted to miss the occasional good chip in the array.

These algorithms illustrate a basic flaw in all the ones which were described in Chapter 5 - namely that, if the same set of directions is selected for OPEN and LOOK, then one of the OPEN directions must be inaccessible to the new chip because it leads back to the penultimate chip in the spiral. A four in/four out algorithm of this type is therefore effectively a four in/three out algorithm.

Another proposal was that the spiral should be permitted to leap over an imprisoning wall of bad chips - perhaps in the manner of Knight's move in chess - rather than backtrack from blind alleys. This would undoubtedly improve the performance of all the algorithms studied by not leaving good chips isolated from the spiral in blind alleys. The most obvious way of incorporating such a "leap" facilty into the design would be as an additional open/look connection on each chip so that chips would contain the logic for the sequential addressing of the normal OPEN directions before initiating "leap". One such leap algorithm is illustrated in Fig. 6.4.. This is based on a single chip type with three OPEN and three (different) LOOK directions and has one direction of leap. It will be noted, however, that a single direction of leap is of restricted usefulness; for example, a dead end to the north-east of the main body of the spiral of Fig.6.3 could not be exited by the spiral. Another point to consider is that the link providing the leap across a bad chip may well itself be faulty.

A study of the interconnection requirements of such "leap" routines suggested that their inclusion on actual devices would cause serious problems of layout and so they were not investigated further.

6.3 TOWARDS THE "IDEAL" ALGORITHM.

Many types of chain-generating algorithms have been investigated and assessed against several criteria. The most powerful of these algorithms - the toroidal 8-way - will tolerate an array yield of ~ 57%. It must be noted, however, that in comparing 4, 6 and 8-way algorithms no account has been taken of the larger chip area (and attendant lower yield and reduced number of chips on the wafer of given size) of chips carrying the additional control logic required to implement the more complex algorithms. The ultimate criterion for algorithm selection is to produce the largest memory on a slice of given size for a particular set of design rules and a particular process line.

The "ideal" algorithm (forgetting for a moment the wasted area occupied by the control logic) would have an infinite number of nearest neighbours and an infinite array size. On any finite array all chips would then become nearest neighbours and the spiral could include every good chip on the array (in any chosen sequence). It is interesting to note that as the number of nearest neighbours increases the lengths of chains of bad chips must also increase (although the cluster size of good chips also increases for a given array yield). In the limit, since all chips are nearest neighbours, the bad chips also form a single cluster on the array, regardless of yield.

While this concept of infinite adjacency is no more than a mathematical abstraction for serial memories the idea immediately becomes a reality if we think instead in terms of a parallel organised array. This concept is now developed further.

6.4 PARALLEL/SERIAL ORGANISED W.S.I ARRAYS.

Two fault-tolerant fixed interconnection structures are now proposed to cover RAM, ROM and associative memory applications.

6.4.1 A Random Access Memory Structure.

It has been shown that one of the major drawbacks of all non-infinite (i.e. practicable) serial algorithms for creating a spiral is that not all good chips on the wafer can be accessed. Furthermore, many of those initially included in the chain are rejected as the spiral backtracks out of blind alleys.

The possibility of directly accessing all chips on the wafer was therefore considered. This is most readily achieved by running X and Y address lines to all chips so that any particular chip can be accessed for purposes of testing. It is then possible to use a minicomputer to construct the optimum spiral through the array. This approach would have the additional advantage that the degree of degradation (and hence tolerance of additional chip failure) of the array could be assessed at any time; the philosophy of not retaining any information on either the distribution of good and bad chips or the track of the spiral does not permit this information to be gained.

However, rather than convert such a parallel array into a serial memory it would seem that the device would offer considerable advantages if left in parallel form. It would appear externally rather like a section of core store but with two important differences. Firstly, some of the sites would be faulty and unable to store data; it would therefore require such faulty sites to be bypassed. Secondly, each W.S.I array would be equivalent not to a plane of core store (as in a RAM, for example) but to a block of store, the multiple Z planes being represented by the serial shift register (probably a serial/parallel/serial ccd chip) at each chip site. We could envisage, for example, a store of between one and two megabits capacity, produced by a 20 x 20 chip array of 5K bit shift registers as shown in Fig. 6.5.

Such a parallel/serial array would require very little on-chip control logic compared with that required for the serial memory because each chip is not required to decode an address field and is not required to interact with its neighbours (although such a feature could possibly be retained if this were sufficiently advantageous).

The device would, of course, require to be fault tolerant; this could be achieved by protecting the X and Y address lines with series resistors of ~ 10 - 100 KΩ between the lines and each chip; such a high resistance would not be detrimental in an MOS design and would provide adequate protection of the address lines against short circuit chips. Short circuits between the address lines themselves would be best eliminated by removing them from the wafer onto the substrate or polyimide superstrate which could be tested independently of the wafer prior to assembly. This point is considered further in the next section.

Such a design has the advantage that not only is it fault/failure tolerant but the degree of degradation is immediately assessible - an important feature for fault tolerant devices.

6.4.2 An Associative Memory/ROM Structure

An alternative memory structure based on the hypothetical model town of Dyadicville* is presented in Fig. 6.6. This provides direct access via an equal line length to all chips in the array.

This memory does not fit exactly into any of the three categories spiral, tree or grid. Although it might appear superficially to have a "tree" configuration, all chips are in fact directly accessed without requiring data to pass through intervening chips. In this respect it is most similar to the true "grid" structure described in the previous section.

The device would seem to be suitable for either associative memory or ROM where the address locations are stored in nonvolatile (e.g. hardwired, FAMOS or EAROM) form as discussed in Section 2.2.2. The inclusion of redundant bits in the serial address code would ensure any chosen degree of protection against faulty address locations responding to other addresses. In the associative memory

* See, for example, H F Harmuth "Transmission of Information by Orthogonal Functions" Springer-Verlag, 1972.

structure, faulty chip sites would never be accessed, the reduction of storage capacity arising from initially faulty chips and those failing during operation being allowed for by redundant chips on the wafer.

ROM applications in which a faulty chip would cause serious disruption could have duplication at two (widely separated) chip sites. An alternative procedure (e.g. in signal processing applications) would be to have the incremental function in adjacent locations equal to one-half or one-third of the required resolution to allow for single or adjacent-pair chip failures.

This array does not require the multiple X, Y address lines of the previous structure and is therefore better suited to small chip sizes. Whereas the optimum RAM structure would probably contain ~ 400 chips of 5K bit capacity this present device could usefully contain many more chips of smaller size thus improving the array yield and accessibility of data.

6.5 FAULT AND FAILURE TOLERANCE.

Fault tolerance occurs at many levels in both hardware and software in computer systems and has been widely discussed in the literature*; this section considers only those aspects relating to semiconductor memory chips and, in particular, to W.S.I arrays.

* See, for example, IEEE Trans.Comput. Vol.C-20 November 1971, Vol.C-22 March 1973, Vol.C-23 July 1974, Vol.C-24 May 1975, Vol.C-25 June 1976, Vol.C-27 June 1978.

Minute imperfections will always be present in integrated circuit chips; these only become critical if they prevent the correct operation of the device or constitute a major reliability hazard. Discretionary wiring and fusible link techniques can, as we have seen, create an apparently perfect array on a flawed wafer; however, any additional major defects occurring in these hardwired arrays will be catastrophic.

The very nature of the "softwiring" (gated interconnection) between chips on this approach to W.S.I endows the serial memory array not only with the attribute of fault tolerance but also of failure tolerance. If a particular chip develops a fault chip Z can cause the spiral to retract to a single chip and regrow, bypassing the new faulty chip to create a fresh spiral. This self-repairing feature, which permits a graceful degradation of the system, is not possible with a Programmed Interconnection approach to W.S.I.

It is essential that the proposed parallel/serial arrays should also possess this fault/failure tolerance. Considering first the faults which may be permitted in the construction of integrated circuit chips, we may observe the following degrees of fault.

a) No faults at all - i.e. an absolutely perfect chip; this cannot be achieved.

b) No faults detectable as being outside the specification limits during routine testing of the chip. This is the normal approach with discrete chips. Minor faults may be present, e.g. nibbles in the aluminium track edges, or oxide pinholes. These, whether they be noticed (e.g. slightly higher leakage current than normal) or remain undetected (e.g. oxide pinholes) must be considered as potential reliability hazards unless the exact nature and location of all such faults is known - an impossible task.

c) The approach adopted for the serial memory device. Here it is accepted that serious faults are present in the device structure but the memory is organised in such a way as to appear perfect to the computer. Once this idea of fault avoidance (as distinct from fault elimination) is accepted then the device may also be endowed with the attributes of failure tolerance and graceful degradation - as in the serial memory.

d) Here no attempt is made to hide the faults from the computer - it is programmed to accept them. The flawed wafer is installed in the machine without any prior attempt to create a perfect memory from the flawed array.

We have seen that categories a) and b) can never (and should never) apply to W.S.I arrays. The proposed parallel/serial RAM array can be organised into either of categories c) and d) above. Taking category d) first, this could be achieved by the following procedure.

i) Write data into X_m, Y_n location from buffer store.

ii) Read data from X_m, Y_n location and check for corruption of data.

iii) If corrupted, move to X_{m+1}, Y_n location and repeat procedure; if uncorrupted use X_m, Y_n location.

This has the disadvantage of increased write time but does not require any information to be held concerning the location of faulty cells. This procedure would be acceptable in applications requiring the reading of data in the sequence in which it was stored. Its use for storage during program execution would however, require a confirmatory associative address to be stored with each block of data and a program instruction to increment the address in the event of a faulty chip or an incorrect data block being accessed.

Category c) is achievable in two ways. Firstly we may test all X, Y locations and store the good addresses in a look-up table; this may be either on-slice or external to the wafer. Only good chips are then accessed by data, the look-up table being updated if faults are detected.

In applications which require a complete X by Y array (without any faulty sites) it is proposed that an address buffer should convert all incoming (virtual) addresses directed towards faulty locations into real addresses which access standby-redundant cells placed alongside the main array on the wafer. This configuration is sketched for a 5 x 5 array in Fig. 6.7.. In this sketch the 1, 4; 3, 3 and 5, 2 input addresses would be converted by the address buffer to avoid the faulty chip sites in the main array. They could, for example, become 1, 6; 2, 6 and 4, 6 (bypassing, of course, any faults in the subsidiary array).

This technique of address buffering was proposed by Sander[45] in relation to the use of faulty memory components on printed circuit boards and is believed to have been adapted by Texas Instruments to create perfect bubble memories. Sander's original suggestion required operator intervention to repair any subsequent faults; it is now proposed to incorporate continuous and automatic updating of the address buffer under (on-chip) microprocessor or external control.

Single open circuits on the address lines could be bypassed using a "ring-main" (toroidal) configuration - feeding both X and Y lines from each side of the array. However, address line open or short-circuits would be unlikely in view of the coarse geometries involved; they could, in any case, be eliminated by the use of pretested polyimide film superstrates as described in Section 6.8.2.

Protection of the address lines against overload by faulty chips is readily achieved by the insertion of series resistors ($\gtrsim$ 10 KΩ is permissible in an MOS technology).

Such resistors would require not to have a junction defect within a few thou. of the contact to the address line but this is not considered a problem in view of the relatively small area occupied by the diffused resistors and the low defect densities essential for VLSI device manufacture.

The possibility of faulty chips attempting to feed spurious signals or stuck-at-one faults onto the address/

read lines must also be considered. Major faults (i.e. short-circuit or excessively leaky chips) are no problem - they will drop so much voltage across the supply protection resistor that they cannot generate a "1" level. Minor flaws could be guarded against by inserting additional series transistors in the READ output circuit of the chip so that two (or more) widely separated transistors must be simultaneously faulty before a chip may output spurious data. If these two (or more) transistors are both (all) faulty, then there must be a major fault on the chip which, as we have seen, precludes the possibility of a "1" being generated anyway.

It should be noted that both the serial and parallel/ serial RAM array memories can be organised in such a way as to prevent loss of data in the event of chip failure. For the serial memory it is proposed that a number of wafers (say 32) are used in parallel, each storing one bit of (say) 1M words. Error detection and correction procedures could then be applied to reconstruct the data and locate the faulty wafer which could be switched off-line and replaced by a stand-by redundant wafer while the spiral is reconstituted. The parallel array may be similarly organised -either using thirty-two separate wafers or, since one chip failure should not interfere with the operation of the memory as a whole (unlike the spiral situation where the failure of any chip will cause corruption of all the data in the spiral), using thirty-two chips on the same wafer to store (say) 5000 thirty-two bit words.

One problem with the serial array is that the extent of degradation is never known until it proves impossible to reconstruct the full spiral. This cuts right across the principles of fault tolerant computer design. One way round this problem is to organise the memory as a set of wafers in series, allowing all but (say) the last five wafers to attain their maximum possible chain length. The last five wafers are initially little used, making up the total required spiral length with short spirals only on each wafer. As chips fail on the earlier wafers, longer and longer chains will be required on these later wafers in the set. It could, for example, be arranged for the wafers to be taken off-line as soon as convenient after the last wafer is called on to supply chips to the memory. The wafers would then be individually assessed and the most degraded ones discarded.

This problem does not apply to the proposed parallel/ serial arrays; all chips can be individually tested and the level of degradation determined at any time.

6.6 RELATIONSHIP OF PARALLEL/SERIAL ARRAY TO OTHER W.S.I STRUCTURES.

The two arrays of Section 6.4 have been proposed primarily to illustrate a coverage of the field of semiconductor memory using W.S.I technology. They are believed to be original structures but, as noted in Section 2.2, a full survey of the literature in this field is outside the scope of this work. Discussions with the industry have aroused considerable interest in the RAM structure; the Associative memory/ROM application of the "Dyadicville" array is a more recent suggestion and has not yet had adequate opportunity for critical assessment.

The programmed interconnection route to W.S.I has been examined and discounted in this work for applications requiring self-reconfiguration in the event of chip failure. The elimination of the 100% yield route as impracticable leaves only the fault tolerant, fixed interconnection route on which the major known work has been discussed in Section 2.2.4. Manning[34] considered "arm" algorithms (equivalent to the spiral of this work), "tree" algorithms (which were considered but not simulated in this work as they were thought to be too difficult to organise into either a serial or a parallel system) and "grid" structures - which are perhaps most closely related to the parallel/serial RAM array of Section 6.4.1. However, he has considered only the concept of "embedding a perfect machine in a flawed array" - as described in Section 2.2.4 and came to the conclusions that "grid embedding is the most difficult repair problem in a checkerboard array" and "Repair efficiency is much smaller for grids than for arms". These conclusions are valid for his approach to grid structures but do not apply when the parallel/serial arrays of Section 6.4 are considered. "Repair" is both easier and more efficient than in spiral arrays so Manning's equation relating optimum repair efficiencies of grid, arm and tree algorithms, noted in Section 2.2.4,

$$\text{"}ORE_G \leqslant ORE_A \leqslant ORE_T\text{"}$$

no longer applies. All good chips may now be used in the "grid" structures proposed so the equation becomes

$$ORE_A \leqslant ORE_T \leqslant ORE_G$$

6.7 POTENTIAL PROBLEM AREAS IN WAFER-SCALE INTEGRATION.

Many potential problem areas specific to W.S.I. and therefore new to current technology have been discussed at relevant points in this work; some, however, have not fitted logically into the sequence so far and are now considered.

6.7.1 Global Power Supply

The general field of power/clock distribution to all chips is now considered.

To reduce the magnitude of the task of protecting the full wafer power supplies and the complexity of interconnections the number of external power/clock supply grids should be reduced as far as possible. The back face of the wafer may be used as an earth plane in many technologies so it is, in principle, possible to generate all the clock signals required for the normal operation of the serial memory on chip from two input master clock waveforms or one clock and a dc supply. Even the dc supply may be eliminated in some technologies, or derived on chip from the input master clock using diodes and capacitive storage (at the expense of a considerable area of silicon). The master clear signal could either be supplied via a separate grid or possibly achieved by holding the clock supply at constant voltage for several seconds. However, this section is not devoted to minimising the number of grids for any particular technology but to ensuring the protection of those which may be required.

The use of current limiting resistors has been discussed in relation to both the serial and parallel arrays in Section 6.5; these have the advantage of simplicity of design and can be arranged (for MOS technologies) to limit the chip dissipation to less than 1 watt, even with all clocks short circuit.

Another possible technique is based on the use of fusible links - any chip trying to draw too much current will blow a series fuse. There are, however, several problems with this approach. Firstly, the technique requires an additional metallisation and photoengraving stage. Secondly, while fusible links can be reliably blown[7 36] when given the ideal current pulses to do this (both total energy and pulse shape being critical parameters) few fuses would be presented with such an ideal overload. When improperly blown the fuse material can be redeposited as debris over nearby metallisation, thus constituting a reliability hazard; alternatively, the fuse may eventually regrow if improperly blown. The third problem is the setting of a suitable threshold limit for fusing. If set too low there may be some tendency for electromigration of the fuse material on good devices, thus imposing a reliability hazard. If set too high the leaky devices will continue to draw excessive current. A further unsatisfactory feature is that <u>all</u> wafer chips not drawing excessive current are left powered up rather than just those required for the spiral.

Using currently available technology, the best way of achieving protection (while simultaneously minimising the power requirements and thermal dissipation of the wafer) is to arrange for the power supply to be switched on to each new chip by the developing spiral. Having added a good chip to the chain the next chip would be momentarily powered up by turning on a transistor isolating this chip from the power supply. Chip Z would decide whether the additional current drawn was acceptable and, if so, proceed to test the new chip. If the current drain was too high the isolating transistor would be switched off again and another chip accessed instead. This technique would ensure that only those good chips actually forming part of the spiral would draw power; the unlikely eventuality of being unable to switch off the bad chip owing to a fault in the switching transistor of the previous chip could be coped with by back-tracking one further chip. It is essential that no chip may switch itself on; the signal to do so must arise from the preceeding chip. A further advantage in powering up chips as they are added to the spiral is that a measurement of increase in supply current will enable marginal chips to be rapidly identified thus reducing test (and spiral configuration) time and improving wafer reliability. This technique could be extended to the parallel array, chips being powered up sequentially during testing of the individual X, Y locations.

Future designs might benefit from a distributed power input over the entire wafer surface by illumination and conversion of the light to electrical energy as described in relation to I^2L by Hart & Slob[20]. This would have the great advantage that faulty chips could only dissipate the power which they themselves generated and could not cause a drain on that of nearby chips. Conversion efficiency of input radiation to electrical energy is ~ 12% for silicon in the visible spectrum. To avoid excessive heating and a large penalty in chip area the total power consumption would require to be limited to ~ 1 watt. This technique may offer considerable promise for future designs.

6.7.2 Double Level Metallisation

While double-level metallisation is practicable on pilot lines it has generally been regarded as an unsuitable technique for quantity production. It is accepted that considerable space could be saved in the device layout if double-level aluminium were available, but no insuperable topological problems arise from a restriction to one layer of metal. If more than one supply grid is required, most circuit technologies will tolerate a small additional resistance arising from diffused cross-unders in at least one of the grids. A second (and even a third) interconnection layer may well arise naturally from the package design - for example the substrate and polyimide superstrate structures to be discussed in Section 6.8.2. Technologies employing polysilicon (e.g. silicon gate MOS) can often use this as a second interconnection layer.

6.7.3 Spiral Branching

If a chip has a particular type of fault in its OPEN address logic, it may attempt to access two adjacent down-stream chips rather than only one. This, if successful, will lead to a branching of the spiral and two parallel spirals will attempt to grow downstream of the faulty chip. This, at best, will lead to a rapid wastage of storage capacity on the wafer. A similar fault on the LOOK address logic will cause the chip to attempt to receive data from two directions simultaneously, resulting in corruption or total loss of data. The nature of the design, however, is such that at least two adjacent chips must have very similar (and improbable) faults for this defect to occur as the spurious LOOK or OPEN direction on the faulty chip should only see a dormant OPEN or LOOK port on the adjacent chip. This built-in safeguard against the transmission of data across such rogue interfaces is believed to be adequate; spiral branching is discussed further in relation to the proposed i/o interface chip in Section 6.8.4.

6.7.4 Thermal Dissipation

This could become a problem with very high speed memories. In addition to the working memory, all chips on the array may be dissipating standby power while faulty chips may well be dissipating several times the normal power level (depending on the nature of the power supply protection as discussed in Section 6.7.1).

A full wafer should be able to dissipate several watts without trouble using natural or forced air convection. To dissipate substantially higher powers it becomes necessary to hold the back of the wafer in intimate contact with a good heat sink. At still higher power levels, phase-change cooling[42] could be implemented although the device would then be reaching the level at which the feeding of power into the wafer becomes a problem rather than the dissipation of the heat produced.

It is, however, proposed that W.S.I should be limited to low power technologies in the first instance to ease the problems of package design; MOS technology, for example, can be limited to ~ one microwatt per bit[2], thus limiting the total wafer power dissipation to a few watts.

6.7.5 Noise and Pattern Sensitivity

In quoting Vaccaro[60], Manning states

> "... the major dilemma facing the user of LSI today is simply that we can build and are building microcircuits today that are more complex than we can adequately test, functionally or parametrically".

It is well known that the close proximity of clock/ power lines carrying large switching transients can cause pick-up and consequent spurious signals in storage elements. It is also generally accepted that RAMs may be pattern sensitive - that is they will appear to function quite correctly under all test conditions but with certain adverse storage patterns they will malfunction. The combination of these two effects in a full wafer memory must be considered as a potentially serious problem requiring evaluation.

It is thought most unlikely that this problem is amenable to calculation and it would therefore require full operational samples to evaluate its importance. The serial memory has the advantage that any corruption of data caused by such noise arising from the addition of a new chip during the setting up of the spiral will be interpreted as a faulty chip. This will cause the new (good) chip to be bypassed in order to reduce the noise problem to an acceptable level by creating a more open array. However, the possibility of pattern sensitivity - that some future adverse combination of data not included in the test sequence may cause a malfunction - still remains.

Lo and Guidry[32] have developed a systematic approach to MOS RAM testing but, in the absence of information on the distribution of good and bad chips on the wafer or the route taken by the spiral it is not possible to simulate likely worst case patterns with the serial memory structure.

No solution to this potential problem is known at this time. However, the parallel array should be considerably more consistent in its performance with regard to pattern sensitivity and, having direct access to every chip should be more amenable to analysis than would the serial array.

Although certain testing problems and pattern sensitivity are expected to be greater for W.S.I than V.L.S.I it should be noted that the reduced chip complexity of the parallel/serial structures enables these to be more readily tested at the chip level than (say) 64k bit shift registers. The problem of testing IC's is not new to the industry. Even a 100-bit shift register (available over a decade ago) would require, at a pulse repetition frequency of 1GHz, a total time to test every one of its possible states of $\sim 3 \times 10^{13}$ years - far longer than the supposed "age" of the Universe ($1\text{-}2 \times 10^{10}$ years on Friedman Model[62]).

This topic of pattern sensitivity is closely linked with the general problem of array testability. Seth[49], in considering complex arrays, concluded that two conditions must be met to detect a fault in such arrays. Firstly it must be possible to apply the test set to every cell in the array. Secondly there must be a means for propagation of the effect of a fault to an observable output(usually at the array boundary).

The problem of implementing this second condition is by no means restricted to two-dimensional chip arrays. Even a small-scale integrated circuit does not generally permit access to the internal circuit nodes where the fault may be directly observable.

Consider, for example an eight-bit MOS shift register. A photoengraving fault has perhaps increased the area of gate metal or a p+ diffused region of (say) the third stage; this causes a capacitance increase at that circuit node. The operation of certain types of MOS shift register depends on a favourable ratio of capacitances between the node storing the signal level and that onto which it is coupled later in the clock cycle. If this latter capacitance is too large a "1" signal at that node may well be degraded to a level very close to threshold voltage. The fault will not, however, be detected at the output as the succeeding stages will reconstitute the signal level before it reaches the output. Normally acceptable parametric changes in the device during use (e.g. a small increase in threshold voltage) may well cause device faulure. Voltage and time margining of the clocks will detect border-line chips, but at the expense of an increase in test time.

To reduce the problems of testing and fault diagnosis in complex circuits, Smith et al[51] have proposed the use of a laser beam as a minute, non-contacting, movable probe. This technique may well, in the absence of the bond-pads required for conventional probe testing, prove useful on W.S.I arrays.

6.8 COMMERCIAL VIABILITY.

An accurate assessment of the commercial viability of W.S.I devices generally - and Catt's serial memory structure in particular - requires a full understanding of all potential application areas and the strengths and weaknesses of competitive devices. Such a study is not appropriate to this work; for example a detailed assessment of the possible application of W.S.I to large Analogue to Digital Converters would require an extensive survey. This section concentrates on the major factors affecting usability of W.S.I arrays - cost (which, in turn, requires a study of assembly technology, processing costs and efficiency in use of silicon area), packing density, power dissipation and reliability.

In order to assess realistically the relative merits of conventional and W.S.I device technology it is essential that the latter should be presented to a wide audience from both the industry and the academic sector so that criticisms and questions concerning this approach to large memories may be considered. The proposed devices have therefore been discussed with workers in the U.K integrated circuit and computer industries while the serial memory structure has been described in P4-P8. (listed on Section 10).

6.8.1 Technology Implications of Wafer-Scale Integration

It must be emphasised that the concept of wafer-scale integration does not depend on any particular technology. It is a technique which can - within the limits of speed and power imposed by slice geometry - combine forces with any integrated circuit technology. The application of W.S.I techniques to charge-coupled device process technology has the capability of producing multimegabit memories on a 3" wafer.

W.S.I is technology-independent in the sense that it may utilise any existing i.c. technology, but it is now shown that benefits may well derive from adapting current practice to fit W.S.I.

In contemplating W.S.I arrays the industry must be prepared to rethink the tradeoffs in chip size, dice yield, testing complexity and assembly areas. Whereas it has become economically justifiable to go for the ultimate in chip size to the point where dice yields are vanishingly small (because chip cost is a negligible fraction of final device cost), serial W.S.I arrays demand smaller chips (to increase dice yield to workable levels of ~ 70%). The question is now "How much useful store can be obtained on a 3" wafer?" and the tradeoff is between the areas to be allocated to on-chip control logic/power distribution and chip storage capacity. Reduced chip size gives increased dice yield and greater utilisation of the array, but at the expense of an increase in on-chip control logic overhead.

The parallel/serial array, similarly, has a tradeoff between address complexity and useful storage capacity, both of which decrease as chip size increases (and yield reduces). The reduced accessibility of data in the larger chips also favours the use of smaller chip sizes.

Rather than attempt to adapt all the ideas and procedures currently in vogue in the industry to W.S.I technology it is suggested that a totally new appraisal of the technology is required. The application of value engineering principles (see, for example [89]) may well suggest radical changes from current techniques. Two specific examples are now quoted to illustrate this point.

i) Projection - printing mask aligners were introduced by the industry to reduce faults arising in photolithography and so increase the chip size which could be employed for a given (low) yield level - in line with the general philosophy quoted in Section 3.1. The industry has had to accept the serious disadvantages of degraded resolution (2µ minimum linewidth and spacings, but typically 4µ acceptable in practice) and reduced field coverage (even 3" diameter field is difficult to achieve at this resolution) compared with in-contact printers. With W.S.I, higher fault densities can be tolerated and one can therefore consider reverting to contact printing procedures for photolithography and so take advantage of the increased gate packing density permitted by the higher resolution (1µ lines can be printed with care).

ii) Slice diameters in the industry have approximately doubled every seven years over the last two decades. 3" diameter is the current industry standard although some UK companies are changing directly from 2" to 4" capability. The advantages of large slices with the conventional technology are relatively minor. With W.S.I, however, there is far more to be gained by pushing towards the limits of silicon crystal technology - currently ~ 10" diameter. Although there are still many unsolved problems for slices over 5", it is anticipated[24] that 5-6" slices will be in production in the early 1980's.
Such large wafers would be an embarrassment to projection printers but, as already suggested, there is a strong case for reverting to contact printers (which require only a collimated light source instead of a high resolution lens) for W.S.I.

6.8.2 W.S.I Device Assembly

The area of device assembly has so far been essentially limited to a paper study owing, primarily, to the vast cost of setting up the required facility to construct the actual devices. The full advantages of W.S.I. can be realised only if a cheap but reliable packaging technology can be developed. It may well prove that the 50W package for 3" slices, noted in Section 2.2.4, provides an acceptable solution but no further information is known on this package.

Hermetic Packaging

Initial ideas in this work were based on the use of a metallised alumina plate with a square hole of approximately the chip size situated near the centre. This "picture frame" would surround either one normal chip or, better, a special input/output chip (as discussed in Section 6.8.4). Such a "picture frame" assembly for a 4ϕ dynamic MOS implementation is sketched in Fig. 6.8.

A flip-slice approach was then considered. In its simplest form the slice metallisation would be connected directly to a premetallised substrate via solder bumps or other form of pillar bond. This would have several advantages. Firstly, connections could be made at every chip site so the orthogonal supply grids could be transferred from the wafer to the substrate along with the current limiting resistors. Any short circuits between the grids could then be eliminated by testing the substrates prior to assembly. The chip size would, moreover, be drastically reduced as a result of removing the grids from the slice. A second advantage would be the relative ease of converting the array to a finite unbounded form by providing links on the substrate between top and bottom rows and (with cross-overs) left and right columns of chips. A third point in favour of this approach is that it provides for multiple inputs to the slice, either to permit multiple spirals to develop or to provide alternative input chips should the first one fail. This proposed assembly route is illustrated in Fig. 6.9.

If all signals are brought into each chip via grids then typical array chips would require seven contacts for 4ϕ MOS; these are ϕ_1, ϕ_2, ϕ_3, ϕ_4, t_o, Master Clear and Earth. The input/output chips would require two additional contacts. To reduce the cross-over problem all X lines could be run on the wafer metallisation with all Y lines on the substrate but additional bonds would then require to be made between these two metal layers.

Although this procedure would require ~ 3000 contacts for a 400 chip array the important point to note is that not all contacts require to be operational. The assembly is also fault tolerant; any contacts which do not function merely give an additional fault pattern to be superimposed on that arising from the slice fault distribution. An exception to this is the intermittent fault; this could permit a spiral to develop through a chip containing such a fault. On the fault becoming apparent the wafer would reconstitute - through the same intermittent fault if this contact were once again operational. The probability of an intermittent fault is not known but it is thought to be unlikely that such faults would survive rigorous environmental test routines. It has been noted in Section 6.7.1 that these 3000 contacts may be substantially reduced - in particular a charge - coupled device design could possibly manage with a single grid supplying a suitable clock waveform. This reduced number of contacts would then require the solder bump process to be only ~ 95% perfect. Two contacts per chip would reduce this to ~ 90% probability of both contacts being good; this could be superimposed on a slice yield of ~ 80% to produce an assembled W.S.I array of ~ 72% yield.

A major problem with a rigid assembly is the mismatch of thermal expansion coefficients between the various package components. For example, quoted values for alumina (5.9×10^{-6} in range 25-200°C) and silicon (2.5×10^{-6}) would give rise to unacceptable stresses; in fact it is generally accepted e.g. Jowett[86] that full compatibility over this temperature range requires a match of TCE to within 4×10^{-7}. In the proposed flip-slice arrangement the silicon metallised surface would be under compression, thus producing a tendency to bow (the active surface of the slice becoming concave). Apart from any tensile stresses so induced there would be shearing forces on the bonds arising from the attempted movement (0.7 thou/inch at 200°C from stress-free condition) of the slice relative to the substrate. It is not known how such an assembly would stand up to this total movement of ~ 1 thou (edge to centre of a 3" slice) over 200°C. If it cannot be accommodated without weakening the solder bumps then it is necessary to use a substrate which can be made to match the silicon TCE tolerably well. A glass substrate would permit two other drawbacks of the alumina to be overcome; firstly alumina substrates tend to be bowed - the camber is quoted as 5 thou/inch while glass substrates are very much better in this respect. Secondly the glass would permit thin film techniques to be used; this would ease the problem of packing density inherent with the thick film procedure in producing several resistors and cross-overs in the space of each chip.

An alternative procedure would be the adaptation of a technique investigated by Kraynak[25] in a technique he called "Wafer Chip assembly". Here the flip-chipping of silicon dice onto a silicon motherboard was proposed, the author noting that

> "... the lack of uniformity which existed on both surfaces generally resulted in at least one dot* on a chip becoming separated from the corresponding substrate dot. To overcome this problem, electroplated gold was selected to obtain a uniform metal buildup on the wafer chip. The electroforming process which is essentially a 100-percent-yield process, produced pedestals with better than adequate height uniformity."

These "wafer chip" structures required all 110 dot contacts to function and working samples were produced. It is reasonable to suppose, therefore, that each dot had a virtually 100% chance of success. Even allowing for additional problems arising due to bow or camber when extending the technique to full wafers it is thought probable that this approach would meet the required limit of 2-5% failure rate per contact.

Some of these ideas have also been considered in the Discretionary Wiring projects. Flip-chipping onto both printed circuit boards and silicon wafers was intended; the latter, it was claimed[59], would provide advantages of "thermal match and small interconnect geometry possible with silicon/ photolith techniques". It is known that a printed circuit board was designed to accommodate both flipped chips and conventional packages but no further information is available.

* Eleven dot contacts per chip.

Non-Hermetic Packaging

To encourage the acceptance of Wafer-Scale Integration by the industry it is advisable that the technique should not require a major package development programme before it can be fully investigated. Further thought on the development of such a hermetic package is therefore probably a waste of effort at this time; if W.S.I proves viable the industry itself will then provide adequate resources for such development.

However, while a fully hermetic package would seem to require one of the expensive types of assembly procedure discussed above, non-hermetic packaging could be relatively easily achieved. Two suggestions are made in this respect although, again, neither has been investigated empirically owing to the high cost of the requisite specialised equipment.

For low power dissipation circuit technologies (to which the technology is best restricted in the first instance) the simplest structure would be to adapt the silicon wafer itself so that it could be plugged directly into a suitable edge connector, relying on forced air circulation for cooling. The conventional slice is unsuitable for this purpose firstly because it is round and secondly because it is too thin - at 5-10 thou thick silicon wafers are very fragile. However slices are produced to this geometry only because this suits the conventional process technology. Although silicon crystals tend to be approximately cylindrical as grown and are centreless ground to bring their diameter to a certain tolerance, they could easily be ground to square cross section and then sliced at an angle to produce large rectangular wafers. Surface orientation - critical for some MOS technologies - could be controlled by careful selection of crystal orientation and angle of cut. Although an increase of slice thickness to ~ 40 thou (to ensure adequate strength in handling) and the oblique cutting angle would mean fewer slices per crystal, less silicon would be wasted as swarf than with conventional slices. The cost of conventional slices (non-epitaxial) is less than £2; even if this increased to £10 for the proposed slices the new technology would still ofer tremendous cost advantages. Square 4 x 4" (polysilicon) wafers are already under investigation to save cost and increase packing density in solar cell technology.

The processing of square wafers would require jigs for furnacing, wet processing and photoengraving to be modified, but the only essential difference from the standard technology would be the addition of edge-connector-compatible contacts. Choice of metal here would be influenced primarily by reliability of contact and avoidance of undesirable intermetallics with the aluminium metallisation.

Another approach which has received serious consideration is illustrated in Fig. 6.10. Here it is proposed that earlier work* on the interconnection of discrete chips into a large array using a premetallised polyimide film could be adapted to a full slice technology. In fact the major problems of this polyimide film technology (difficulty in

* ACTP Contract K/78b/332 with ICL.

accurate positioning/alignment of chips and the requirement for all bonds to make good contact) would be automatically overcome in applying it to a fault tolerant full-slice technology. This technique avoids the problem of mismatch of thermal expansion coefficients by the introduction of a flexible substrate and superstrate to replace the rigid components required for the assemblies described earlier.

The polyimide film superstrate could be designed to carry the power supply/clock grids with all X lines formed on one surface and all Y lines on the other. The additional interconnections required for the toroidal configuration could also be made via this superstrate. Apart from simplifying the on-slice metallisation (and thereby increasing yield and permitting closer packing of chips) this procedure would eliminate the requirements for diffused crossovers in the power/clock grids. A further advantage would be the facility for testing the global (full wafer) metallisation independently to ensure the absence of interclock short-circuits before committing each polyimide film to a wafer.

While these techniques do not have the advantage of hermeticity it must be noted that hermetic packaging is not an end in itself but a means to an end - that of reliable device operation in an alien environment. It may well prove more economical when dealing with such highly complex units as full slice arrays to achieve the necessary degree of hermeticity by suitable packaging of subsystems containing many such (non hermetic) wafers rather than by the hermetic encapsulation of each W.S.I device.

6.8.3 Gate Protection of W.S.I Devices

The high impedance of the gate dielectric of an isolated MOS transistor ($\sim 10^{14}\ \Omega$) allows charge - carried by circulating air, for example - to build up on the gate metal over a prolonged period (seconds to hours); the gate-substrate capacitance may thus eventually attain a voltage sufficient to exceed the maximum field which the gate dielectric can withstand (typically 100V for 1000 Å silicon dioxide). This problem and the more obvious cause of device destruction due to discharge of static electricity from a person's fingers are well known to MOS device manufacturers. The usual solution is to incorporate a system of gate protection which may vary from a simple avalanche diode (to preclude the possibility of a slow build-up of charge and, hopefully, to prevent spark discharge from destroying the device) to an extensive diode/resistor/ capacitor network. Using this it is possible to slow the voltage rise across the gate dielectric sufficiently for the diode to shunt the charge to ground before a disastrous field is reached across the gate dielectric; it is thus possible to raise the tolerance of MOS devices to static electricity to the levels at which bipolar devices fail[30, 82].

In general the greater the capacitance at a circuit node the less is the chance of damage by static electricity. Internal circuit nodes do not normally require protection in packaged devices although it is feasible that wafer-scale memories may (dependent on packaging) be more prone to damage owing to the possibility of air circulation over the chips, with consequent build-up of static charge on the slice surface. However, the presence of the power supply grids on the slice surface may well inhibit such charge accumulation and so render protection unnecessary, particularly if the slice is overlaid with a thick protective oxide.

The input/output terminals of the first chip in the chain are exposed to the outside world and, in a conventional device, would each carry protection networks. This, is primarily a concession to the fact that users in general do not take adequate precautions in handling MOS devices - certainly when very high input impedance MOS transistors are required for specialised applications these can be made and used quite satisfactorily without any gate protection provided that care is exercised in their handling.

The inclusion of gate protection networks on every array chip would be very wasteful of space. However, the next section considers the advantages of a special input/output interface chip, which could carry protection networks for the first (exposed) array chip without an expensive overhead in silicon area on all other chips.

6.8.4 The Input/Output Interface

Two distinct approaches may be made to the assembly of W.S.I serial memory devices; one may either standardise on a particular location for the input/output or one may choose to vary the input/output position to guarantee connecting to a good chip. The latter approach has the advantage that no wafer need be rejected owing to a faulty first chip (e.g. at 75% dice yield ~ 25% of wafers would be rejected) but the former allows a standard package to be used and requires no wafer testing prior to assembly; this is the preferred approach.

If the decision is taken in favour of a fixed i/o location in the array it is further proposed that a special i/o interface chip should be included at this preset site in the array of chips. This would have several advantages, thus offsetting the minor drawbacks of the loss of one memory chip and the committal of the fate of the wafer to one chip which has to be good.

Firstly, the interface chip could carry substantially larger i/o bonding pads at larger separation than would be practicable if all chips were to be identical. Secondly it would permit elaborate input gate protection networks and output expansion buffers to be incorporated on the wafer. Thirdly, test components could be included at one site only, thus saving space on the array chips.

This i/o chip will appear to the developing spiral as an additional bad chip in the array and so could well assist in extinguishing the spiral during its critical early stages of growth. It should be noted, however, that provided the i/o chip is sited (say) to the North of the first spiral chip it will not restrict the choice of OPEN directions available to that chip provided that its instructions from chip Z assume that the signal has originated from a chip to the North. (The control logic prevents any spiral chip from attempting to select the same output (OPEN) direction as its input (LOOK) direction).

It has so far been assumed that this i/o chip is dedicated to a fixed adjacent first chip; in this situation the failure rate for wafers (owing to a bad i/o chip and/or a first spiral chip) will be higher than 25% at a 75% dice yield level. However, the probability of the (relatively simple) i/o chip being good is very high and so the number of wafers rejected owing to a faulty first chip can be drastically reduced if the i/o chip is permitted to access all four adjacent chips. This requires the i/o chip to carry the OPEN/LOOK control logic so that the four adjacent chips can be sequentially accessed. Note that it is not permissible to abandon this logic on the i/o chip and feed the data to all four directions simultaneously as each adjacent good chip, although prevented from implementing commands, will accept the data stream - in parallel with the true spiral - when its relevant i/o port is enabled, thus corrupting the data.

Summarising, the advantages of a special i/o buffer chip at a fixed point in the array are

i) Standardised assembly

ii) No wafer testing prior to assembly

iii) Larger i/o bonding pads, more widely separated

iv) Gate protection networks on i/o lines

v) Test components and production line aids on all wafers

vi) Output expansion capability on wafer

vii) Reduced wafer rejection rate (for standardised assembly)

6.8.5 Area Requirements of W.S.I Control Logic and Power/Clock Grids

The detailed layout of the p-channel enhancement mode 4ϕ dynamic MOS chip showed that the on-chip control logic for the serial memory required an area of silicon approximately equivalent to 256 shift register stages, thus demonstrating that it becomes a negligible overhead on shift register chips of ~ 5K bit complexity. The arrays of section 6.4 require considerably less on-chip control logic. The power/clock grids must, however, also be included in any calculation of "wasted" space in W.S.I. arrays. Fig. 6.11 sketches the layout of a W.S.I "chip" and shows that the residual area after allowing for five orthogonal grids on a 100 thou square chip is ~ 8,464 sq. thou. Estimating the serial memory control logic at 1600 sq. thou for a CCD process leaves an area of 6,864 sq. thou available for memory; or alternatively, 31.4% of the chip area is wasted. This may be compared with a conventional device (sketched in Fig. 6.12) as follows.

A conventional chip must be scribed; this requires grid lines which are typically 4 thou wide. A 3 thou oxide border is normally left between the outer edge of the aluminium bond pad and the grid line; this allows some misplacing of the wire bonds without short-circuiting to the silicon in the grid line region and reduces the chance of microcracks (which spread in from the scribed edge of the chip) penetrating under the bonding pad area. The bond pads themselves are normally 4 thou square and are separated from the active device by a 3 thou wide oxide border - again to allow for misplacing of the wire bonds. As bond pads are normally placed along all four chip edges this reduces the active area to 76 thou square or 5,776 sq. thou. That is, the conventional route wastes 42.4% of the silicon area. Protagonists for the conventional technology would argue that not all this border is wasted, active components are often sited between the bond pads. This is true, but these components tend to be test devices, alignment marks, chip identity codes, gate protection networks and output driver stages - none of which are required on the array chips of W.S.I. devices.

One might therefore suggest that the control logic and associated global interconnections on W.S.I arrays fit easily into the grid line area wasted on conventional chips. This is not strictly a true comparison as the chip size on W.S.I. arrays will generally be smaller (and hence the overhead per chip larger) than the discrete chips which the array replaces. However if this factor is taken into account then the greater wastage of silicon by these lower yield larger chips must also be considered.

6.8.6 Comparison of Conventional and W.S.I Memory

It is important to note that W.S.I is not in competition with ccd or any other semiconductor technology in the manufacture of large memories. Within the limitations imposed primarily by power, W.S.I is complementary to and can be used in conjunction with any such technology to enhance its capability for large arrays. However, in stressing this technology independence it is accepted that W.S.I is, of course, in direct competition with techniques to which W.S.I is inapplicable - for example bubble memories. The relative costs, advantages and applications of relevant memory types are now compared.

Proebsting[44] notes that MOS memory first became cost-competitive with core store with the advent of 4096 bit n-channel MOS devices. The general decline in memory cost per bit with increasing chip complexity is noted by Noyce[40]; his predictions for the costs of 1K, 4K, 16K and 65K RAMS to 1983 are reproduced in Fig. 6.13. The memory size increase offered by W.S.I would extend this plot downwards by two further curves (each a factor of 4 increase in complexity) on a 1M bit system, although it must be expected that the high initial investment for W.S.I assembly technology would place the starting point of the "learning curve" above that of the conventional 65k bit devices.

In comparing microelectronic memories, Hodges[25] notes that CCD and magnetic bubble devices are aiming at the same gap between MOS RAMs and moving surface magnetic memories - as summarised in Fig. 6.14, extracted from this paper. The recent massive increases in magnetic bubble memory sizes* and consequent reduction in cost per bit must give a highly competitive edge to magnetic bubble technology which can only be balanced by a similar increase in complexity of CCD memory chips.

Baker[3], however, suggests that CCD and magnetic bubbles are complementary rather than competing technologies. He notes that the main advantages magnetic bubbles offer

* as demonstrated, for example, at Electronica 1978 Exhibition, November, 1978.

over CCD's are

1) Nonvolatility in the event of stopping the clocks or even of power failure.

2) Potentially two to four times more bits per chip (no longer the case if W.S.I is applied to CCD technology.)

3) Possible reliability advantage arising from (a) simpler processing and (b) fewer chips for a given storage capacity. (Point (b) would become invalid with point (2)).

He quotes fast random access of data blocks as the main advantages of ccds. Such block accessibility renders the device ideal as a "swapping memory", useful as an interface store between the very large (but slow) multimegabit disc or drum and the very fast (50 ns) caché memory accessed by the cpu. A page of memory is dumped into the serial parallel serial CCD memory, thus avoiding the CPU having to wait for data availability.

For manufacturers offering both technologies this claim of non-competition may be valid. It should be noted, however, that as manufacturers of moving magnetic media memories find their markets increasingly eroded by ccd and magnetic bubble technology they may well find it far easier to develop an in-house facility based on magnetic bubbles rather than ccd or other semiconductor technology, thus designing semiconductor memory out of this market sector.

Magnetic bubble memories have achieved their relatively large size (in comparison with ccd structures) partly by greater sophistication in design. The inclusion of redundancy using a special bubble loop to store the locations of defective areas in the main bubble loop has enabled the fact that faults are present to become transparent to the user. The absence of the corresponding refinement on conventional semiconductor memory (obtainable with W.S.I) has restricted their size to currently "64k bit"; even this size is causing problems - Intel have recently withdrawn their 64k bit ccd device.

However, semiconductor memory has two major advantages over magnetic bubble devices. Firstly, the potential packing density of 10^8 bits per square inch of ccd is far greater than the 10^6 bits per square inch of magnetic bubbles. These figures are quoted by Toombs[56] who also suggests that 4M bit ccd memories will be available as discrete chips 330 thou square by 1986.

The second major advantage of semiconductor technology over other storage techniques is that data processing logic can be far more efficiently implemented in this technology than in any other. If, for example, the parallel/serial W.S.I array contains data processing logic on each chip then it is elevated from passive memory to a digital array processor. Alternatively, a microprocessor chip can be included in the array, updating the look-up table or address buffer (see Section 6.4.1) if it detects a fault in any storage location, thus maintaining an apparently perfect memory for its data processing. This type of structure, although dependent on the survival of the microprocessor and therefore no longer strictly failure tolerant, would seem to offer great advantages over discrete microprocessor chips and associated 64K bit RAM's. Applications of this nature would guarantee for the semiconductor industry a substantial market sector, safe against attack by such techniques as magnetic bubbles. Unless memory sizes are increased very rapidly (perhaps by wafer-scale integration) the future looks bleak for passive semiconductor memory over the lower end of the speed range. Rockwell International has recently announced* a 256k bit bubble memory and expects "to produce a 1M bit device during 1979 and is developing 4 and 16M bit devices for production in the latter part of the 1980's". The company anticipates a market for bubble memories of "$ 500M annually by 1985[39]".

While such devices have reduced volatility and different operating power level, temperature range, speed, radiation resistance etc. and are therefore ideally suited to a different market sector from semiconductor memory, the availability of cheap bubble memory must lead to erosion of the semiconductor memory market unless the latter can be made cost-competitive by substantial size increases.

6.8.7 A Cost Comparison of a 1M-bit Serial Semiconductor Memory in W.S.I and Conventional Technology

The major costs to the stage of saleable product are now compared for the conventional and W.S.I routes to a 1M bit serial memory. These figures are based on information supplied during discussions with the industry.

* Electronica 1978 Exhibition, November 1978.

The costs are extrapolated to a proposed 64K bit CCD memory - anticipated for 1980 - and it is assumed that 15 working chips can be obtained on a 3" slice. Table 6.1 compares these costs.

Process Stage	Cost in £	
	Conventional Route	W.S.I. Route
Processed Slice	S	S
Slice Test	T_1	-
Scribe and Break	Scr	-
Assembly and Package (chip)	A_c	-
Assembly and Package (slice)	-	A_s
Device Test	T_2	-
Wafer Test	-	T_3

Table 6.1

Cost advantage of the W.S.I 1M bit memory to the stage of tested, assembled device (ignoring additional cost of devices found to be faulty after assembly) is

$$T_1 + Scr + N A_c - A_s + N T_2 - T_3$$

(where N ≡ number of good packaged devices).

This assumes that the two routes provide equal amounts of memory (~ 1M bit). Estimated figures for the costs in the above equation are:-

$T_1 = £\ 0.50$	$A_S = £\ 4$
$Scr = £\ 0.50$	$T_2 = £\ 0.1$
$N = 10^6/64 \text{ x } 10^3 = 15$	$*T_3 = £\ 2$
$A_C = £\ 1$	

* assumes that one chip Z (external control electronics, costing ~ £100) can set up and maintain 100 wafers and allows £1 overhead per wafer to cover running costs.

$$
\begin{aligned}
\therefore \quad \text{Cost saving of W.S.I} &= 0.5 + 0.5 + (15 \text{ x } 1.0) - 4 + (15 \text{ x } 0.1) - 2 \\
&= 1 + 15 - 4 + 1.5 - 2 \\
&= 17.5 - 6 = £\ 11.5 \text{ (for 1M bit)} \\
&= 1.15 \text{ millipence per bit.}
\end{aligned}
$$

The percentage saving on manufacturing costs depends critically on the slice cost, S, to be applied in the equation:-

$$
\begin{aligned}
\text{Cost (conventional route)} &= S + T_1 + Scr + NA_C + NT_2 \\
&= S + 17.5\ (£) \text{ for 1M bit of memory}
\end{aligned}
$$

Estimates of cost of processed slice vary from £10 to £50 depending on how the overheads of running a production line are amortized. Taking an estimate of £20 as the "true" cost of a processed slice, projected cost for the conventional route becomes

£ 20 + 17.5 = £37.5 or 3.75 mp/bit.

Thus the W.S.I approach represents a cost saving of ~ 30% in terms of manufacturing cost of saleable product on the basis of these figures. The profits will be further increased when the advantages to the user are partially discounted in an increased selling price for the W.S.I product.

The cost advantages to the user are now estimated in relation to a 32M bit system consisting of, firstly, 16 printed circuit boards, each containing 32 integrated circuits (64K bit CCD's) and, secondly, 32 W.S.I devices, each of 1M bit. Memory purchase price per bit is assumed equal for the two approaches

i) Conventional Approach

Printed circuit board for 32 packages	£ 50
Control board for 16 p.c.bs	£150
Edge connectors for 16 p.c.bs	£ 20

Total installation cost = £(50 x 16) + 150 + 20 = £970

ii) W.S.I Approach

Control board for 32 wafers	£150
Edge connectors for 32 wafers	£ 40
Chip Z (one-third committed)	£ 33

Total installation cost = £150 + 40 + 33 = £223

The installed cost is therefore reduced by a factor of > 4. This rises to a factor > 8 for 2M bit spiral.

It should be noted that these calculations provide a pessimistic figure for the cost advantage of W.S.I. In particular it is assumed that only the same storage capacity is available to the W.S.I device as is obtained on the 15 discrete chips - yet one of the great advantages of W.S.I is its much-improved utilisation of silicon area arising from the higher yield of the smaller chips employed. It is further assumed that no chips or packages are damaged during the assembly of the 15 integrated circuits. Taking these factors into account, one may expect the installation cost advantages of 2M bit - plus serial W.S.I memories to be at least a factor of 20 and to be ever-increasing as slice diameters increase.

In assessing relative costs of W.S.I and conventional memory it is important to consider also the reduction in post-installation costs arising from W.S.I. It has been noted (Table 3.1) that system failure in the field is an expensive item. It is therefore suggested that additional cost savings will arise from the large potential improvement in reliability offered by W.S.I (discussed in Section 3.2). In addition to this increased MTBF - arising from fewer individual packaged components - the graceful degradation feature of such devices (noted in Section 6.5) will often permit renewed operation (without manual intervention) of "failed" components.

A further system cost reduction will arise from the improved packing density and reduced wiring complexity offered by W.S.I, as discussed in Chapter 3.

6.9 INDUSTRIAL ACCEPTABILITY OF W.S.I.

The major difficulty in persuading manufacturers to take up the serial memory structure is the likely high cost of developing a suitable package to the stage of a fully proved, marketable product. While there may well be great advantages to be gained from the proposed changes in computer architecture described in Section 3.4.2, it must be accepted that such drastic changes can occur only slowly and the industry cannot be expected to invest large sums in a crash development programme on such devices. Manning[34] notes that

> ".... many proposed arrays remain paper-studies for various reasons, including impracticality and IC industry inertia."

It is ironic that Manning's work should also have remained - as far as can be ascertained - a paper study. "IC industry inertia" derives primarily from the fact that no manufacturer can afford to make a serious error in the choice of its major technology. Those companies for example which sat back and waited to see the outcome of discretionary wiring did not lose money on that technology.

However it is essential that the general concept of full-slice technology should not also be branded as unworkable as a result of the costly venture into discretionary wiring; the concept of the full wafer as the packaged unit was then, and still is, a very worthwhile goal.

It is a tenet of this thesis that, while the structures proposed in this work may also remain paper studies, wafer-scale integration must eventually be applied to realise the full capabilities of semiconductor technology in memory (and other)applications.

W.S.I technology is ideally suited to iterative arrays of logic-in-memory. The potential advantages of Distributed Array Processing are well known and many structures have been proposed. Such highly parallel processing systems have been discussed by Lewin[67] and are currently under investigation by ICL[47]. However, it must be accepted that requirements involving a simultaneous change of both product function and computer architecture will be treated with great caution by device manufacturers who do not have an in-house requirement for such devices. Whereas IBM, for example, may decide autonomously to move into a new field of computer design and change their device structures to suit, those (many) IC manufacturers not involved in the design of computers must tread more warily.

To overcome the industry's natural apprehension of a new technology it is therefore important to present a structure of minimum costs and problems of implementation with obvious application and immediate appeal to the industry. It is believed that the parallel/serial RAM array described in Section 6.4.1 would meet these requirements. For applications where associative memory is more appropriate than RAM (see, for example, Lea[29]) a structure of the type proposed in Section 6.4.2. would provide associative access to large blocks (say 1-5K bits) of data.

Lewin[31] notes the potential utilisation of ROM devices in the implementation of combinational and sequential circuits. The direct implementation of many - variable problems would perhaps be eased with the large structures available in W.S.I.

Another possible introduction of W.S.I technology is via a product which cannot currently be made economically with conventional technology. Analogue to Digital (A/D) converters may provide such a vehicle. Timko and Holloway[55] note an "increasing number of applications requiring fast converters with resolution and accuracy of 12 bits or more".

Even their 12-bit A/D converter requires two chips. Saul[46] notes that "solutions to the problems of cascading chips have now been found", so it is suggested that W.S.I technology may well have application to this field of high speed, high accuracy converters. For example a 128 chip array of 8-bit converters could perhaps provide 15 bit accuracy.

While there may well prove to be undiscovered problems associated with wafer-scale integration of semiconductor devices it is believed that the enormous potential return on the investment justifies the risks involved.

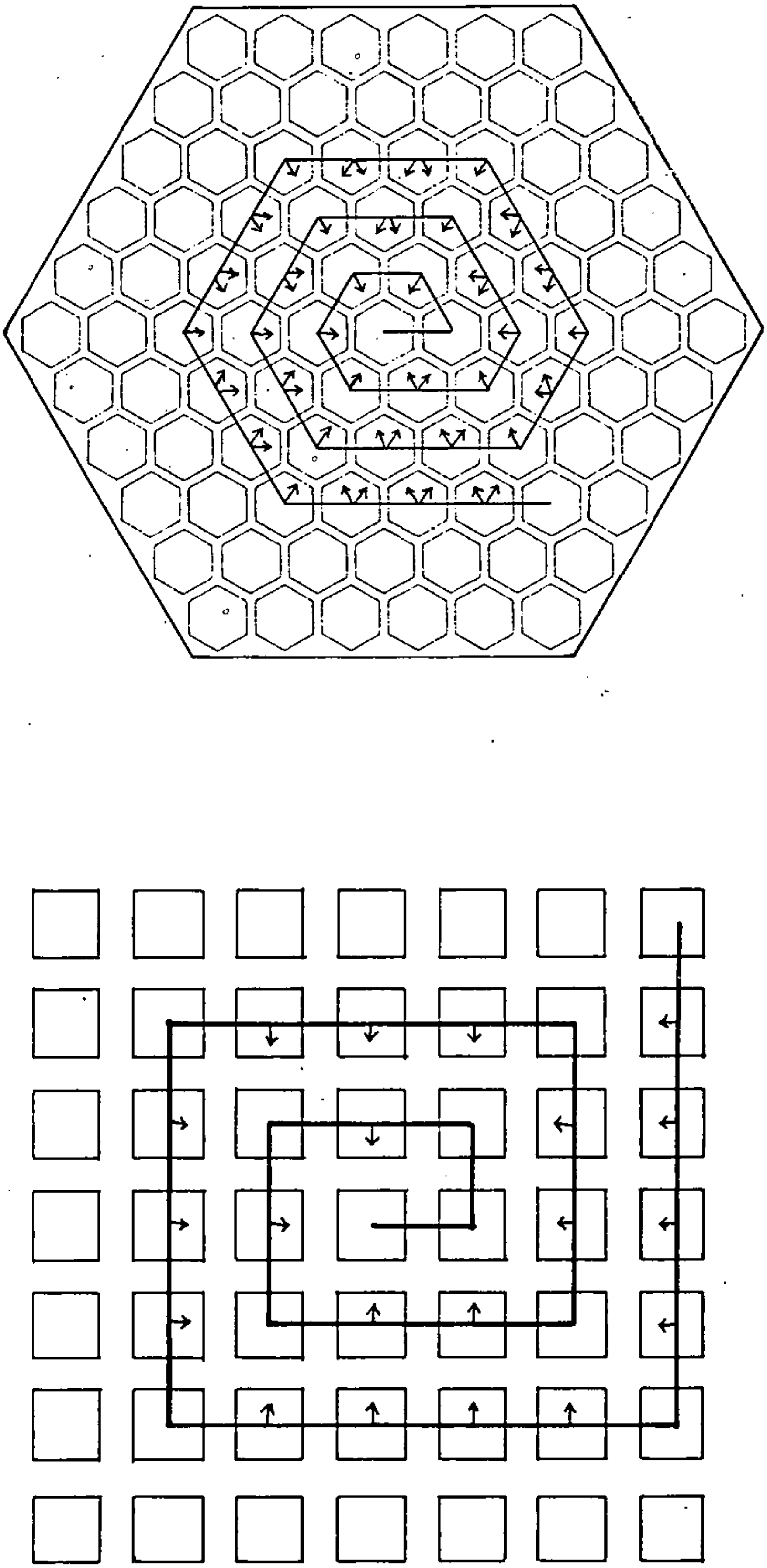


FIG. 6.1. Wasted attempts during algorithm progression.

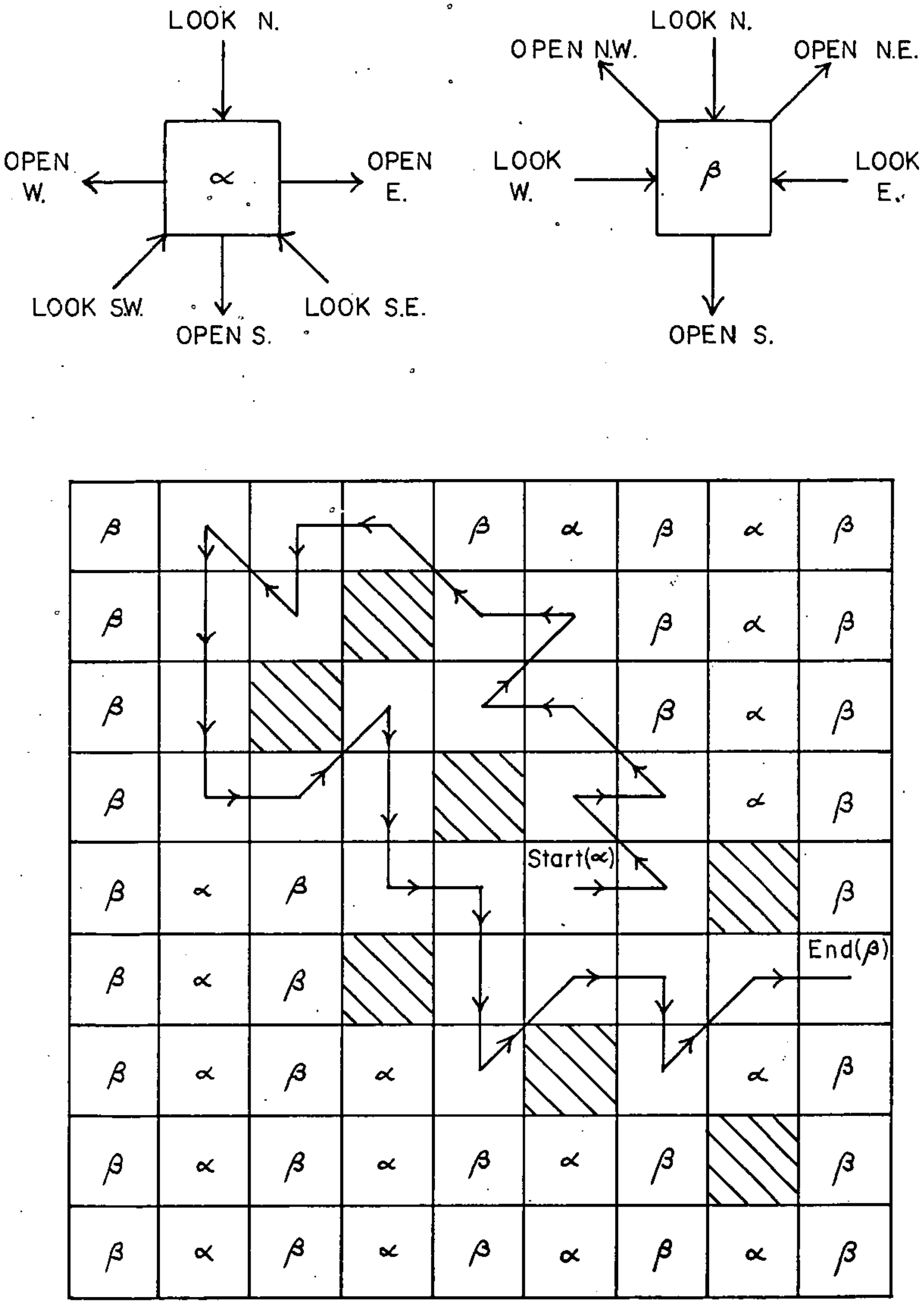


FIG. 6.2. Two-chip type of 3-way algorithm.

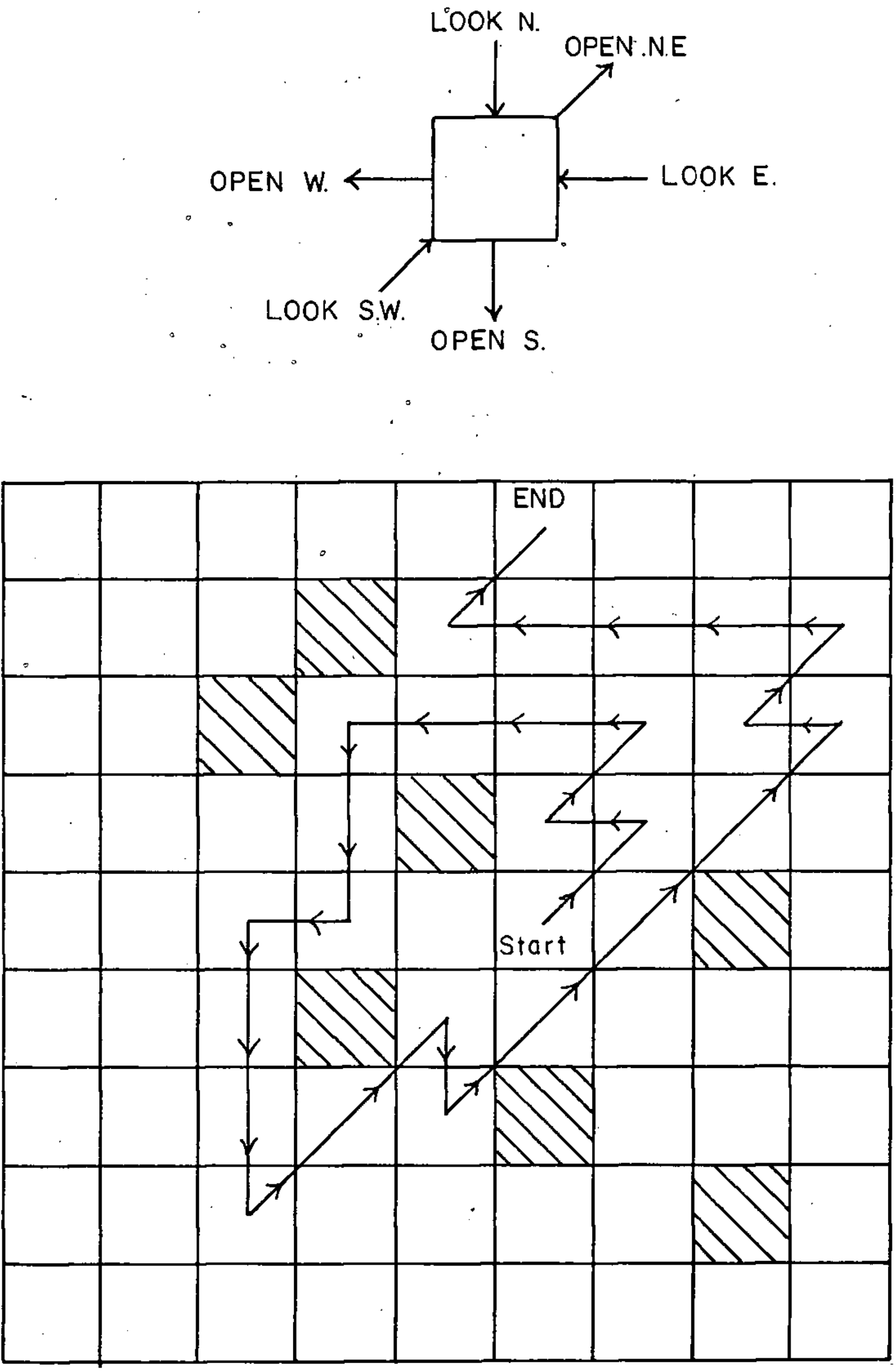
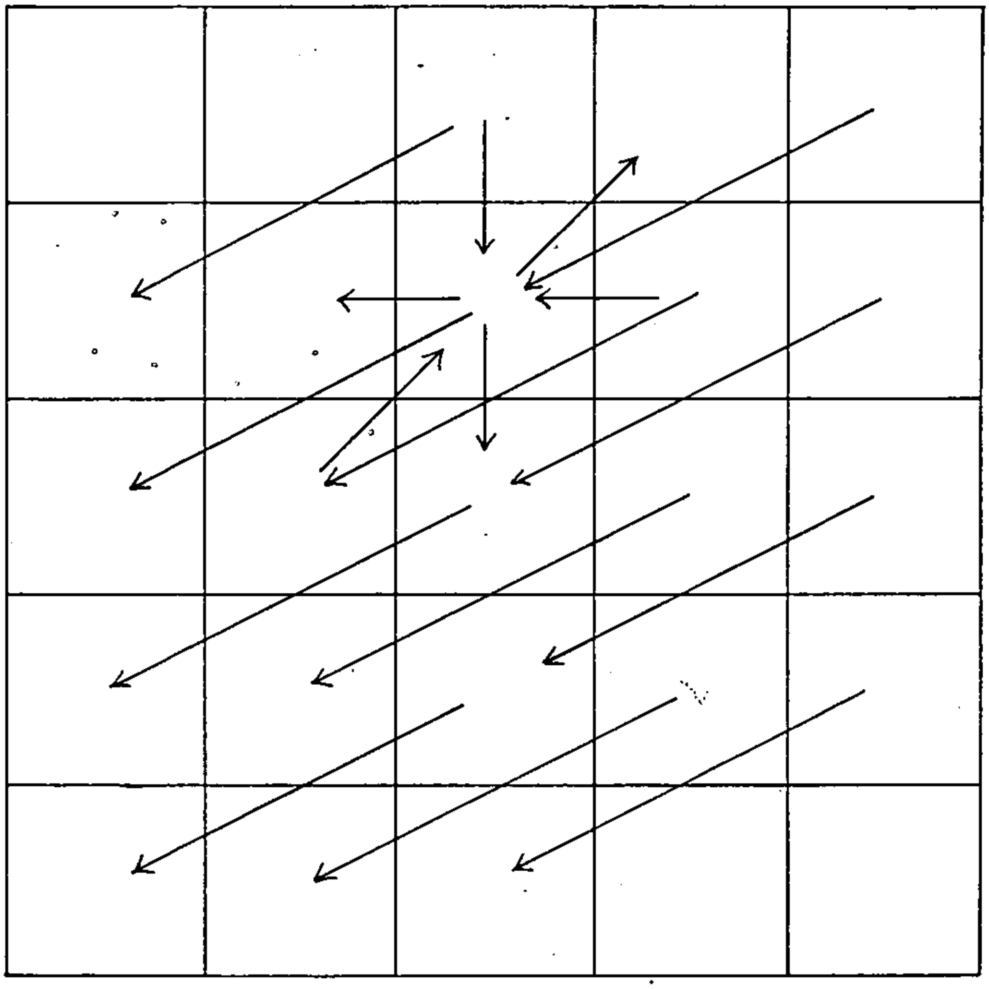
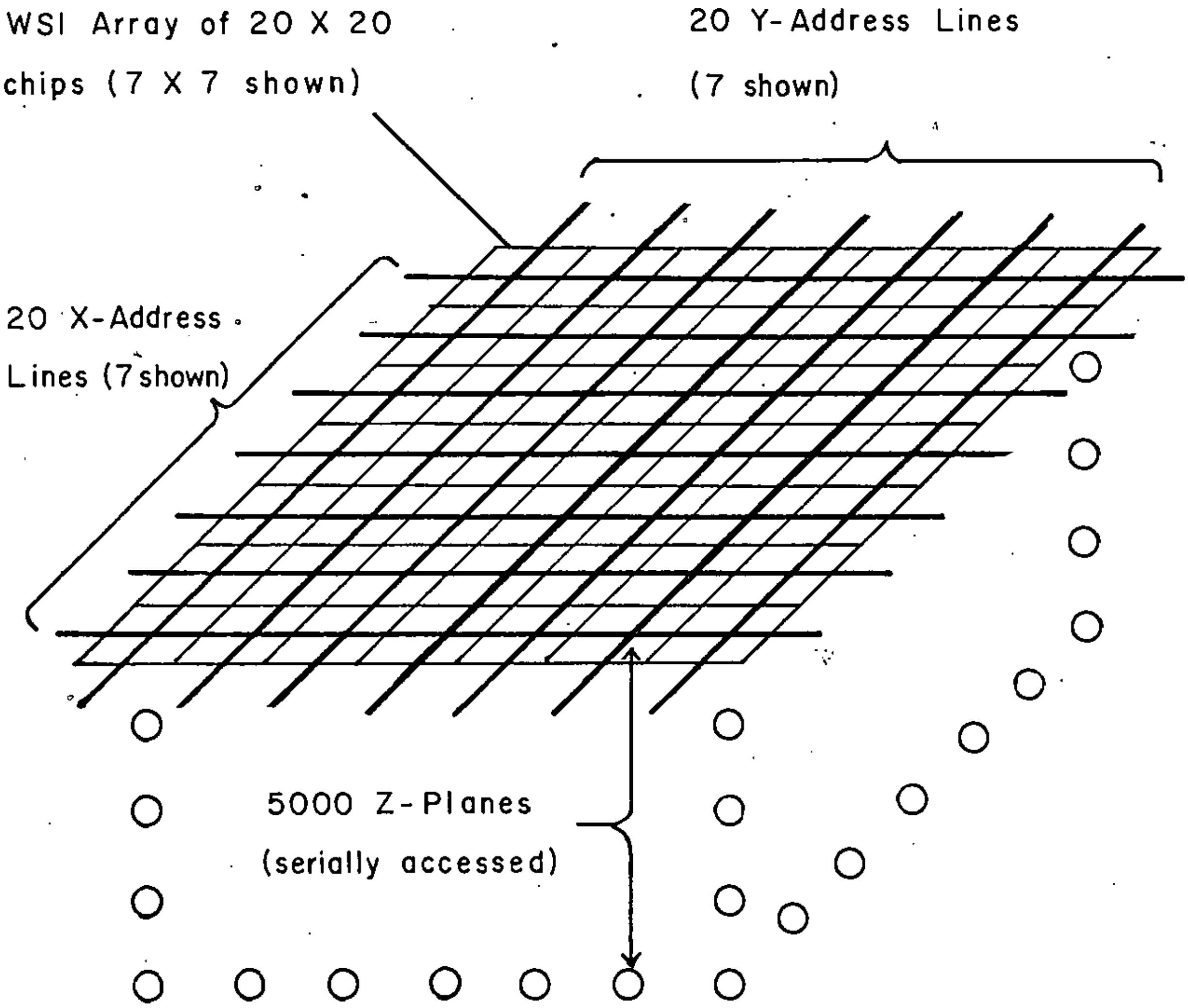


FIG. 6.3.   Single chip type of 3-way algorithm.



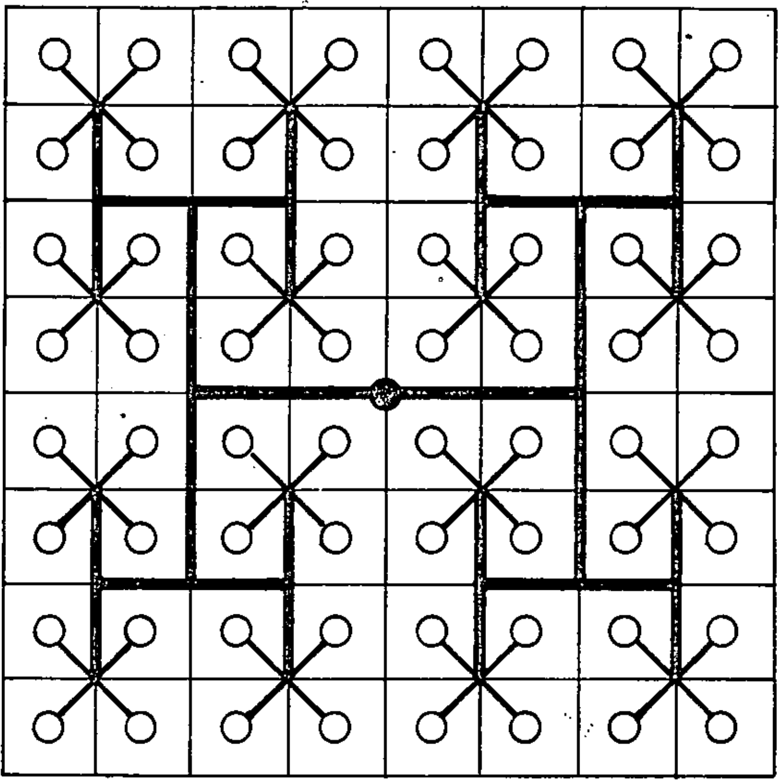
Note: This is for a single chip type - as used in Fig 5.27 - and hence one direction of leap. This is a serious limitation; in the above case a blind alley to the East of the main body of the spiral could not be exited by leap without extensive backtracking. With two chip types (giving two leap options) the interconnection problem increases considerably.

FIG. 6.4. Effect of Leap on interconnection complexity.



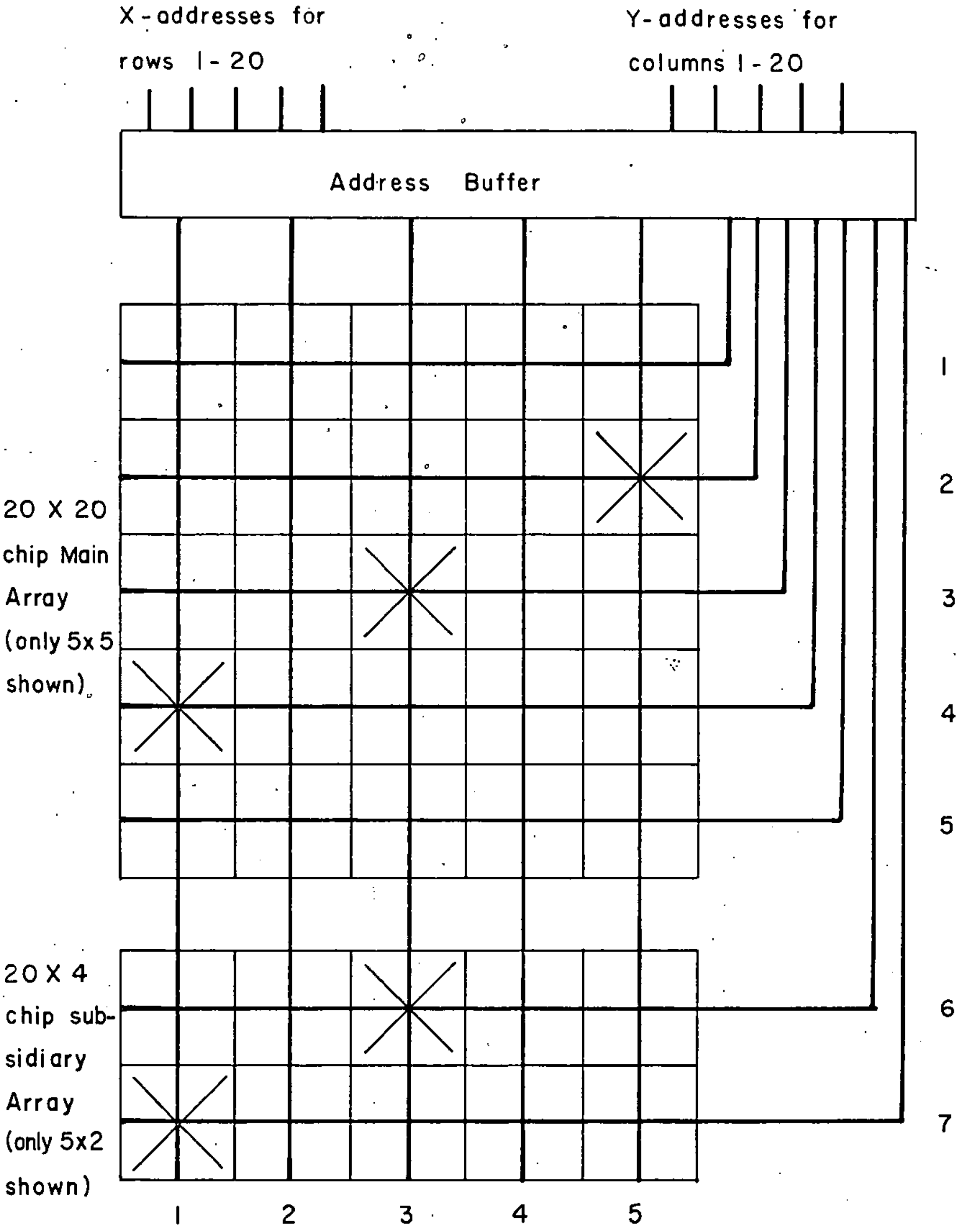
Each chip contains a 5k-bit serial shift register and has a unique X,Y address which accesses that 5k-bit serial character string. A 7 X 7 array is shown for simplicity but 20 X 20 would be more probable.

FIG. 6.5.. Parallel/serial WSI RAM array.



"Dyadicville" has several interesting properties - e.g. there are no cross-roads (all intersections being T-junctions on the true Dyadicville plan) and the route instructions for finding any house from any other are implicit in the two addresses (house numbers). These properties may well prove to be advantageous to logic-in-memory devices. However, the main feature of possible advantage to the passive memory described here is the fact that all houses are equidistant from the school (located at the centre of the town) - i.e. signals reach all chip locations simultaneously.

FIG. 6.6. 64-chip array structure based on the mathematical model town of Dyadicville.


X-addresses for rows 1-20
Y-addresses for columns 1-20
Address Buffer
20 X 20 chip Main Array (only 5x5 shown)
20 X 4 chip subsidiary Array (only 5x2 shown)
1
2
3
4
5
6
7
1
2
3
4
5

Input addresses may be coded on 1 (serial) or 5 (parallel) X-lines and 1 or 5 Y-lines. Address Buffer activates relevant wafer X_m and Y_n line.

FIG. 6.7. Address Buffer for Parallel/Serial RAM array.

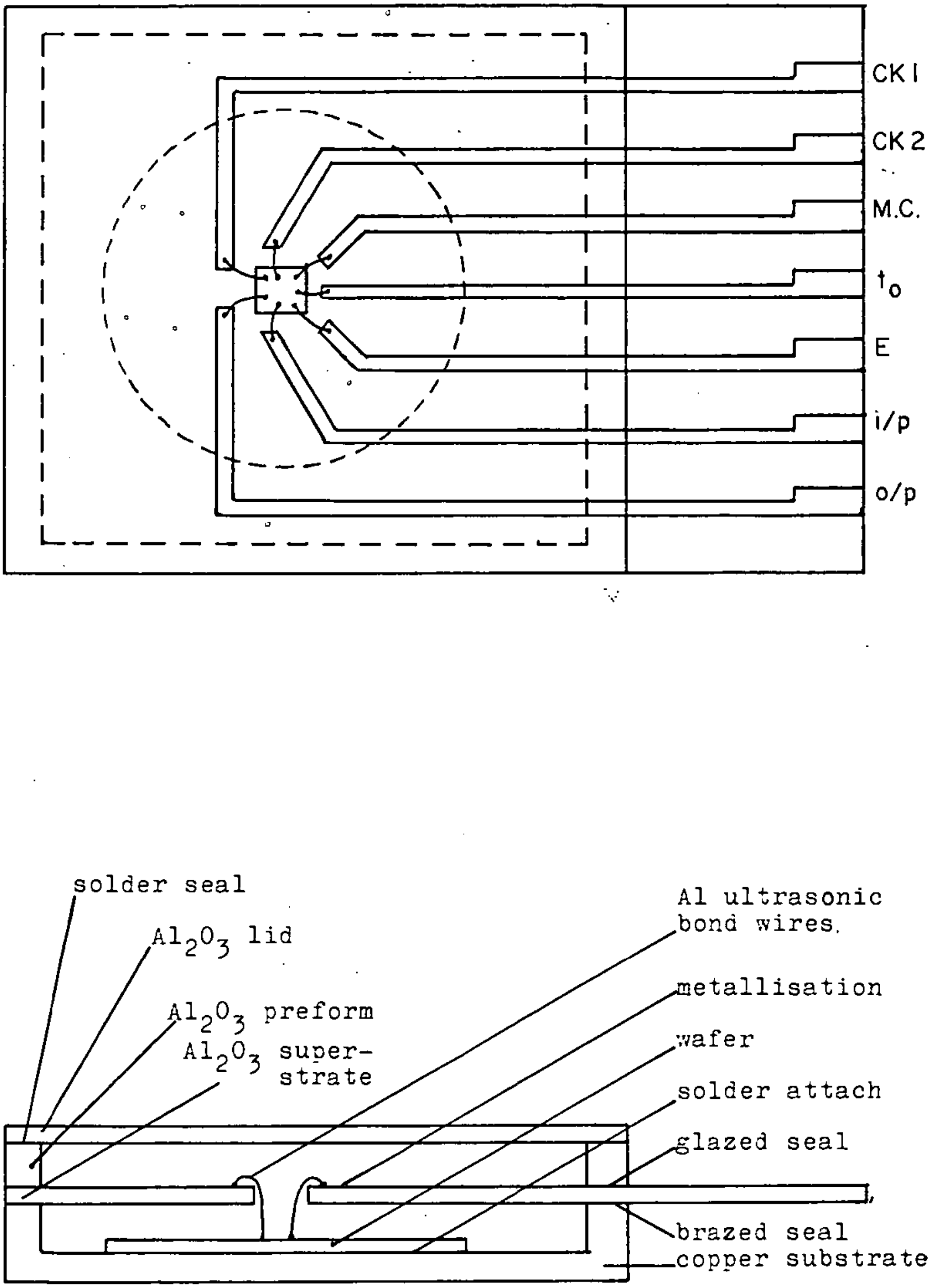
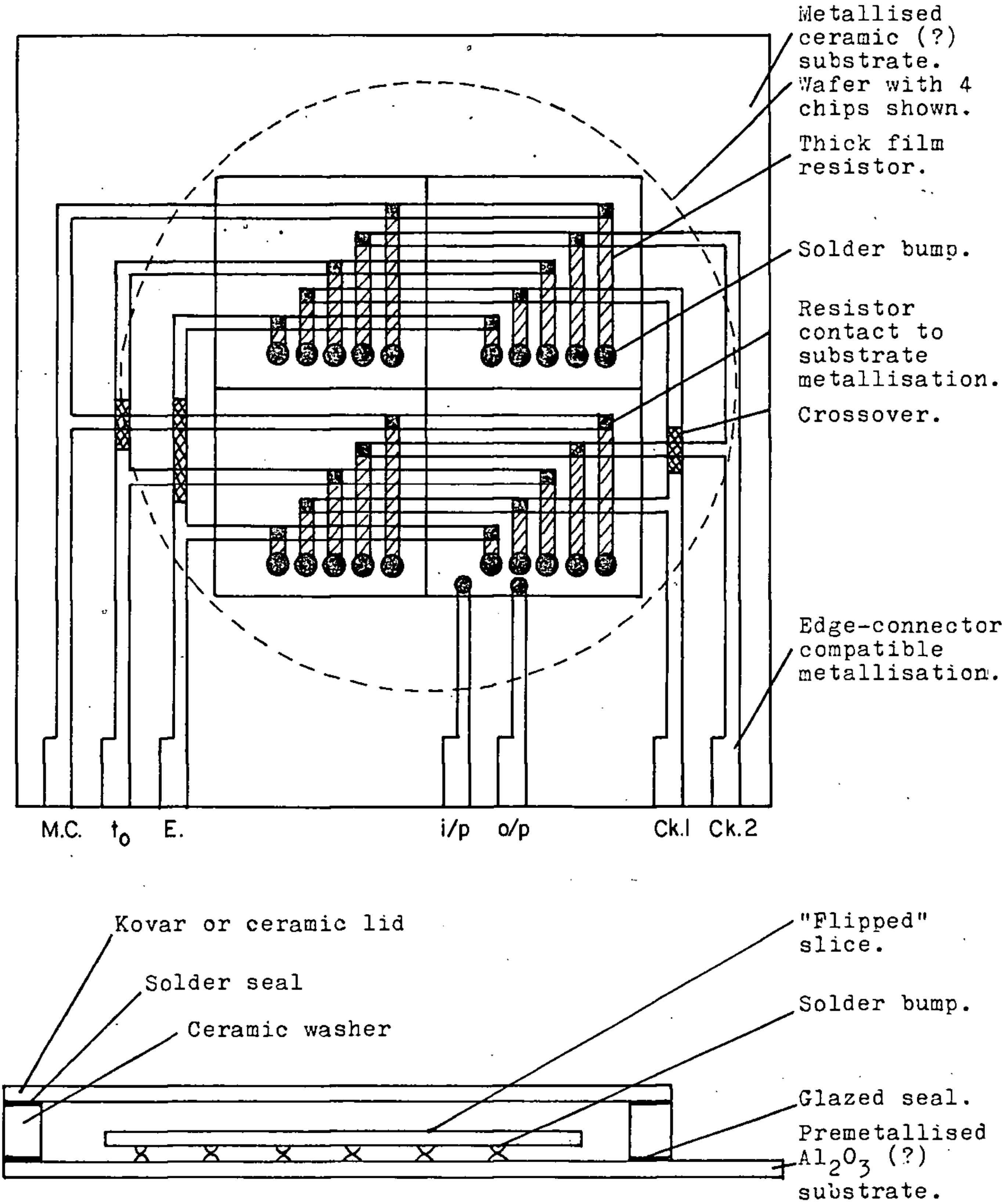


FIG. 6.8. "Picture-frame" assembly.


Metallised ceramic (?) substrate.
Wafer with 4 chips shown.
Thick film resistor.
Solder bump.
Resistor contact to substrate metallisation.
Crossover.
Edge-connector compatible metallisation.
M.C.
t_0
E.
i/p
o/p
Ck.I
Ck.2
Kovar or ceramic lid
Solder seal
Ceramic washer
"Flipped" slice.
Solder bump.
Glazed seal.
Premetallised Al_2O_3 (?) substrate.

Note:- Metallisation to convert array to toroidal configuration not shown for the purpose of clarity.

FIG. 6.9. Proposed "flip-slice" assembly.

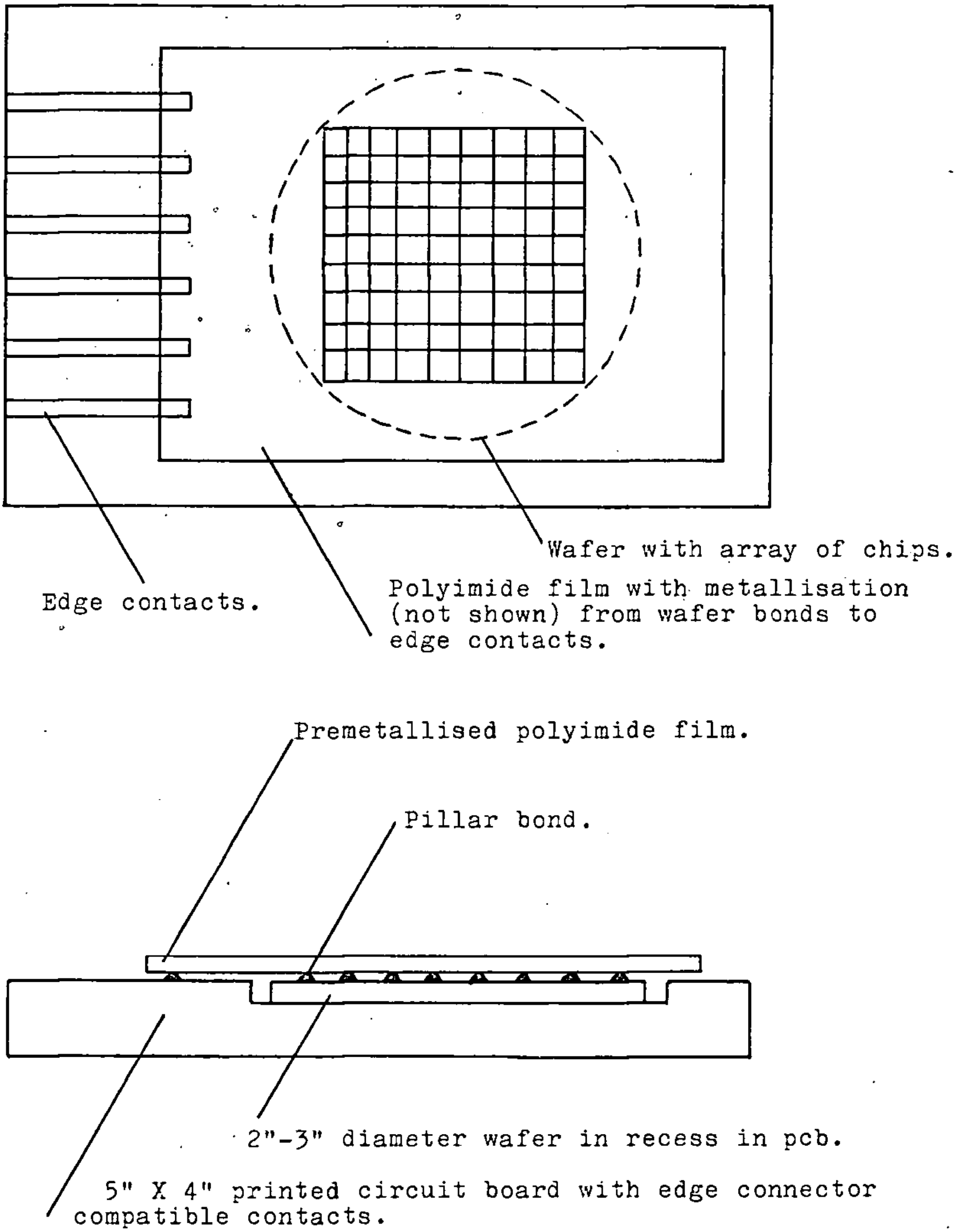
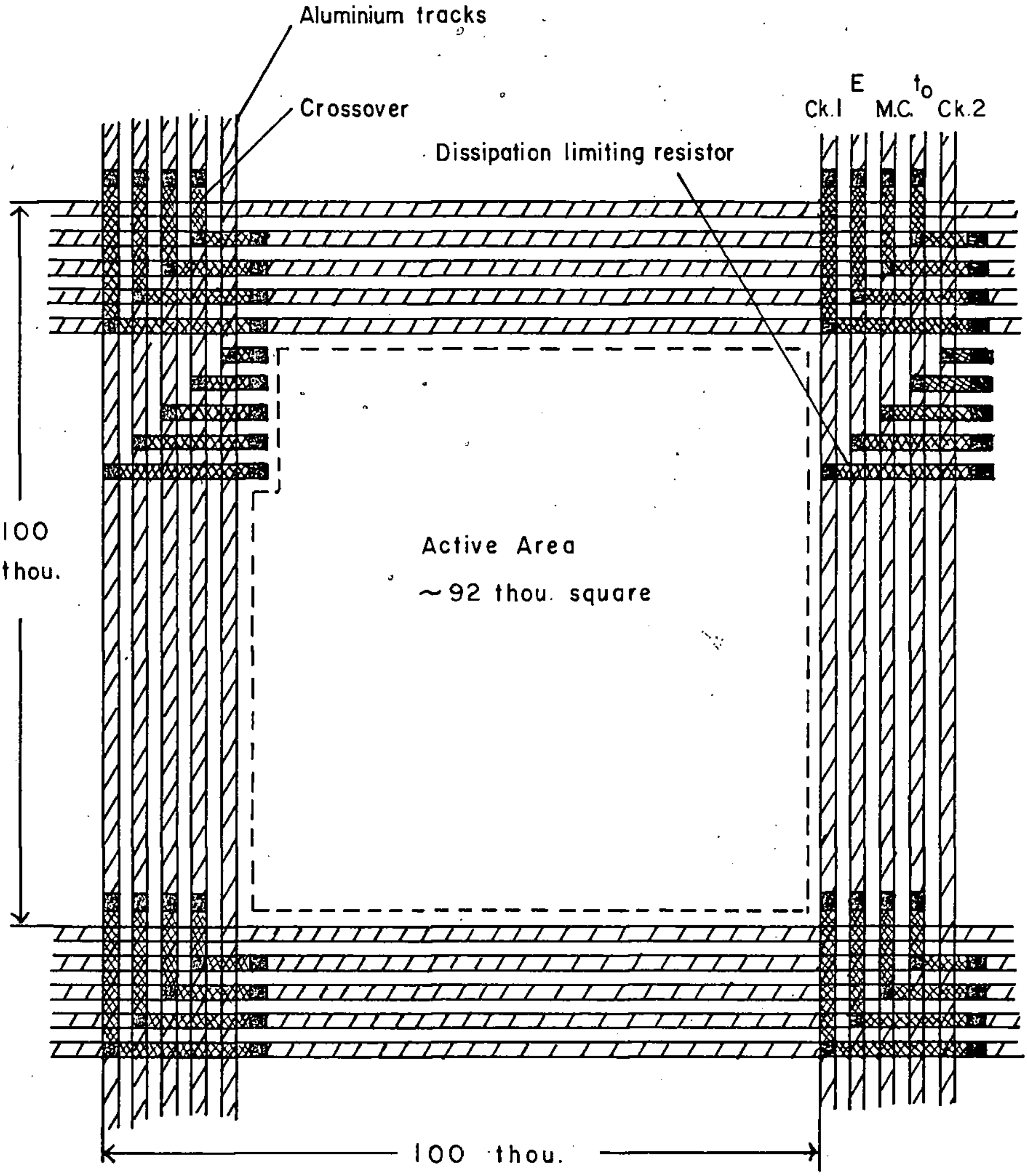

Wafer with array of chips.
Edge contacts.
Polyimide film with metallisation (not shown) from wafer bonds to edge contacts.
Premetallised polyimide film.
Pillar bond.
2"-3" diameter wafer in recess in pcb.
5" X 4" printed circuit board with edge connector compatible contacts.

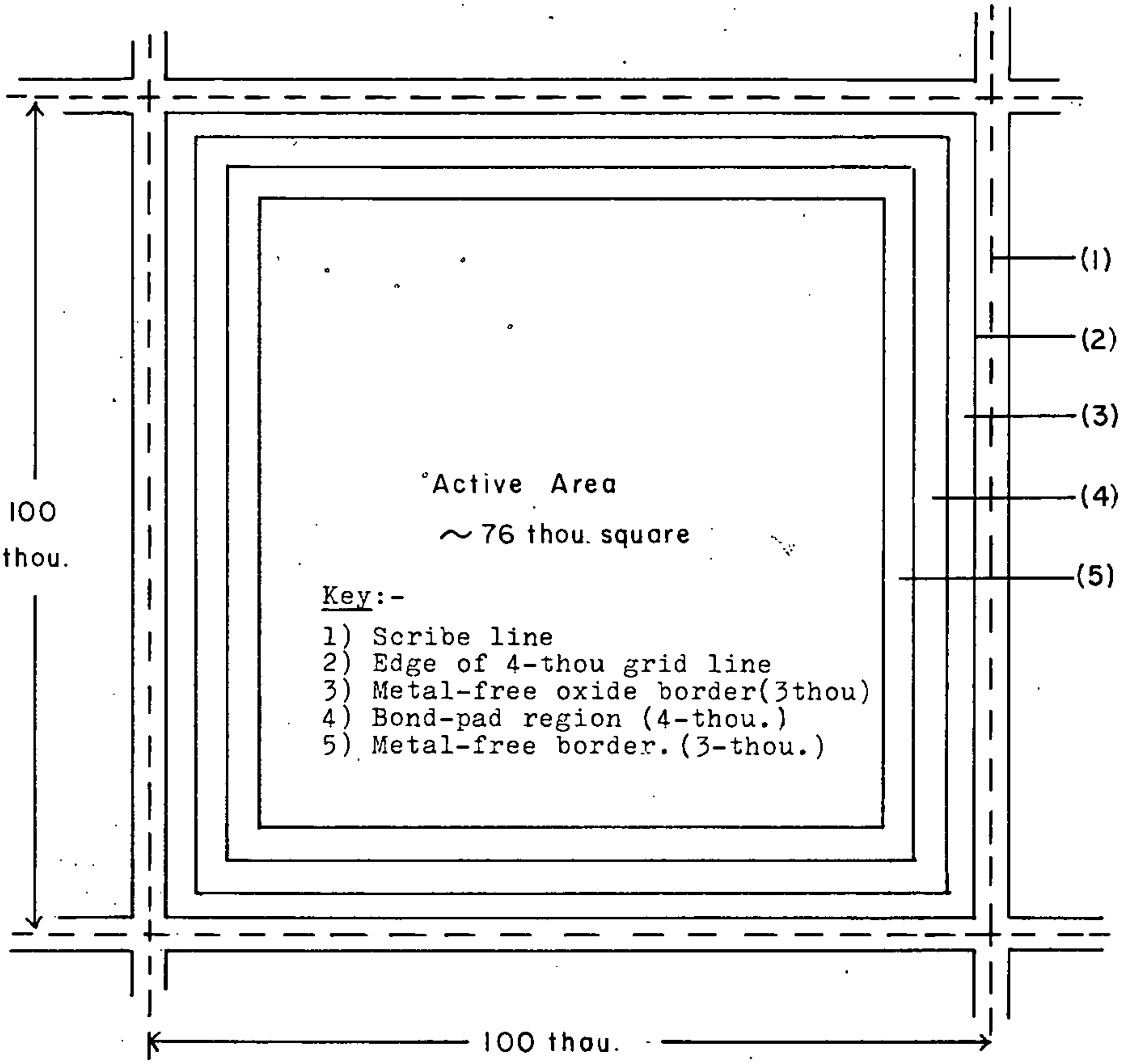
FIG. 6.10. Proposed polyimide film / pcb assembly.



NOTE:- Diagram not to scale.

Chip area = 10,000 sq. thou.
Active area = 8,464 sq. thou.
Control logic 1,600 sq. thou. (estimated for ccd).
Memory area 6,864 sq. thou.

FIG. 6.11. Residual chip area - WSI approach.


Active Area
~ 76 thou. square
Key:-
1) Scribe line
2) Edge of 4-thou grid line
3) Metal-free oxide border(3thou)
4) Bond-pad region (4-thou.)
5) Metal-free border. (3-thou.)
(1)
(2)
(3)
(4)
(5)
100 thou.
100 thou.

NOTE:- Diagram not to scale.

Chip area = 10,000 sq. thou.
Active (available) area = 5,776 sq. thou.
"Wasted" area = 4,224 sq. thou. (42.2%).

FIG. 6.12. Residual chip area - conventional assembly.

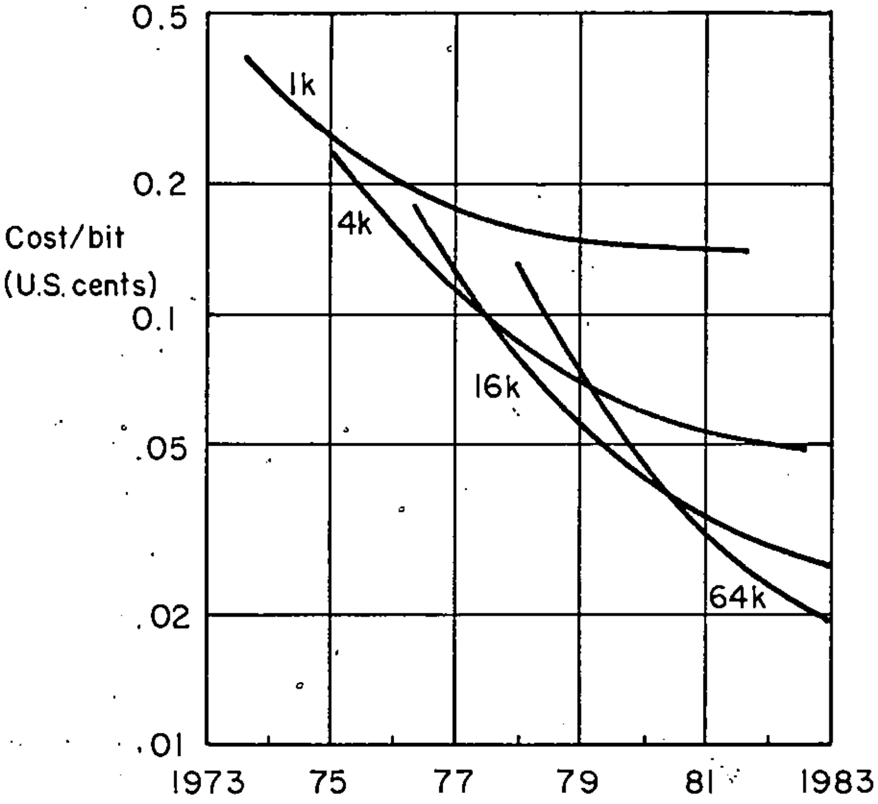


FIG. 6.13. Decline in cost per bit of semiconductor memory.

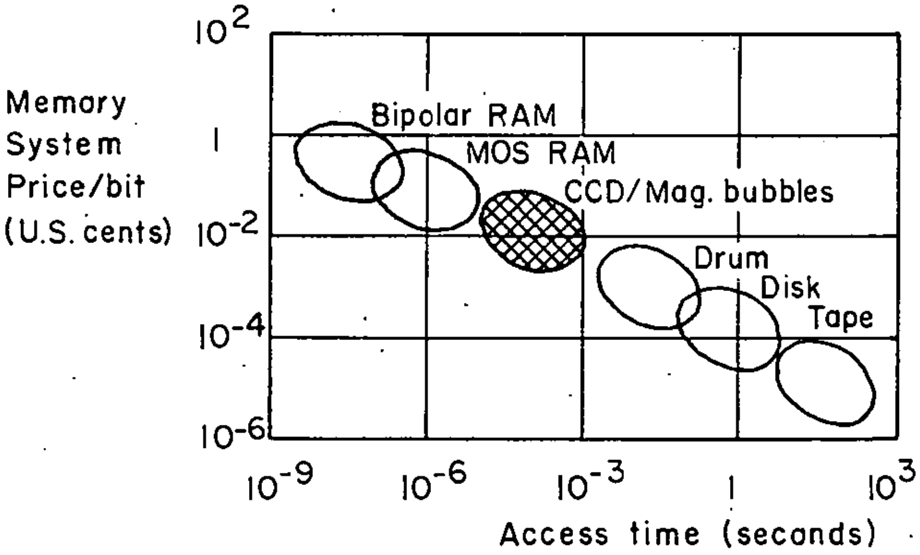


FIG. 6.14. Access time and bit price of various memory technologies.

7. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER WORK.

The technical feasibility of producing large serial memories on iterative chip arrays at reasonable yield levels has been demonstrated by computer simulation. The design of both the on-chip control logic and the external control electronics required for the serial memory structure of British Patent 1,377,859 have been proved by TTL simulation. This model has demonstrated that only 76 gates of control logic are required on each chip and has illustrated more subtle features of the design; for example the building of two parallel spirals has been achieved without any interaction between the two. It has been found to be of invaluable assistance in explaining the general concept of this approach to the industry.

The chip layout has shown that the control logic does not require an excessive amount of chip area in its layout; this non-optimised design requires an area approximately equal to that taken by 256 (optimised) shift register stages - a negligible overhead on large shift registers.

The large potential cost advantage to systems based on such structures rather than conventional semiconductor memory has also been demonstrated.

A move towards linear array distributed processing - as outlined in Section 3.4.2 - could well create a large market for these structures. Although the existing market requirements for such devices have not been fully surveyed, a substantial market sector for large serial memories is evidenced by the high current level of industry activity in the field of complex bubble memories as discussed in

Section 6.8.6. Semiconductor memories could supply much of this market if, perhaps by using the techniques described in this work, they became cost-competitive with bubble memory.

It has been shown, however, that the potential applications of W.S.I arrays may well extend far beyond the limits of the serial memory to which most of this work has been devoted. This structure should be viewed as one of a whole family of fault tolerant fixed interconnection procedures. The device may be extended to content addressable memory as described in Section 3.4.2. It has been shown that parallel/serial members of this family extend the application area to RAM, ROM and associative memory. It has further been noted that this fault-tolerant, fixed interconnection approach may well lead the way to integrated A/D converters of unprecedented ($\geq$ 15-bit) accuracy.

Further work is therefore recommended, both within the restricted area of these studies to date and within the general field of W.S.I technology. The construction of prototype W.S.I arrays - perhaps based on the chip design described in Section 4.2 - would provide test vehicles for investigation of noise and pattern sensitivity problems as well as providing more direct evidence of the technical feasibility of such devices. More thought could profitably be given to the area of power supply protection against faulty chips, as discussed in Section 6.7.1. The extension of the serial memory structure to a pseudo-core structure, as discussed in Section 3.4.2, is also worthy of further investigation. The area of device package design and assembly technology is another where major effort is required. In fact it is suggested that wafer-scale integration should be accepted as a major new technology and a design team of carefully selected and wide-ranging skills should be devoted to its investigation in both memory and wider applications.

Viewing the field of fault tolerant computer design from a semiconductor process/device design engineer's point of view it seems that there has been too little collaboration between these aspects and the more theoretical studies by logic designers in investigation of fault tolerant computing. For example, a considerable amount of thought has gone into the analysis of the effects of stuck-at-0 and stuck-at-1 faults (e.g.(9)) whereas (as noted by (12)) other faults - such as a loss of stage inversion are probably equally, if not more, likely in practice. Workers in these fields must collaborate more closely to ensure the best approach to a fault tolerant system as a whole; some problems are more suited to solution by fault tolerant semiconductor device design, others by sophisticated treatment of data - involving error detection/correction techniques.

We should remember, in striving for this ultimate goal of a totally secure computing system, the advice of Von Neumann[811]

> "There can be no question of eliminating failures or of completely paralysing the effects of failures. All we can do is to try to arrange an automaton so that, in the vast majority of failures it can continue to operate."

Only close liaison and understanding by workers in each field of the capabilities of the other can create such an optimised design.

8. REFERENCES

1. Agajanian,A.H. "A bibliography on charge-coupled devices" Solid State Technology, vol 19, no.5, pp 48-54, May 1976.

2. Arai,E. and N.Ieda, "A 64-kbit dynamic MOS RAM" ESSCIRC '77 Conference digest, pp 74-75.

3. Baker,K. "Solid-state serial memories - the role of bubbles and ccds" IEE Electronics and power, vol 24, no.9, pp. 647-652, September 1978.

4. Barsuhn,H. "Functional wafer - a new step in LSI" ESSCIRC '77 Conference digest, pp 79-80.

5. Bell,C.G. "The effect of technology on near term computer structures" Computer, vol 5, no.3, pp 29-38, March 1972.

6. Bremer,J.W. "Superchip boosts RAM size to 40-k" Electronics, vol 48, no.6, pp 30-31, March 20th, 1975.

7. Butcher,J.B. and M.Cullinan, "Nichrome fusible links for programmable integrated circuit memories" Internepcon 1977 Proceedings, pp 97-100.

8. Canning,M. et al, "Active memory calls for discretion" Electronics, vol 40, no.4, pp 143-154, February 20th, 1967.

9. Carter,W.C. and C.E.McCarthy, "Implementation of an experimental fault-tolerant memory system" IEEE Trans. Comput., vol 25, no.6, pp 557-568, June 1976.

10. Catt,I. "Improvements relating to digital integrated circuits" British Patent Specification no.1,377,859.

11. Catt,I. "Property 1A" Seminar at Imperial College, London, December 13th, 1978.

12. Dias,F.J.O. "Truth table verification of an iterative logic array" IEEE Trans Comput., vol 25, no.6, pp 605-613, June 1976.

13. Dingwall,A.G.F. "High yield processing for fixed-interconnect large-scale integrated arrays" IEEE Trans. ED. vol 15, no.9, pp 631-637, September 1968.

14. Dingwall,A. and G.B.Herzog, "High yield redundant processing for large-scale integration" IEEE NEREM Record, pp 134-135, November 1967.

15. Elmer,B.R. et al, "Fault tolerant 92,160 bit multiphase ccd memory" IEEE ISSCC Digest, pp 116-117, February 1977.

16. Frohman-Bentchkowsky,D. "A fully decoded 2048-bit electrically programmable FAMOS read only memory" IEEE Jnl. Solid State Circuits, vol 6, no.5, pp 301-306, October 1971.

17. Gledhill,R.J. "Automated IC layout" Internal communication, May 1976.

18. Goldberg,J. et al, "An organization for a highly survivable mrmory" IEEE Trans Comput., vol 23, no.7, pp 693-705, July 1974.

19. Grundy,D.L. et al, "Collector diffusion isolation isolation: the bipolar LSI" Microelectronics, vol 4, no.3, pp 10-30, Autumn, 1972.

20. Hart,K. and A.Slob, "Integrated injection logic: a new approach to LSI" IEEE Jnl Solid State Circuits, vol 7, no.5, pp 346-351, October 1972.

21. Hatt,R.J. et al, "Four-phase logic circuits using integrated m-o-s transistors" Mullard Technical Communications, no.99, pp 266-276, May 1969.

22. Hazlett,L. "Computer accelerates design and production of large arrays" Electronics, vol 40, no.4, pp 166-168, February 20th, 1967.

23. Herndon,W.H. et al, "A static 4096-bit bipolar random acces memory" IEEE Jnl. Solid State Circuits, vol 12, no.5, pp 524-527, October 1977.

24. Herring,R.B. "Silicon wafer technology - state of the art 1976" Solid State Technology, vol 19, no.5, pp 37-42, May 1976.

25. Hodges,D.A. "Microelectronic memories" Scientific American, vol 237, no.3, pp 130-145, September 1977.

26. Kraynak,P. and P.Fletcher, "Wafer-chip assembly for large-scale integration" IEEE Trans. ED. vol 15, no.9, pp 660-663, September 1968.

27. Kukreja,S.N. and I.Chen, "Combinational and sequential cellular structures" IEEE Trans. Comput., vol 22, no.9, pp 813-823, September 1973.

28. Lathrop,J.W. "Large scale integration through discretionary interconnection" IEEE NEREM Record, pp 214-215, November 1966.

29. Lea,M. "The comparative cost of associative memory" The Radio and Electronic Engineer, vol 46, no.10, pp 487-496, October 1976.

30. Lenzlinger,M. "Gate protection of MIS devices" IEEE Trans. ED., vol 18, no.4, pp 249-257, April 1971.

31. Lewin,D. "Outstanding problems in logic design" The Radio and Electronic Engineer, vol 44, no.1, pp 9-17, January 1974.

32. Lo,T.C. and M.R.Guidry, "An integrated test concept for switched-capacitor dynamic MOS RAMs" IEEE Jnl. Solid State Circuits, vol 12, no.6, pp 693-703, June 1977.

33. Manning,F.B. "An approach to highly integrated, computer-maintained cellular arrays" IEEE Trans. Comput., vol 26, no.6, pp 536-552, June 1977.

34. Manning,F.B. "Automatic test, configuration and repair of cellular arrays" PhD dissertation, Massachusetts Institute of Technology, May 1975.

35. Mo,R.S. and D.M.Gilbert, "Reliability of NiCr fusible link used in PROMs" Jnl.ECS, vol 120, no.7, pp 1001-1003, July 1973.

36. Moralee,D. "IBM reveals more details of new RAM" IEE News, January 1979.

37. Muehldorf,E.I. "Fault clustering: modrling and observation on experimental LSI chips" IEEE Jnl.Solid State Circuits, vol 10, no.4, pp 237-244, August 1975.

38. Murphy,B.T. "Cost-size optima of monolithic integrated circuits" Proc. IEEE, vol 52, no.12, pp 1537-1545, December 1964.

39. Northrup,M. "Rockwell set for 1Mbit bubble memory" IEE News, p5, December 1978.

40. Noyce,R.N. "Microelectronics" Scientific American, vol 237, no.3, pp 62-69, September 1977.

41. Petritz,R.L. "Current status of large scale integration technology" IEEE Jnl.Solid State Circuits, vol 2, no.4, pp 130-147, December 1967.

42. Preston,S.B. and R.N.Shillabeer, "Direct liquid cooling of microelectronics" INTERNEPCON Proceedings 1970, pp IX 10-IX 31.

43. Price,J.E.. "A new look at yield of integrated circuits" Proc. IEEE, vol 58, no.8, pp 1290-1291, August 1970.

44. Proebsting,R. "Dynamic MOS RAMs" New Electronics, vol 10, no.18,. pp 46-56, September 20th, 1977.

45. Sander,W.B. "Yield enhancement techniques in semiconductor memories" IEEE Jnl.Solid State Circuits, vol 7, no.4, pp 298-300, August 1972.

46. Saul,P. "A novel approach to high speed A-D conversion" Electronics Industry, pp 25-27, February 1978.

47. Scarrott,G. and S.Reddaway, "ICL's distributed array processor: world lead in number crunching" New Electronics, vol 11, no.8, pp 26-28, April 18th, 1978.

48. Schroeder,J.E. and R.L.Goslin, "A 1024-bit, fused-link CMOS PROM" IEEE ISSCC Digest, pp 190-191, February 1977.

49. Seth, S.C. "Fault testing in combinational cellular arrays" PhD dissertation, Univ. Illinois, Urbana, May 1970.

50. Shoup,R.G. "Programmable cellular logic arrays" PhD dissertation, Carnegie-Mellon Univ., Pittsburgh, PA, March 1970.

51. Smith,J.G. and H.E.Oldham, "Laser testing of integrated circuits" IEEE Jnl. Solid State Circuits, vol 12, no.3, pp 247-252, June 1977.

52. Stein,K.U. "Noise-induced error rate as limiting factor for energy per operation in digital IC's" IEEE Jnl. Solid State Circuits, vol 12, no.5, pp 527-530, October 1977.

53. Sutherland,I.E. and C.A.Mead, "Microelectronics and computer science" Scientific American, vol 237, no.3, pp 210-228, September 1977.

54. Tammaru,E. and J.B.Angell, "Redundancy for LSI yield enhancement" IEEE Jnl. Solid State Circuits, vol 2, no.4, pp 172-182, December 1967.

55. Timko,M. and P.Holloway, A fast 12-bit A/D converter on two chips" New Electronics, vol 12, no.3, pp 30-35, February 6th, 1979.

56. Toombs,D. "An update: CCD and bubble memories" IEEE Spectrum, vol 15, no.4, pp 22-30, April 1978.

57. USAF Contract AF 33(615)-3491 - Radio Corporation of America, 1966-1967.

58. USAF Contract AF 33(615)-3546 - Texas Instruments Incorporated, 1966-1967.

59. USAF Contract AF 33(615)-3620 - Philco-Microelectronics Division, 1966-1967.

60. Vaccaro,J. "Semiconductor reliability within the Department of Defense" Proc. IEEE vol 62, no.2, pp169-184, February 1974.

61. Waddell,J.M. "Achieving quality and reliability in memories" New Electronics, vol 12, no.5, pp 30-40, March 6th 1979.

62. Wallmark,J.T. "Fundamental physical limitations in integrated circuits" ESSDERC Conference, September 1974. (Reprinted in abbreviated version in Electronic Engineering, pp 52-55, February 1975.)

63. Wegener,H.A.R. et al, "The variable threshold transistor, a new electronically alterable, non-destructive read only storage device" International Electron Devices Meeting Digest, October 1967.

64. Wilcock,J.D. "Semiconductor memory review" New Electronics, vol 12, no.3, pp 70-76, February 6th, 1979.

9. BIBLIOGRAPHY

B1. Carr, W.N. and J.P.Mize, "MOS/LSI Design and Application" McGraw-Hill, 1972.

B2. Encyclopaedia Brittanica.

B3. Glaser, A.B. and G.E.Subak-Sharpe, "Integrated Circuit Engineering" Addison-Wesley, 1977.

B4. Hnatek, E.R. "A Users' Handbook of Integrated Circuits" Wiley Interscience, 1973.

B5. Hodges, D.A. "Semiconductor Memories" IEEE Press, 1972.

B6. Jowett, C.F. "Materials in Electronics" Business Books Limited, London, 1971.

B7. Lewin, D. "Theory and Design of Digital Computers" Nelson, London, 1972.

B8. Mavor, J. "MOST Integrated Circuit Engineering" IEE, 1973.

B9. Oughton, F. "Value Analysis and Value Engineering" Pitman, London, 1968.

B10. Stern, L. "Fundamentals of Integrated Circuits" Hayden Book Co. Inc., 1968

B11. Von Neumann, J. "Theory of Self-Reproducing Automata" Univ. of Illinois Press, Urbana and London, 1966.

B12. Wolfendale, E. "MOS Integrated Circuit Design" Butterworth and Co., London, 1973.

10. PUBLICATIONS

This section lists, in chronological order, the major publications by the author. Also included are reprints of the two papers (P5, P6) of greatest relevance to this present work.

P1. Aubusson, R.C. and J.W.Richer, British Patent No.1,278,414 (A compound transistor structure to reduce the chip area occupied by ECL input transistor gating.)

P2. Aubusson, R.C. "Gate protection of MIS devices" Fourth Annual Conference on Solid State Devices, Exeter University, 15-18th September, 1970.

P3. Aubusson, R.C. "The design and fabrication of a 2GHz power transistor" Colloquium on Microwave Devices, Imperial College, London, December 13th, 1972.

P4. Aubusson, R.C. and I.Catt, "Wafer-scale integration: a new approach" European Solid State Circuits Conference, Ulm University, September 20-22nd, 1977. (ESSCIRC '77 Conference Digest pp 76-78.)

P5. Aubusson, R.C. and I.Catt, "Wafer-scale integration - a fault-tolerant procedure" IEEE Jnl. Solid State Circuits, vol 13, no.3, pp 339-344, June 1978.

P6. Aubusson, R.C. and R.J.Gledhill, "Wafer-scale integration - some approaches to the interconnection problem" Microelectronics Journal, vol 9, no.1, pp 5-10, September 1978.

P7. Aubusson, R.C. "The CAM Project" Invited paper at ACTP advisory meeting, National Physical Laboratory, Teddington, June 16th, 1978.

P8. Aubusson, R.C. "Wafer-scale integration of semiconductor memories" Seminar at Brunel University, Uxbridge, November 22nd, 1978.

Wafer-Scale Integration—A Fault-Tolerant Procedure

RUSSELL C. AUBUSSON AND IVOR CATT

Abstract—This paper considers a new approach to full-slice technology in relation to existing procedures for achieving this goal. Under external control a chain of good chips is created to form a long serial memory from an array of identical chips on a full slice. Bad chips are automatically bypassed without requiring any pre- or post-programming of the metallization and without any prior knowledge of the distribution of faulty chips on the wafer. Computer simulations of chain formation are described which demonstrate the feasibility of creating such serial memories at practicable dice-yield levels.

The proposed logic design is summarized and its verification by TTL simulation is noted. The inherent fault and failure tolerance of the design are discussed and the potential problem areas of short-circuit chips, double-level metallization, spiral branching, thermal dissipation, and noise/pattern sensitivity are described together with suggested solutions. The current status of our 4ϕ dynamic MOS design is noted. Technology independence of the concept is stressed and the implications of the radical changes to the relative chip and assembly costs of wafer-scale integrated (WSI) structures are summarized.

I. Introduction

MAJOR COST reductions and improvements in reliability and packing density have always occurred each time the number of external circuit connections between components has been reduced. The advent of small-scale integrated (SSI) circuit technology and the development of this through medium-scale integrated (MSI) and large-scale integrated (LSI) to current state of the art very large-scale integrated (VLSI) devices have each made substantial contributions to these important areas of cost, size, and reliability.

The further potential cost reduction and improvements in packing density and reliability of full-slice technology have long been known. To quote from [1], R. Petritz stated in 1967:

> We at Texas Instruments feel that the full potential of semiconductor technology for integrated electronics will be realized only when the entire semiconductor slice constitutes the packaged product.

This claim for the potential advantages of a full-slice technology, made at the height of the investigations into discretionary wiring, remains equally true today. Developments in wafer-processing technology during the ensuing ten years, and the continuing high costs of chip assembly have only served to increase the attractiveness of a full-slice technology.

Discretionary wiring, however, proved not to be an economically viable way of achieving this aim. The basic idea of interconnecting the required number of chips on the wafer into a complex array using a second level, custom designed metallization pattern suffered from several disadvantages in its implementation. Firstly, each wafer required a tailor-made mask, the layout of which could only be specified after probe testing the full wafer; this added a substantial overhead to the processed slice cost. Secondly the requirement to test the wafer before completion of processing is undesirable as it must introduce some degree of contamination and damage to the first aluminum layer in addition to requiring each chip to carry a set of bonding pads (for probe test purposes) which will be redundant in the final discretionary-wired array. The major problem, however, was the implicit assumption that the processing of the second level, full wafer metallization would be 100 percent perfect and that no chips tested as good would fail during the additional processing schedules. The problems of ensuring that all vias make adequate contact while no crossovers short to the underlying metal are well known to the industry. A vast amount of research effort was expended on discretionary wiring prior to 1969 when the technique appears to have been virtually abandoned.

It is essential that the general concept of full-slice technology should not also be branded as unworkable as a result of this disastrous venture into discretionary wiring; the concept of the full wafer as the packaged unit was then, and still is, a very worthwhile goal.

Three other techniques for full-slice technology are listed in relation to discretionary wiring in Fig. 1. We see that another programmed interconnection technique has recently been investigated (2); this is based on the selective blowing of fusible links to incorporate shift-register stages or eliminate them from the array. This post-programming of the final interconnection pattern is certainly preferable to the pre-programming procedure of discretionary wiring in that it permits further levels of surgery on the wafer if the first attempt fails to achieve the desired result. It still suffers from the drawback, however, that each wafer must be treated individually in conjunction with elaborate and expensive testing procedures.

We believe that to be economically viable and acceptable in a production environment, a full-slice technology must postpone the decision on which chips are to be connected into the array until all wafer processing is complete. It should not require a detailed knowledge of the actual fault distribution on any slice so that all wafers may be treated identically throughout the processing. This suggests a fixed-interconnection procedure and

Manuscript received November 2, 1977; revised January 30, 1978. This work was supported by the Advanced Computer Technology Project, Department of Industry, UK.

R. C. Aubusson is with Middlesex Polytechnic, Queensway, Enfield, Middlesex, UK.

I. Catt is with C. A. M. Ltd., St. Albans, Hertfordshire, UK.

0018-9200/78/0600-0339$00.75 © 1978 IEEE

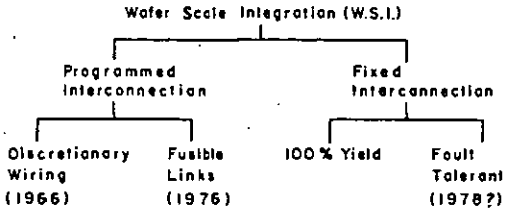

Wafer Scale Integration (W.S.I.)
Programmed Interconnection
Fixed Interconnection
Discretionary Wiring (1966)
Fusible Links (1976)
100 % Yield
Fault Tolerant (1978?)

Fig. 1. WSI techniques.

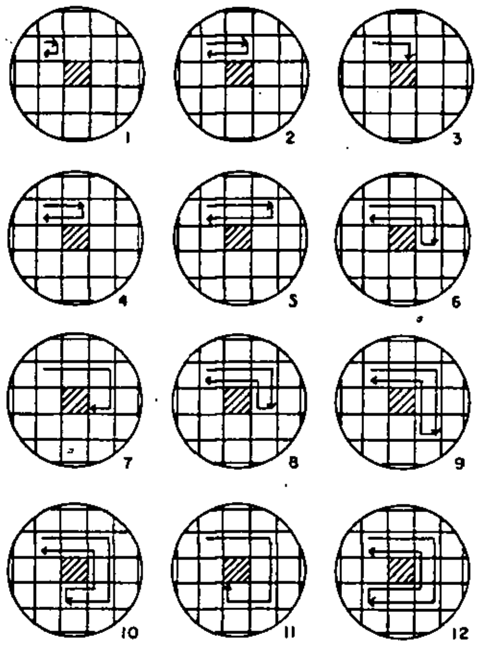


Fig. 2. Twelve stages in the early formation of a spiral.

two approaches are listed in Fig. 1. The 100 percent yield situation must, however, be regarded (for the present at least) as a trivial and unattainable special case, listed only for the sake of completeness. As a fixed interconnection approach will of necessity include faulty chips in the full-wafer metallization pattern, the chosen procedure must be fault tolerant.

II. Design Philosophy

We now describe the fault-tolerant fixed interconnection procedure first described in [4], which has been under investigation at Middlesex Polytechnic since early 1975.

The essential feature of this design is the capability of linking together the good chips on a wafer without requiring any additional mask or even a prior knowledge of which chips on the wafer are good and which are faulty. To achieve this each chip is given the capability of addressing, under external control, its four nearest neighbors. Connections are made to the input and output of one chip on the wafer and to the power supply/clock grids supplying all chips on the wafer.

We will now discuss the creation of a spiral in relation to Fig. 2. Let us suppose that the chip is a 1K bit shift register. A known bit pattern of 1K bits is fed into the chip and the output pattern is compared with the input to ensure that the memory is functioning correctly. This chip is then instructed to access the adjacent chip due East and 2K bits of data are fed into this two-chip serial memory. If the data are returned uncorrupted then the second chip is also known to be good and is instructed to access the chip to the South. 3K bits of data are fed in and we will now suppose that an error is detected in the returned data, thus indicating that the third chip is faulty. The second chip is now instructed to close its links with the existing third chip and to access the chip to the East, which now becomes the third chip in the chain. In this way it is possible to build up long chains of good chips, thus producing very large serial memories.

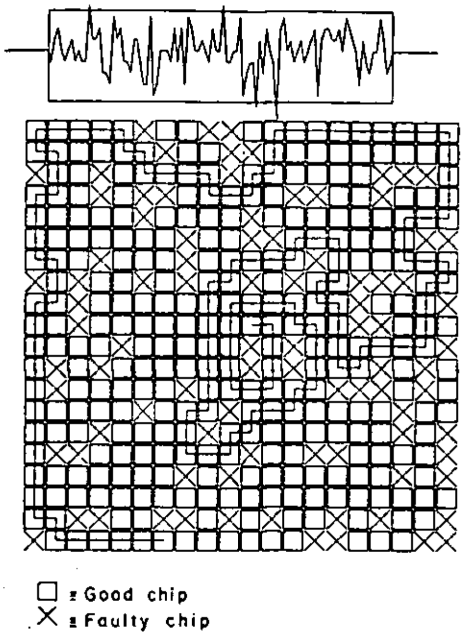


Fig. 3. Computer simulation of square array with spiral.

When we presented this structure at ESSCIRC 1977 [3] we were unaware of the parallel work undertaken at the Massachusetts Institute of Technology, Laboratory of Computer Science. This is described in a paper by F. B. Manning [5]. In relation to this paper we, too, have considered "tree" and "grid" algorithms although most of our work to date has been directed towards the "arm" algorithm (called "spiral" in our terminology), and our paper relates exclusively to this type of array.

To predict the length of the chain of chips which is possible for a given dice yield on the array, a computer program was developed [6] to investigate the probability of generating a 128 chip spiral on a 20 X 20 chip array at yields in the range 60-90 percent. A sample computer-generated plot is shown in Fig. 3.

A random array of good and bad chips is generated to a target yield by assigning to each site on the 20 X 20 matrix a random number in the range 00-99. If our target yield is, say, 73 percent, then numbers 00-72 represent good chips and 73-99 represent bad ones. A check is kept on the frequency distribution of the numbers so generated to ensure that no detectable order is built into the array.

We may expect, on average, that four numbers will occur either less than once (i.e., not at all) or greater than 8 times

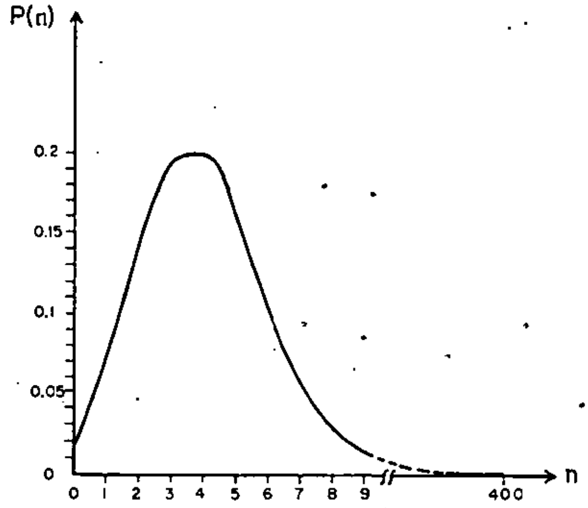


Fig. 4. Probability of any particular number in range 00–99 occurring *n* times from 400 random selections.

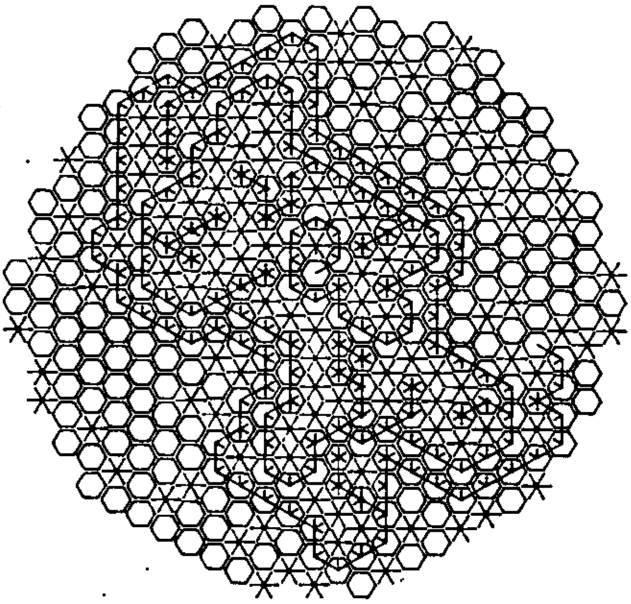


Fig. 5. Computer simulation of hexagonal array with spiral.

when selecting 400 numbers at random from the range 00-99, the general form of the distribution for each number being as shown in Fig. 4. The frequency of occurrence of each number is plotted at the top of the diagram, the horizontal lines representing frequencies of one-half and $8\frac{1}{2}$ so that we expect; on average, four peaks or troughs outside these limits. It is accepted that this is not a rigorous test of randomness of the array so plotted but the distribution of bad chips on actual slices—which the array represents—is not truly random either. The clustering of defective chips due to large process induced faults (e.g., photoresist tears), and towards the edge region of actual slices, will assist the propagation of the spiral. If we discount the possibility of very large area defects (e.g., tweezer scratch across slice) our random array therefore provides a pessimistic indication of the performance of our spiral algorithm on actual slices.

Having generated the required 20 X 20 pseudorandom array the program then attempts to generate a spiral of good chips on the array, bypassing faulty ones and backtracking out of blind alleys as necessary until either the required 128-chip spiral has been generated or the spiral has failed to reach the 128-chip target in the preset time limit. The maximum and final chain lengths of unsuccessful spirals are also noted.

It will be observed that the spiral starts near the center of the array; if the 11, 11 point is a bad chip the program permits the starting point to wander on the array until a good chip is found. If this chip happens to be walled in by faulty ones, however, the run will terminate. The spiral progresses in a clockwise direction until it reaches the array edge when, because it is always attempting to turn clockwise (and hence off the array), it proceeds in a retrograde (anticlockwise) direction, hugging the array edge. Note that the computer has no prior knowledge of either the fault distribution on the array or the position of the spiral in relation to the array boundaries. Every new chip is added only after testing for a faulty chip and an array boundary.

We note that Manning's spiral [5] is launched at a corner of the array. This has the advantage that the spiral cannot become trapped between itself and the array edge; however, this is not generally a serious problem unless grossly nonsquare arrays are considered. We believe the better course is to launch the spiral well away from the slice edge where it will be less impeded during its critical early stages of growth.

It is desirable to reduce the required slice yield for a given chain length if this can be done without substantially increasing the on-chip control logic complexity and hence chip size. The hexagonal array, shown in two forms in [4], permits the new chip to access five others rather than the three others of the conventional square array. Both these structures are fully compatible with current step and repeat procedure (without interlacing) if the array is stepped as a block of two staggered chips. Any chip shape is acceptable in wafer-scale integration as the slice will not be scribed.

We have modified the computer program to fit the hexagonal array. Fig. 5 shows such a computer generated plot. In this format the good chips are represented by hexagons and the bad ones by six-pointed stars. The spiral is launched near the center of the array and proceeds in an anticlockwise direction, once again bypassing faulty chips, backing out of blind alleys and turning retrograde at the array edge. On this plot we also indicate, as short spurs on the spiral leading from the center of each good chip, directions which have been accessed and rejected either because a chip was faulty or it had already been included in the spiral or it was outside the array boundary.

Several hundred such spirals have been simulated for both the square and hexagonal arrays.

Fig. 6 indicates the proportion of spirals which reach the 128-chip target for a range of dice yields. This plot assumes that the input chip is good. Each data point represents a minimum of 20 spirals and the error bounds are for 95 percent confidence levels. The best smooth curve fit to the data is also shown. These curves show the marked reduction in the necessary yield for the hexagonal array compared with the square

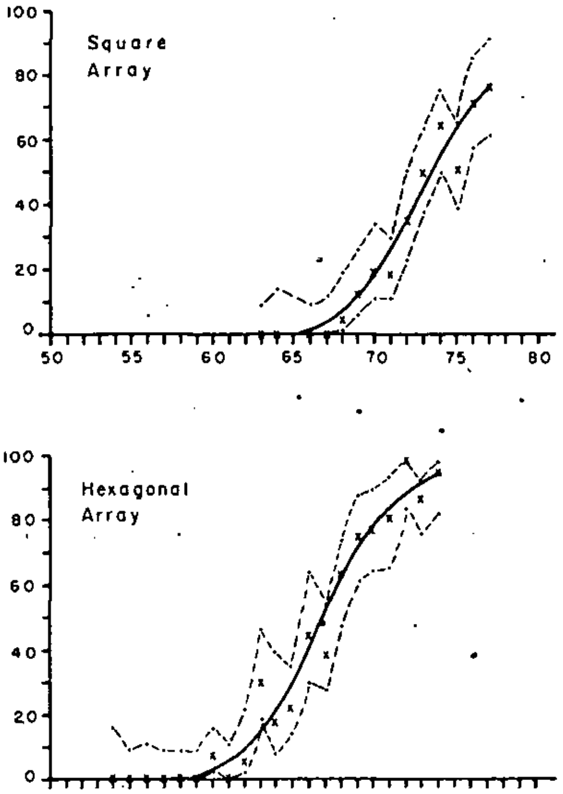


Fig. 6. Percentage of spirals reaching target of 128 chips on array of ~ 400 chips (vertical) versus dice yield (percentage) for square and hexagonal arrays.

array. Some of this advantage will, however, be offset by the increased logical complexity and hence chip size of the 6-way algorithm.

Considerably more powerful algorithms for spiral generation have been proposed and are currently under investigation; however, sufficient results are not yet available to warrant their inclusion in this paper.

III. Logic Design

The logic required to achieve the recognition and implementation of external commands during the setting up of the spiral is fully described in [4]. The current design embodies some improvements on this early scheme but these are in details rather than basic philosophy and will not be discussed here. We will, however, now summarize the broad concepts of the logic design.

All chips in the array are connected to a master grid or grids; in our 4ϕ dynamic p-channel MOS design these will provide signals from which ϕ_1, ϕ_2, ϕ_3, ϕ_4, and two timing waveforms—designated t_o and Master Clear—may be derived. In the absence of a spiral, all good chips on the wafer will be "looking" in the same direction at any given time; that is, their interchip address register will cause only one of the four input gates—one each on the North, East, South and West edges of the chip—to be enabled to receive data at that instant. Similarly all good chips will be "open" to one direction; that is, their address registers will allow only one of the four output gates—one each on the N, E, S, and W edges of the chip—to be enabled to output data at that instant. OPEN and LOOK directions on all chips will sequence through the cycle N, E, S, W under the control of t_o.

In order to feed data into the input/output chip it is essential that this chip is "looking" in the direction from which the data stream will appear. The Master Clear pulse is arranged to reset both OPEN and LOOK address registers on all chips to "OPEN N" and "LOOK N" so that the OPEN/LOOK direction of all free-running chips on the wafer is known at any time from the number of t_o pulses since the last Master Clear.

All chips on the wafer are identical except for a special interface chip which carries bond pads, protection networks (if necessary), and discrete process-control test structures if desired. This is desirable to eliminate the requirement for bonding pads on any of the actual array chips. It is sited to the North of the first chip in the spiral (chip A) as this is designed to receive its input from the North, and so the interface chip will not reduce the choice of three downstream directions (E, S, W) available to chip A.

The principle of the design is to use part of the shift-register storage element itself to detect and decode commands by tapping off certain key bits near the beginning of the shift register. Only two commands are required in our implementation. The first, which we call "FREEZE," causes the chip addressed to cease following the cyclic N, E, S, W directions of all free-running chips and remain looking in the current direction until either power failure or a further Master Clear pulse. This will cause it to lock onto the upstream chip. The second command, "STEP 90," causes the OPEN address register to disable the current OPEN direction (which will also have been frozen during the previous FREEZE command) and access the next one in N, E, S, W cycle. Both commands are represented by an all zeros word (to ensure that no spurious "1" bits are present in the shift register) followed by a "1" then a "0." FREEZE contains only zeros in the rest of this new word while STEP 90 contains an extra "1" at a key point in the word. The appearance of this "1" at a predetermined point in the shift register during a command word triggers the STEP 90 routine.

The command must, in general, pass down the spiral through other chips to reach the intended chip at the end of the spiral; a further requirement of the on-chip control logic is therefore to ensure that only the intended chip implements the STEP 90 command. To achieve this, part of the command word is allocated to a field which contains a number equal to the number of chips currently in the spiral up to the chip to be addressed. Each chip subtracts one from this number so that by the time the word reaches the addressed chip it has been reduced to zero. The last chip attempts to subtract one from zero, causing an unrequited borrow to propagate up this field. The control logic detects this unrequited borrow and enables a gate which would otherwise inhibit the execution of the STEP 90 command.

In our design the above control is achieved using 76 gates—a negligible overhead on, say, a 5K bit shift register. As might be expected, however, the routing of the signals around the chip requires a substantial area of silicon that is considerably larger than the random logic gates themselves. However the area occupied by the control logic and associated metallization fits easily into the space which is wasted on conventional chips by the scribe channel, oxide-free border, bond pads, and ancillary components which are not required for WSI arrays.

IV. Design Verification

Both the on-chip control logic and the external control electronics required to govern the spiral formation have been simulated for the square array using TTL. A rack was built to represent a matrix of 4 X 4 chip sites and ten sets of printed circuit boards representing good chips may by plugged into any ten of the 16-chip sites in a pseudorandom array. Bad chips have so far been represented simply be empty chip sites; it is intended to construct additional "bad" chips to simulate specific faults which may occur in the on-chip control logic. Eight light-emitting diodes at each chip site on the rack indicate to which and from which direction each chip is sending and receiving data; a further three LED's indicate the state of other key points on the logic. A purpose-built clock generator supplies the required waveforms (also indicated by LED's) to both on-chip and off-chip control logic so the only external supply required is a 5 V dc input.

This rack has demonstrated the building of short spirals, bypassing faulty chips and backtracking out of blind alleys, thus proving the logic design for both on-chip and external-control electronics.

V. Fault and Failure Tolerance

Fault tolerance occurs at many levels in computer systems but we will consider only those aspects directly relevant to WSI arrays. Minute imperfections will always be present in integrated-circuit chips; these only become critical if they prevent the correct operation of the device or constitute a major reliability hazard. Discretionary wiring and fusible link techniques can, as we have seen, create an apparently perfect array on a flawed wafer; however, any additional major defects occurring in these hardwired arrays will be catastrophic.

The very nature of this "soft-wiring" of interconnections between chips on our appproach endows the device not only with the attribute of fault tolerance but also of failure tolerance. If a particular chip develops a fault the spiral retracts back to a single chip and regrows, bypassing the new faulty chip to create a fresh sprial. This self-repairing feature, called graceful degradation, is not possible with any of the other approaches to wafer-scale integration so far considered.

VI. Potential Problem Areas

We have considered several aspects of the implementation of such WSI arrays in which we feel problems new to the technology may arise. Some of the more potentially serious problem areas are now briefly discussed.

1) *The Short Circuit or Leaky Chip:* The grid supplying power to all chips on the wafer must be protected from short circuit or excessively leaky chips. The best way of achieving this would be to isolate such chips from the supply, for example, by the use of fusible links. For our initial investigations we have chosen merely to limit the current which may be drawn by any chip in the array using a diffused series resistor between the power/clock lines on each chip and the full-wafer supply grid. This presupposes that no such resistor will have a serious defect within a few mils of the grid contact but we do not consider this to be a major limitation.

2) *Double-Level Metallization:* While double-level metallization is practicable on pilot lines it has generally been regarded as an unsuitable technique for quantity production. We accept that considerable space could be saved in the device layout if double-level aluminum were available, but no insuperable topological problems arise from a restriction to one level of metal. If more than one supply grid is required most circuit technologies will tolerate a small additional resistance arising from diffused crossunders in at least one of the grids.

3) *Spiral Branching:* If a chip has a particular type of fault in its OPEN address logic it may attempt to access two adjacent downstream chips rather than only one. This, if successful, will lead to a branching of the spiral and two parallel spirals will attempt to grow downstream of the faulty chip. This, at best, will lead to rapid wastage of storage capacity on the wafer. A similar fault on the LOOK address logic will cause the chip to attempt to receive data from two directions simultaneously, resulting in corruption or total loss of data. The nature of the design, however, is such that at least two adjacent chips must have very similar faults for this defect to occur as the spurious LOOK or OPEN direction on the faulty chip will only see a dormant OPEN or LOOK port on the adjacent (good) chip.

This built-in safeguard against the transmission of data across such rogue interfaces is believed to be adequate as the probability of two adjacent chips containing similar control logic faults must be small.

4) *Thermal Dissipation:* This could become a problem with very high speed memories. In addition to the working memory, all chips on the array are dissipating standby power while the faulty chips may well be dissipating several times the normal power level. Power wastage occurs both with fusible links and with current limiting resistors, although it could perhaps be avoided by feeding the power supply around the wafer with the developing spiral so that only chips involved in the actual spiral are powered up.

A full wafer should be able to dissipate several watts without trouble using natural or forced air convection. To dissipate substantially higher powers it becomes necessary to hold the back of the wafer in intimate contact with a good heat sink. At still higher power levels (hundreds of watts) phase-change cooling could be implemented although we would then be reaching the level at which the feeding of power *into* the wafer becomes the problem rather than the dissipation of the heat produced. With MOS technology, however, it is possible to limit the total wafer power dissipation to the order of a few watts.

5) *Noise and Pattern Sensitivity:* The close proximity of several million shift register stages switching in synchronism on a full wafer must be expected to create a substantial noise problem. The structure has a built-in tolerance in that during the setting up of the spiral any noise-induced corruption of data will be interpreted as a faulty chip. This will cause the

spiral to bypass good chips, if necessary, in order to reduce the noise problem to an acceptable level by creating a more open array. However, as the possibility of pattern sensitivity cannot readily be checked by the use of pseudorandom data sequences and voltage and time margining of the clock waveforms supplied to the wafer, some thought has been given to the design of special test structures to assess its importance.

VII. Conclusions and Recommendations

To suit our in-house p-channel MOS process line at Middlesex Polytechnic we have selected 4-phase dynamic MOS technology (with overlapping clocks). The device has initially been designed as a conventional chip so that it may be more readily evaluated. Masks have been generated for this design and some completed wafers are now being evaluated. Once the circuit design has been proved the layout will be stripped of bonding pads, gate protection networks, and test structures, and will be restepped as a true WSI array.

It must be emphasized that this approach does not depend on any particular technology; it should not, for example, be considered to be in competition with CCD technology. In fact it is suggested that a serial shift register of greater than 1 Mbit complexity could be manufactured on a single wafer using standard CCD processing facilities already available in the industry.

In contemplating WSI arrays the industry must be prepared to rethink the tradeoffs in chip size, dice yield, testing complexity, and assembly areas. Whereas it has become economically justifiable to go for the ultimate in chip size to the point where dice yields are vanishingly small because chip cost is a negligible fraction of final device cost, WSI arrays demand smaller chips (to increase dice yield to workable levels of ~ 70 percent). The question is now "How much useful store can I get on a 3 in wafer?" and the tradeoff is between the areas to be allocated to on-chip control logic and chip storage. Reduced chip size gives increased dice yield and greater utilization of the array, but at the expense of an increase in on-chip control-logic overhead.

As described in this paper the structure is suitable only for serial memories. Compared with existing techniques for long serial semiconductor memory fabrication, we believe our approach to offer many advantages. However the addition of a fast data line, as described in [4], creates a medium speed memory which compares favorably with core in terms of performance while offering a considerable cost advantage.

We believe that the structure described is the first in a family of fault tolerant, fixed interconnection WSI devices. Work will proceed at Middlesex Polytechnic on further development of algorithms and structures for such arrays but we believe that the device is already useful to the industry in the form described

Acknowledgment

We would like to thank our colleagues both within the Polytechnic and in the industry who have contributed to this project.

References

[1] R. L. Petritz, "Current status of LSI technology," *IEEE J. Solid State-Circuits*, vol. SC-2, pp. 130-147, Dec. 1967.

[2] B. R. Elmer *et al.*, "Fault tolerant 92160 bit multiphase CCD memory," in *IEEE Int. Solid-State Cir. Conf. Dig.*, Feb. 1977, pp. 116-117.

[3] R. C. Aubusson and I. Catt, "Wafer scale integration: A new approach" in *3rd European Solid-State Cir. Conf. Dig.*, Sept. 1977, pp. 76-78.

[4] I. Catt, British Patent Specification 1 377 859, Dec. 1974.

[5] F. B. Manning, "An approach to highly integrated, computer-maintained cellular arrays," *IEEE Trans. Comput.*, vol. C-26, pp. 536-552, June 1977.

[6] R. J. Gledhill, Middlesex Polytechnic, private communication.

Russell C. Aubusson was born in Liverpool, England, in November 1942. He received the B.Sc. degree in physics from Manchester University, Manchester, England, in 1964 and is currently working towards the Ph.D. degree at Middlesex Polytechnic, Middlesex, UK.

He joined the Semiconductor Division of Ferranti Ltd. to work on RTL, DTL, ECL, and MNOS integrated-circuit development. In 1971 he was appointed to head the company's R & D group on RF transistors, and in 1973 was made responsible for liaison with the Ministry of Defence on all Ferranti Electronic Component Division's defence contracts. He joined Middlesex Polytechnic as a Research Fellow in 1975 to take charge of this memory project and was appointed as a Senior Lecturer in Microelectronics in September 1977.

Mr. Aubusson is a member of the Institution of Electrical Engineers.

Ivor Catt was born in Plymouth, England, on December 19, 1935. He received the M.A. degree in engineering from Trinity College, Cambridge University, Cambridge, England, in 1959.

From 1959 to 1962 he worked in the Ferranti (now I.C.L.) Computer Labs, Manchester, England. After working in computer peripheral companies in Los Angeles, he went to Motorola Semiconductor Products Division, Phoenix, AZ, where he pioneered the interconnection of 1-ns logic gates. After a period of being in charge of the LSI Research Group at Sperry Semiconductor, Norwalk, CT, he returned to England in 1968 and worked in R & D at Computer Technology Ltd., Hemel Hempstead. More recently he has worked in R & D on radar systems at G.E.C. and telex exchange at International Telephone and Telegraph. He is presently with C.A.M. Ltd., St. Albans, Hertfordshire, UK. He has published two books: *Computer Worship* (London: Pitman, 1974), and *The Catt Concept* (New York: Putnam, 1971). Currently, three research projects funded by the British Government to develop his computer inventions are in progress or about to start. He is consulting on these, and also is writing a textbook on digital design.

Wafer scale integration - some approaches to the interconnection problem

by R. C. Aubusson and R. J. Gledhill †

This paper describes computer simulations of various algorithms applied to the problem of interconnecting individual good chips on an imperfect wafer. A fault-tolerant, fixed interconnection scheme is employed and algorithms are described for structures ranging from a simple rectangular block of chips through a hexagonal array to a pattern having eight nearest neighbours. Simulations of finite but unbounded arrays are also considered.

1. Introduction

Integrated circuit technologies normally construct an array of similar devices on a wafer of silicon. The geometry of the array of chips is constrained by the requirement of scribability to be a rectangular grid. The yield of such devices depends on the chosen technology and on the size of the individual devices typically as illustrated in Fig. 1. Additional losses are imposed during scribing and assembly operations.

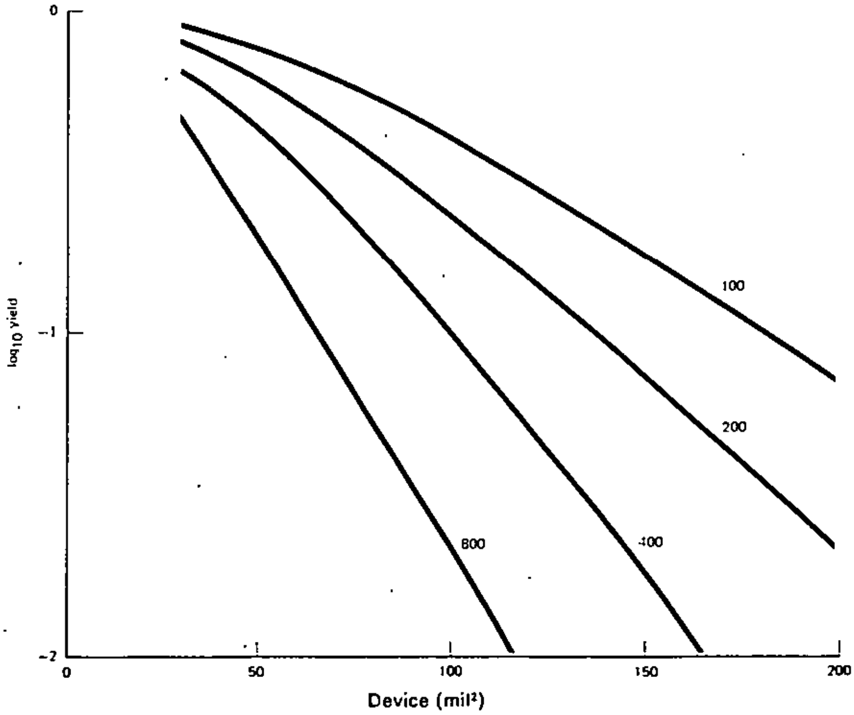


Fig. 1 Yield vs device area parametrised by defect density (inches⁻²).

The high cost of assembly and testing of the individual chips has caused manufacturers to increase chip complexities to the point where device yields are vanishingly small in their attempt to reduce the cost per gate to the lowest possible figure. The current practicable limit on chip size for bipolar technologies is probably about 6mm square.

Wafer-Scale Integration (WSI) is an alternative technology which may be applied to certain types of integrated circuit – for example the interconnection of shift

† Middlesex Polytechnic, Queensway, Enfield, UK.

MICROELECTRONICS, Vol. 9 No. 1 © 1978 Mackintosh Publications Ltd., Luton.

register chips to create very large memories. In WSI the individual chips are interconnected on the wafer into a large array, thus eliminating the scribing and chip assembly operations and thereby drastically reducing assembly costs and improving reliability. The techniques for WSI have been compared in refs. 1 and 2 and are summarised in Fig. 2.

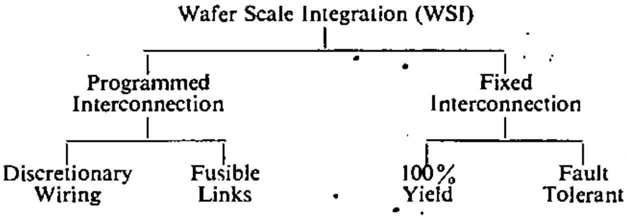


Fig. 2 Techniques for WSI.

Attempts to eliminate the scribe, dice and encapsulation processing steps have in the past concentrated on either avoiding faulty devices by discretionary wiring[3] (requiring a customised metal interconnection for each wafer) or the use of fusible links[4] to isolate faulty devices.

The fault tolerant, fixed interconnection procedures for full slice integration described in references [1, 2, 5, 6] retain the advantage of fixed metal interconnection by including control logic on each device to enable selection of a signal path avoiding faulty devices.

Since such a path may be constituted at any time by external control logic this approach embodies both fault tolerance and graceful degradation – as faults manifest themselves the faulty devices may be avoided by reconstituting a new path using only functional devices.

The main features which characterise these alternative approaches are the spatial configuration of devices which may input signals to or receive output signals from a given device, called its neighbours, and the sequence of selected input and output connections as an assembly of devices is configured, called the algorithm.

2. Algorithms studied

The schemes discussed in this paper are illustrated in Fig. 3.

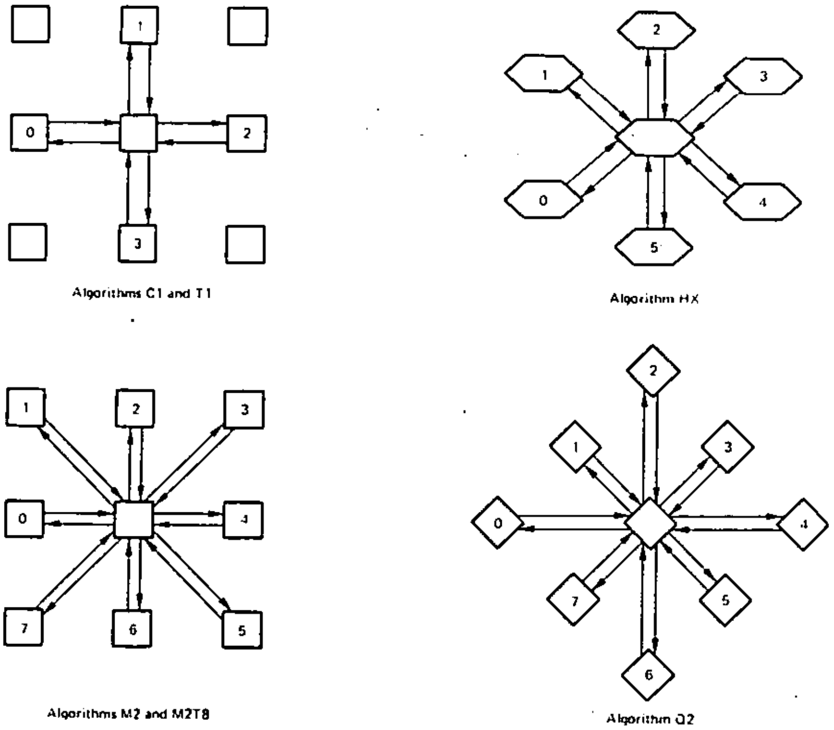
Algorithms discussed here were all designed to assemble devices into a spiral avoiding faulty devices as illustrated in Fig. 4; the sequence of neighbours to be tested is as indicated in Fig. 3, the central device being accessed from the device labelled zero.

Although the performance of an algorithm may be improved by permitting non-adjacent chips to be accessed (e.g. Knight's move in chess) to extricate the spiral from blind alleys, it is desirable to reduce interconnection crossover problems by requiring a device's neighbours to be physically adjacent to that device.

In assessing the relative performance of different algorithms one may either set an arbitrary target length for the spiral to achieve or, alternatively, require the algorithm to produce the longest possible spiral length on the array. It is probable that practical applications of such memory devices will require a fixed length for the spiral and, with this point in view, our earlier results[2] related to a fixed (arbitrary) spiral length of 128 chips. The results quoted in this paper, however, relate to arrays where the spirals have been permitted to extend to their maximum possible length.

A schematic of the algorithm requirement is given in Fig. 5.

Criteria of practical importance in assessing algorithm performance include the area of device required to implement the on-chip control logic, the variation of spiral length as a function of device yield, the yield required to achieve particular target lengths reliably and the time



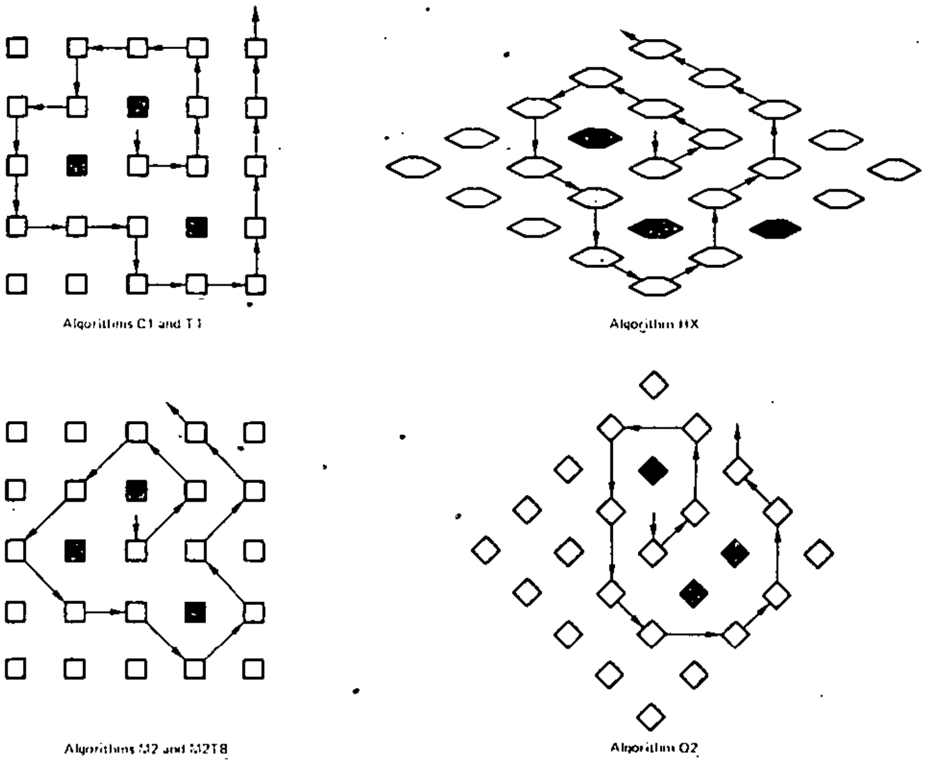
Algorithms C1 and T1

Algorithm HX

Algorithms M2 and M2T8

Algorithm Q2

Fig. 3 The sequence in which neighbours of the central device are tested. Data are received from the device marked O.



Algorithms C1 and T1

Algorithm HX

Algorithms M2 and M2T8

Algorithm Q2

Fig. 4 The growth of a spiral avoiding faulty devices (shown solid).

required to carry out the test and spiral generation.

3. Results

The first criterion – the silicon area required to implement the on-chip control logic – is obviously highly dependent on the technology chosen for device fabrication. As an example, a non-optimised layout of the on-chip control logic required to implement algorithms Cl and Tl in p-channel four-phase dynamic MOS technology occupied an area roughly equivalent to 256 shift register stages – less than the area occupied by the bonding pads used on these test devices. (Bonding pads and scribe channels are not required on each chip in WSI technology.)

The remaining criteria were evaluated by computer simulation of 20 × 20 arrays of devices for algorithms Cl, Tl, M2 and M2T8, for a nearly circular array of 392 devices for algorithm HX and 421 devices for the Q2 array.

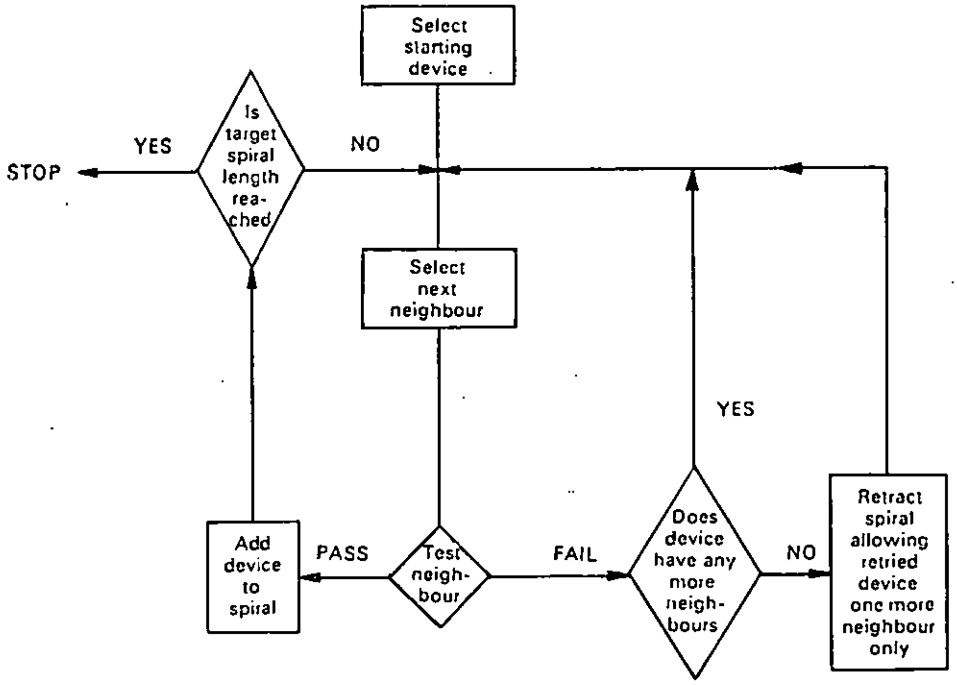

Select starting device
Is target spiral length reached
YES
STOP
NO
Select next neighbour
YES
Add device to spiral
PASS
Test neighbour
FAIL
Does device have any more neighbours
NO
Retract spiral allowing retried device one more neighbour only

Fig. 5 Schematic of the algorithm requirement.

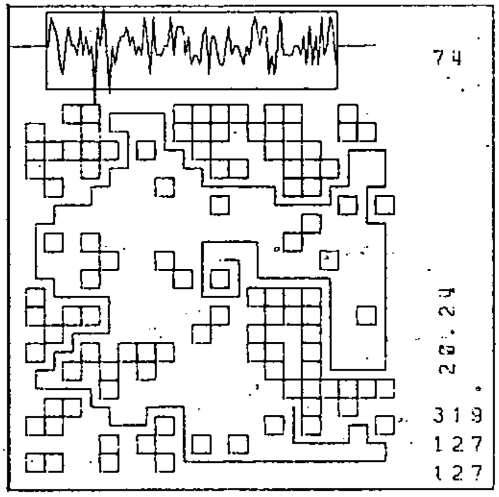


Fig. 6 Simulation of the C1 algorithm.

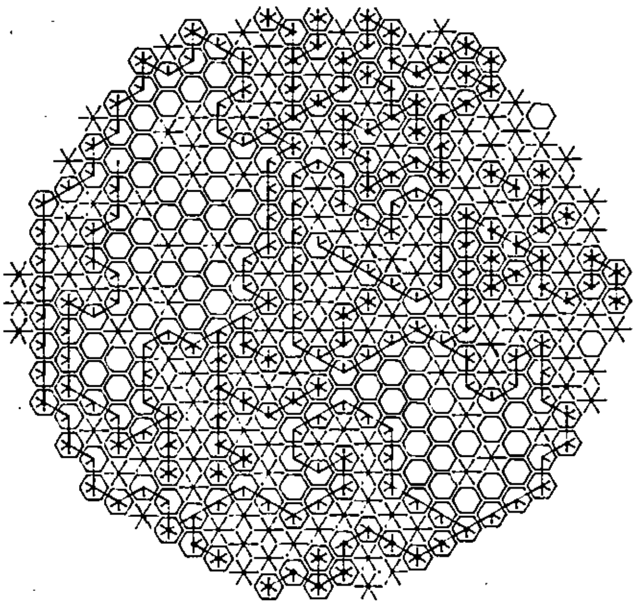


Fig. 7 Simulation of the HX algorithm.

For each algorithm, six thousand simulations of slices with pseudorandomly distributed faulty devices were carried out at yields ranging from 40 per cent to 100 per cent. Graphical output was used to verify that each algorithm was being simulated correctly; this is reproduced in Figs. 6 and 7 for the C1 and HX algorithms.

A fairly severe restriction on the maximum attainable spiral length is imposed when the growing spiral encounters an array edge, many functional devices being rejected as the spiral retracts to try new neighbours. To avoid this constraint the toroidal algorithms T1 and M2T8 were proposed and simulated; here the signal paths for corresponding devices on the north, south, east and west edges of the array were interconnected to provide an effectively unbounded but finite array of devices.

The maximum spiral length attained was recorded for each simulation and these data are summarised in Figs. 8 to 10. The time taken to assemble a spiral of given target length is given in Fig. 11 where the time is in multiples of the time required to test a single device.

4. Conclusions

These results demonstrate that long spirals of good chips can be assembled at practicable dice yield levels. A comparison of the HX and T1 array results suggests that the elimination of the array boundary – achieved with the

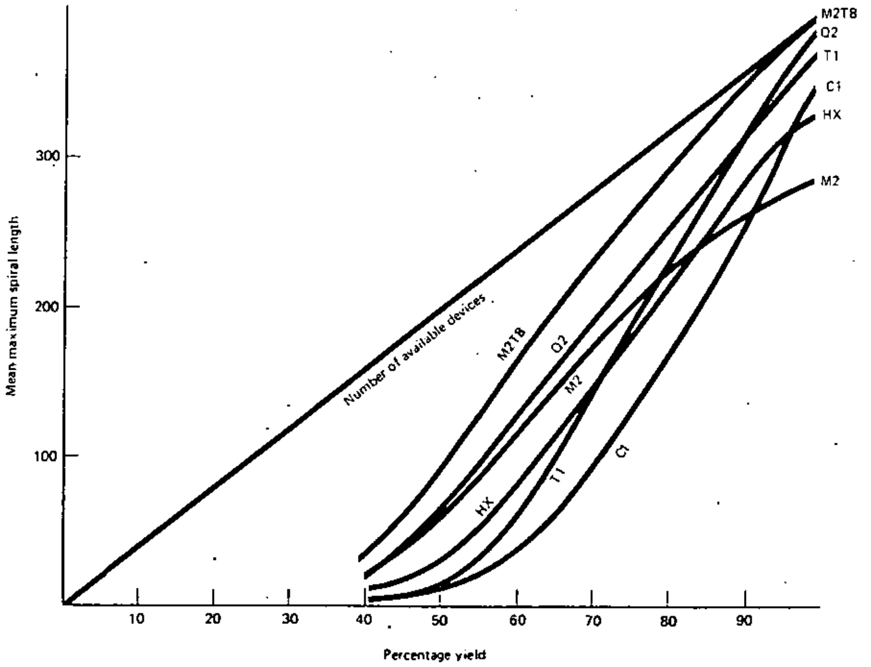


Fig. 8 Mean maximum spiral length vs % yield.

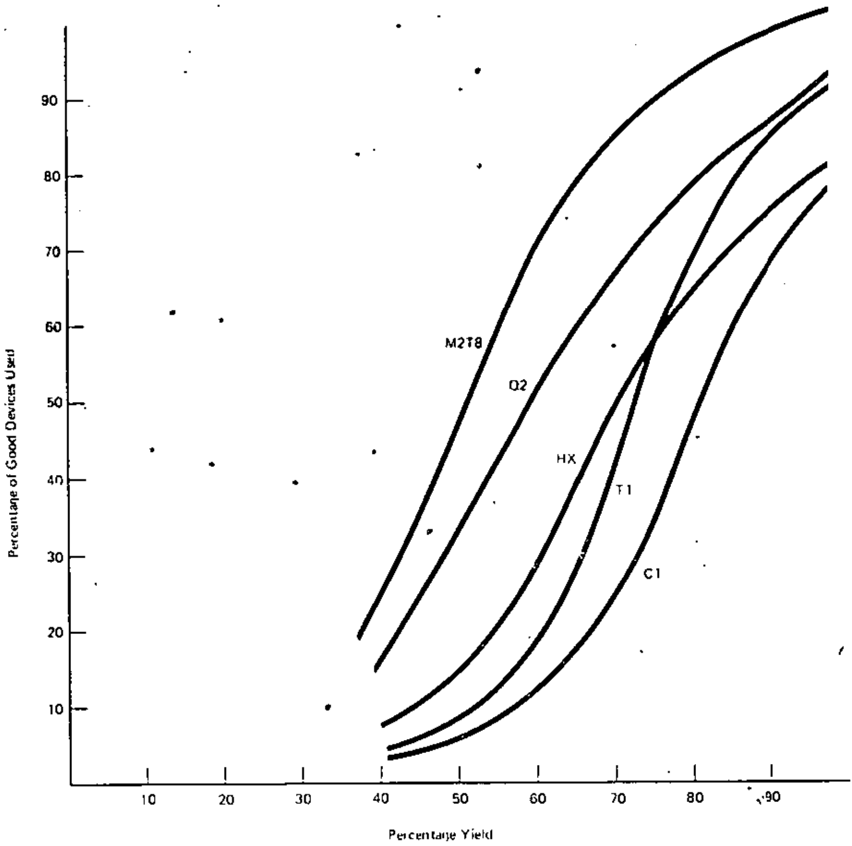


Fig. 9 Percentage usage vs yield.

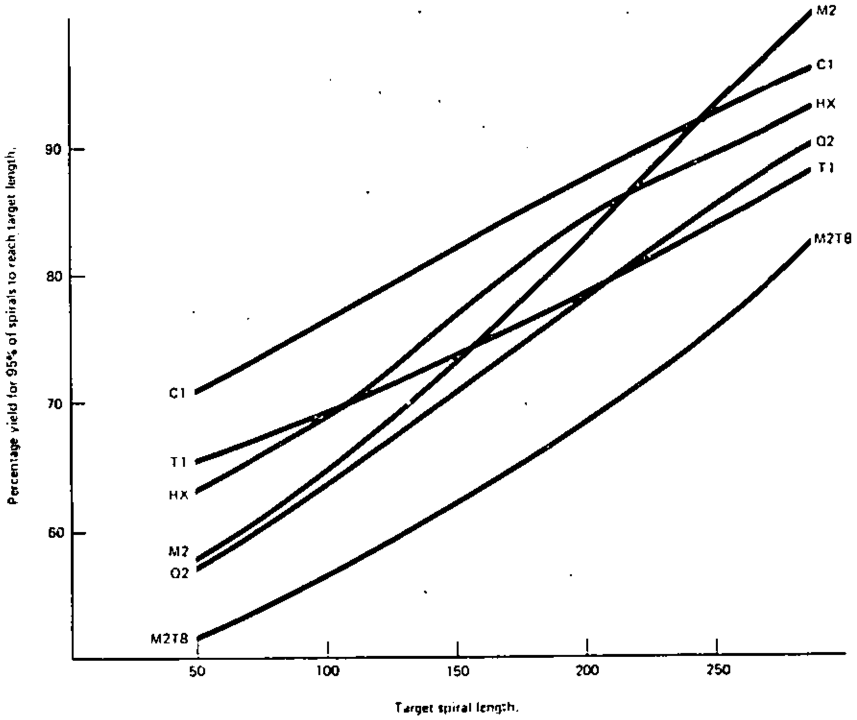


Fig. 10 Effect of variation of target length.

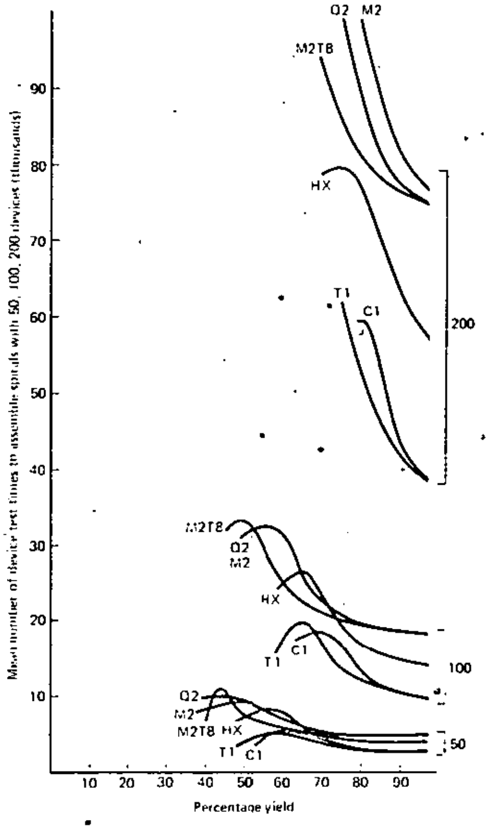


Fig. 11 Speed of spiral formation.

toroidal configuration – is equivalent to an increase of two nearest neighbours, suggesting that on-chip control logic complexity may be reduced at the expense of increased metallisation. However, even the basic four-way algorithm (C1) is considered adequate in performance for chip sizes up to ~ 140 mils square with dice yields currently available in the industry.

5. References

[1] Aubusson, R. C. and Catt, I., "Wafer-Scale Integration: A New Approach", 3rd European Solid-State Circuits Conference Digest, p..76-78, Sept. 1977.

[2] Aubusson, R. C. and Catt, I., "Wafer-Scale Integration – A Fault-Tolerant Procedure", *IEEE Journal of Solid-State Circuits*, Vol. SC-13, pp.339-344, June 1978.

[3] Petritz, R. L., "Current Status of LSI Technology", *IEEE Journal of Solid-State Circuits*, Vol. SC-2, pp.130-147, Dec. 1967.

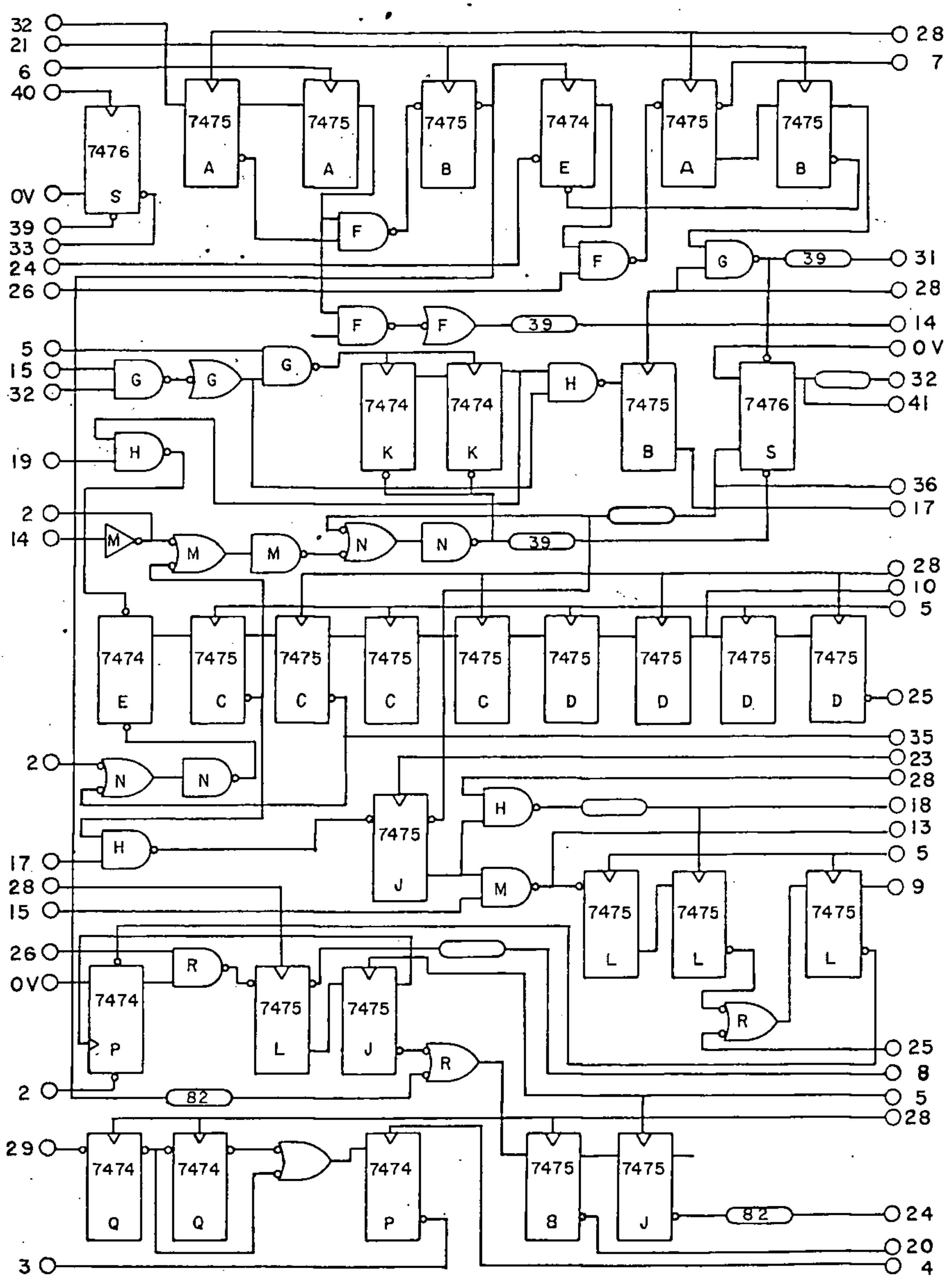
[4] Elmer, B. R., *et al*, "Fault Tolerant 92160 bit multiphase CCD Memory", in IEEE Int. Solid-State Circuits Conference Digest, pp.116-117, Feb. 1977.

[5] Catt, I., British Patent Specification 1,377,859, December 1974.

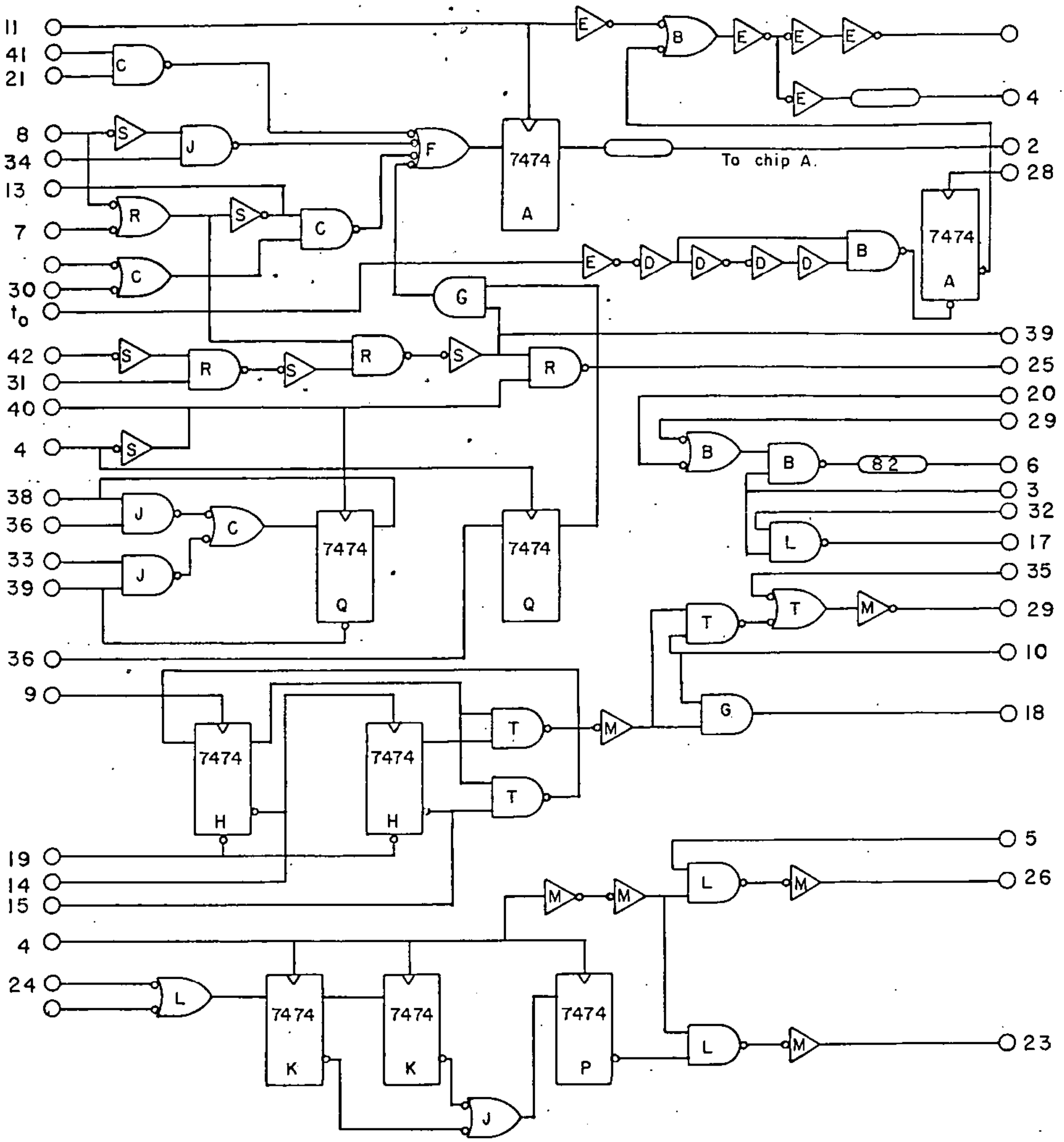
[6] Manning, F. B., "An Approach to Highly Integrated, Computer-Maintained Cellular Arrays", *IEEE Trans. Computers*. Vol. C-26, pp.536-552, June 1977.

6. Acknowledgments

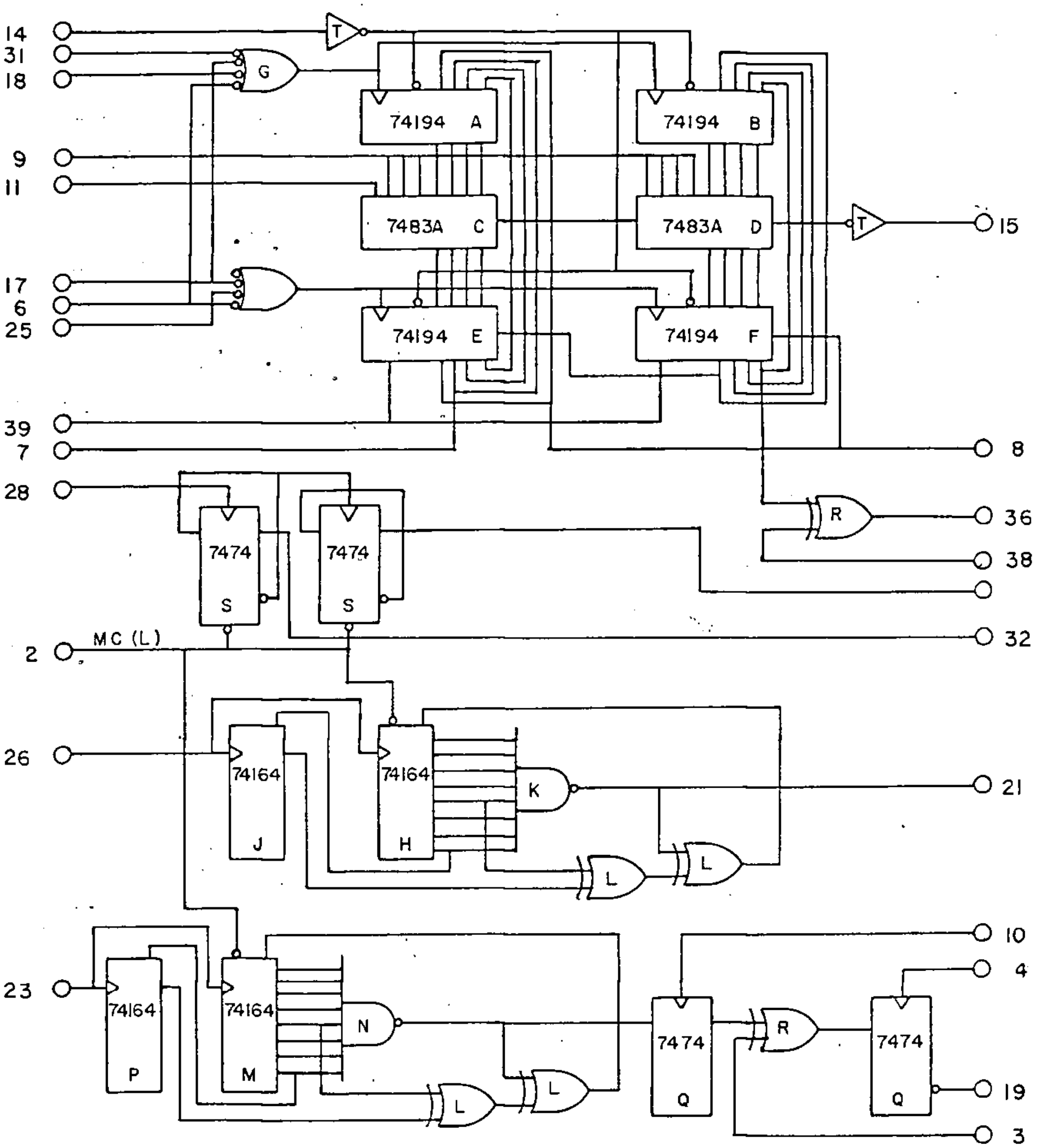
This work was supported by ACTP, Department of Industry, to whom permission to publish this paper is gratefully acknowledged.



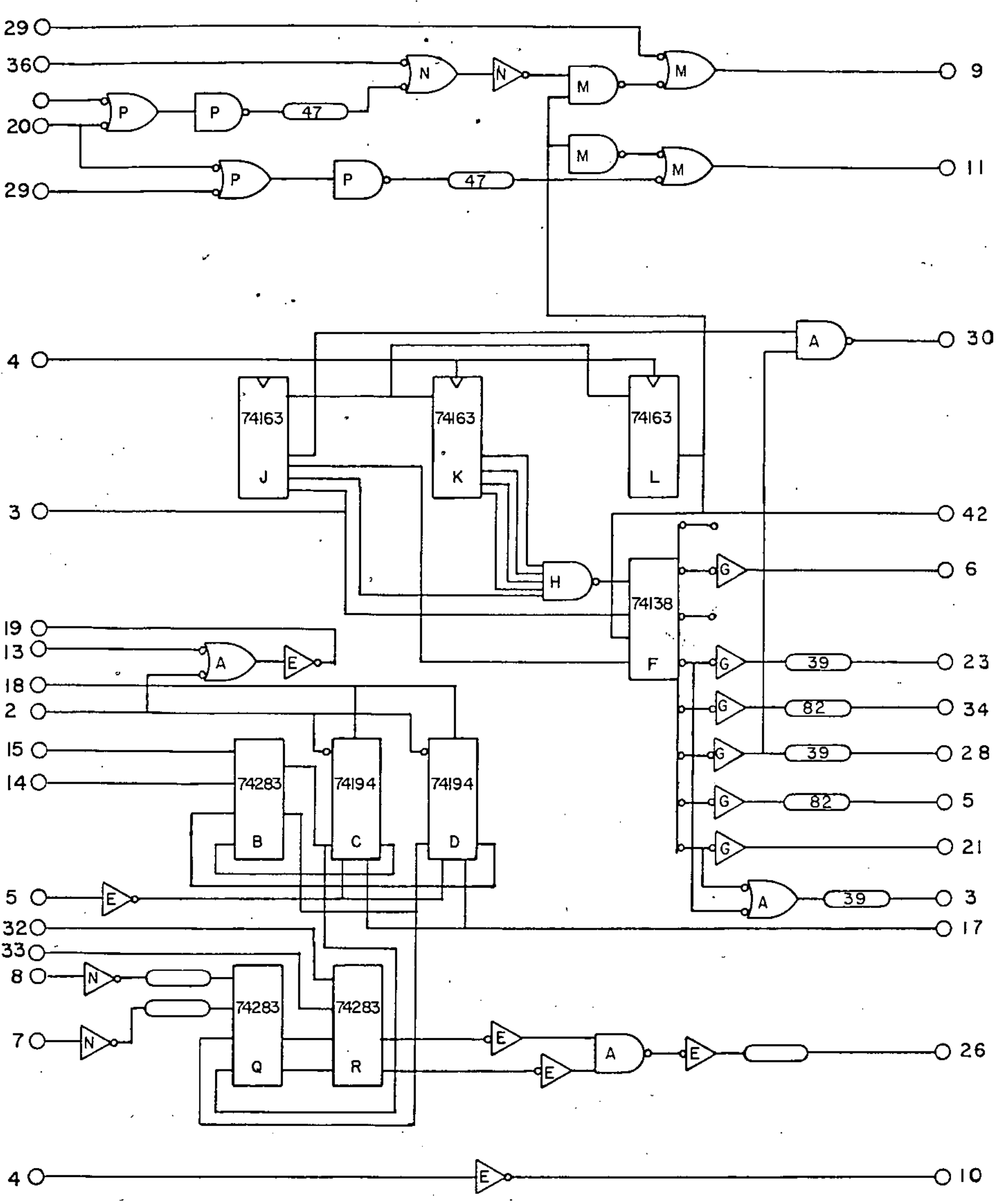
External control electronics (Chip Z)- Board I.



External control electronics (Chip Z) - Board 2.



External control electronics (Chip Z) - Board 3.



External control electronics (Chip Z) - Board 4.

INTER-BOARD CONNECTIONS FOR CHIP-Z.

Pin No.	Card Number			
	1	2	3	4
2	3,4/2	To chip A	1,4/2	1,3/2
3	2/5	4/3	From chip A	2/3
4	2,3,4/4	1,3,4/4	1,2,4/4	1,2,3/4
5	4/5	1/3		1/5
6	4/6	3/6	2/6	1/6
7	2/7	1/7	4/7	3/7
8	2/8	1/8	4/8	3/8
9	2/9	1/9	4/9	3/9
10	2/10	1/10	4/10	3/10
11	M.C.	$\varnothing_1$.	4/11	3/11
13	4/13	3/39		1/13
14	3/14	4/14	1/14	2/14
15	3/15	4/15	1/15	2/15
17	4/17	3/17	2/17	1/17
18	3/18	4/18	1/18	2/18
19	3/19	4/19	1/19	2/19
20	2,4/20	1,4/20		1,2/20
21	4/21	3/21	2/21	1/21
23	4/23	3/23	2/23	1/23
24	2/24	1/24		
25	2,3/25	1,3/25	1,2/25	
26	4/26	3/26	2/26	1/26
28	2,3,4/28	1,3,4/28	1,2,4/28	1,2,3/28
29	2,4/29	1,4/29		1,2/29
30		4/30,t_o(L)		2/30
31	3/31	4/31	1/31	2/31
32	2/32	1/32	4/32	3/32
33	2/33	1/33	4/33	3/33
34		4/34		2/34
35	2/35	1/35		
36	4/36	3/36	2/36	1/36
38		3/38	2/38	
39	2/39	1/39	2/13	
40	2/40	1/40		
41	2/41	1/41		
42		4/42		2/42

Pins not detailed in the preceding table are used for H.T. and Earth rails (except for pin 37 which is recessed to accommodate the polarisation key and is therefore not available for external connection).

The device types not identified on the drawings I.1-I.4 are as follows:-

Type SN7400.

Board 1.	F,G,H,M,N,R.
Board 2.	B,C,J,L,R,T.
Board 4.	A,M,N,P.

Type SN7404.

Board 2.	M,S.
Board 4.	E,G.

Type SN7408.

Board 2.	G.

Type SN7414.

Board 2.	D,E.
Board 3.	T.

Type SN7420.

Board 2.	F.
Board 3.	G.

Type SN7430.

Board 3.	K,N.
Board 4.	H.

Type SN7486.

Board 3.	L,R.

APPENDIX II

Computer Program to simulate spiral generation in square arrays.

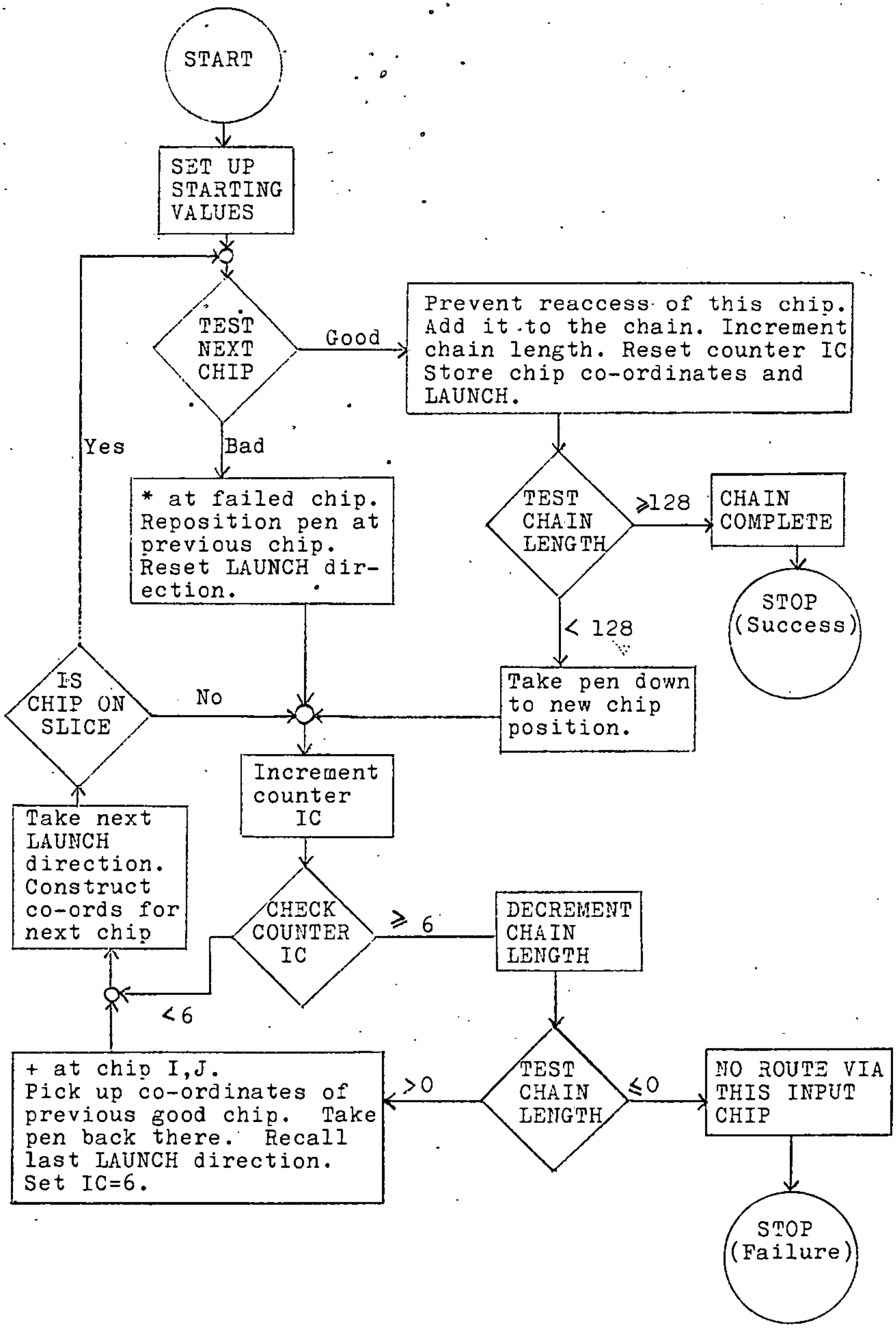
```
      DIMENSION KK(26),INEWS(128),IA(20,20),JF(101),JP(101),
      JX(128)
      DIMENSION JY(128),II(4),JJ(4)
      DATA II,JJ/0,1,0,-1,1,0,-1,0/
      DATA JP,JF/75*1,127*0/
      CALL PLOTS(2)
      R=2.182818
      YIELD=75.
      NCHIPS=4
      LX=-501
      LX=0
      LY=1503
      GOTO2
    1 CALL SYMBOL(LX,0,.1,KK,90.,78)
    2 READ(2,3)KK
    3 FORMAT(26A3)
      LX=LX+25
      IF(KK-3H$   )1,4,1
    4 LX=LX-501
  300 CONTINUE
      RR=R
      DO 301 I=1,101
      JF(I)=0
  301 CONTINUE
      LY=LY+501
      IF(LY-2004)53,51,51
   51 LY=0
      LX=LX+501
   53 CONTINUE
      NF=0
      RR=R
      DO 102 I=1,20
      DO 100 J=1,20
      R=AMOD(R*37.,1.)
      XR=R
      XR=R*100
      IR=XR
      IR=IR+1
      IA(I,J)=JP(IR)
      NF=NF+IA(I,J)
      JF(IR)=JF(IR)+1
  100 CONTINUE
  102 CONTINUE
      WRITE(3,103)
  103 FORMAT(///)
      DO 201 I=1,20
      DO 201 J=1,20
      IX=I*19+LX+1
      IY=J*19+LY+1
      IF(IA(I,J))199,200,199
  199 CONTINUE
      CALL DRAW(IX,IY,3)
      CALL DRAW(IX+17,IY,2)
      CALL DRAW(IX+17,IY+17,2)
      CALL DRAW(IX,IY+17,2)
      CALL DRAW(IX,IY,2)
      GOTO201
  200 CONTINUE
      CALL DRAW(IX,IY,3)
      CALL DRAW(IX+17,IY+17,2)
      CALL DRAW(IX,IY+17,3)
      CALL DRAW(IX+17,IY,2)
  201 CONTINUE
      KY=LY+460
      KY45=KY-45
      L341=LX+341
      LX40=LX+40
      KY35=KY+35
      CALL DRAW(LX+1,KY,3)
      CALL DRAW(LX40,KY,2)
      CALL DRAW(LX40,KY35,2)
      CALL DRAW(L341,KY35,2)
      CALL DRAW(L341,KY45,2)
      CALL DRAW(LX40,KY45,2)
      CALL DRAW(LX40,KY,2)
      DO 220 I=1,100
      IX=LX+39+3*I
      IY=LY+500-10*JF(I)
      CALL DRAW(IX,IY,2)
  220 CONTINUE
      CALL DRAW(LX+341,KY,2)
      CALL DRAW(LX+380,KY,2)
      WRITE(3,103)
      NN=0
      N=1
      I=11
      J=11
      JX(1)=I
      JY(1)=J
      DO 399 K=2,128
      JX(K)=0
      JY(K)=0
  399 CONTINUE
      IANGLE=0
      NEWS=1
      MAX=0
  400 CONTINUE
      IF(I)510,510,405
  405 IF(I-20)410,410,510
  410 IF(J)510,510,415
  415 IF(J-20)416,416,510
  510 CONTINUE
      GOTO465
  416 CONTINUE
      NN=NN+1
      IF(NN-400)419,506,506
  419 CONTINUE
  420 IF(IA(I,J)-1)465,425,465
  425 CONTINUE
      INEWS(N)=NEWS
      IA(I,J)=-1
      IF(N-127)440,440,430
  430 CONTINUE
      WRITE(3,435)
  435 FORMAT(22H STOP. SPIRAL COMPLETE )
      GOTO515
  440 CONTINUE
      IF(MAX-N)431,431,432
  431 MAX=N
  432 CONTINUE
      CONTINUE
      IANGLE=0
  450 CONTINUE
      IANGLE=IANGLE+1
  455 CONTINUE
      NEWS=NEWS+1
      N=N+1
      IF(NEWS-5)461,460,460
  460 NEWS=1
  461 CONTINUE
      I=I+II(NEWS)
      J=J+JJ(NEWS)
      JX(N)=I
      JY(N)=J
      GOTO400
  465 CONTINUE
      N=N-1
      IF(N)485,485,466
  485 CONTINUE
      IF(NN-1)491,491,515
  491 CONTINUE
      NN=0
      N=1
      I=JX(1)
      J=JY(1)
      IF(J-20)495,492,492
  492 J=0
      IF(I-20)494,493,493
  493 I=0
  494 CONTINUE
      J=J+1
  495 J=J+1
      JX(1)=I
      JY(1)=J
      GOTO400
  466 CONTINUE
      I=JX(N)
      J=JY(N)
      NEWS=NEWS-2
      IF(NEWS)468,468,469
  468 CONTINUE
      NEWS=NEWS+4
  469 CONTINUE
      IF(IANGLE-3)450,470,470
  470 CONTINUE
      IANGLE=2
      NEWS=INEWS(N)
      IF(MAX-N)800,801,801
  800 MAX=N
  801 CONTINUE
      GOTO465
  506 CONTINUE
      WRITE(3,507)
  507 FORMAT(50H STOP. EXCESSIVE TIME TAKEN IN TESTING
  515 CONTINUE
      N=N-1
      WRITE(3,802)RR,N,NN,MAX
  802 FORMAT(1X,F14.8,3(1X,I3))
      X=N
      CALL NUMBER(440+LX,10+LY,.1,X,0.,-1)
      X=MAX
      CALL NUMBER(440+LX,40+LY,.1,X,0.,-1)
      X=NN
      CALL NUMBER(440+LX,70+LY,.1,X,0.,-1)
      CALL NUMBER(460+LX,120+LY,.1,RR,90.,2)
      EN=NF
      EN=0.25*EN
      CALL NUMBER(440+LX,440+LY,.1,EN,0.,-1)
      IPEN=3
      IF(N)796,796,797
  797 CONTINUE
      DO 799 K=1,N
      I=JX(K)
      J=JY(K)
      IX=I*19+LX+10
      IY=J*19+LY+10
      CALL DRAW(IX,IY,IPEN)
      IPEN=2
  799 CONTINUE
  796 CONTINUE
      NCHIPS=NCHIPS-1
      IF(NCHIPS)106,106,300
  106 CONTINUE
      CALL PLOT(0.,0.,999)
      WRITE(3,105)R
  105 FORMAT(1X,F20.12)
      STOP
      END
```

APPENDIX III

Computer program to simulate spiral generation in hexagonal arrays.

```
      DIMENSION IA(29,29)
      DIMENSION ID(6),JD(6),NEXT(6),IX(128),JY(128),
      KL(128)
      DIMENSION LAST(6)
      DIMENSION XXC(8),YYC(3)
      DIMENSION ITIT(3)
      COMMON XC,YC
      DATA IAA/3HA  /
      DATA ITIT/3H(  ,3H / ,3H  )/
      DATA XXC/10.,40.,70.,100.,130.,160.,190.,220./
      DATA YYC/0.,25.,50./
      DATA ID,JD/1,1,0,-1,-1,0,0,1,1,0,-1,-1/
      DATA NEXT/3,4,5,6,1,2/
      DATA LAST/4,5,6,1,2,3/
      R=.31312
      YIELD=0.70
      XC=0.
      YC=0.
      H=0.1
      H=0.5
      CALL PLOTS(2)
      NC=29
      DO 101 NX=1,4
      XC=XXC(NX)
      DO 101 NY=1,3
      RR=R
      YC=YYC(NY)
      CALL NEWPEN(1)
      NG=0
      NT=0
      DO 4 J=1,NC
      DO 4 I=1,NC
      R=AMOD(R*37.,1.)
      IF(R-YIELD)2,2,3
    2 IA(I,J)=1
      NG=NG+1
      NT=NT+1
      IF(ITEST(I,J))4,4,200
  200 CONTINUE
      CALL HEX(1.,I,J,H)
      GOTO4
    3 IA(I,J)=0
      NT=NT+1
      IF(ITEST(I,J))4,4,300
  300 CONTINUE
      CALL STAR(1.,I,J,H)
    4 CONTINUE
      YIELD=NG
      DUMMY=NT
      YIELD=YIELD/DUMMY
      CALL NEWPEN(2)
      I=12
      I=NC/2
      J=I
      IPEN=3
      LAUNCH=6
      II=I
      JJ=J
      IC=0
      N=0
      NN=0
      MAX=0
   10 CONTINUE
      NN=NN+1
      IF(NN-999)43,41,41
   41 WRITE(3,42)
   42 FORMAT(8H TIME UP)
      GOTO30
   43 CONTINUE
      IF(ITEST(I,J,))100,100,40
   40 CONTINUE
      IF(IA(I,J))100,100,20
   20 CONTINUE
 CHIP I,J PASSED TEST
      IA(I,J)=-1
      II=I
      JJ=J
      IC=0
      N=N+1
      IX(N)=1
      JY(N)=J
      KL(N)=LAUNCH
      CALL CART(I,J,X,Y,H)
      CALL DRAW(X,Y,IPEN)
      IPEN=2
      IF(N-128)50,30,30
 CHAIN COMPLETE
   30 CONTINUE
      IF(MAX-N)31,99,99
   31 MAX=N
      GOTO99
      STOP
 CHAIN INCOMPLETE
  100 CONTINUE
      LAUNCH=LAST(LAUNCH)
 C    CALL STAR(.5,I,J,H)
      CALL CART(II,JJ,X,Y,H)
      CALL DRAW(X,Y,3)
   50 IC=IC+1
      IF(IC-6)60,70,70
   60 LAUNCH=NEXT(LAUNCH)
      I=II+ID(LAUNCH)
      J=JJ+JD(LAUNCH)
      CALL CART(I,J,XN,YN,H)
      XN=0.3*(XN-X)+X
      YN=0.3*(YN-Y)+Y
      CALL DRAW(XN,YN,3)
      CALL DRAW(X,Y,2)
      GOTO10
 CUL DE SAC
```

```
  70 CONTINUE
     IF(MAX-N)71,72,72
  71 MAX=N
  72 CONTINUE
     LAUNCH=KL(N)
     N=N-1
     IF(N)80,80,90
  80 CONTINUE
     WRITE(3,81)
  81 FORMAT(9H NO ROUTE)
     GOTO99
     STOP
  90 CONTINUE
     CALL CROSS(II,JJ,H)
     II=IX(N)
     JJ=JY(N)
     CALL CART(II,JJ,X,Y,H)
     CALL DRAW(X,Y,3)
     IC=7
     IC=6
     LAUNCH=LAST(LAUNCH)
     GOTO60
CHIP I,J FAILED TEST
  99 CONTINUE
     YA=YC+2.
     XA=XC-11
     X=N
     CALL SYMBOL(XA,YA,0.5,ITIT,0.,9)
     CALL NUMBER(XA+0.5,YA,0.5,X,0.,-1)
     X=MAX
     CALL NUMBER(XA+2.5,YA,0.5,X,0.,-1)
     X=NN
     CALL NUMBER(XA+4.5,YA,0.5,X,0.,-1)
     CALL NUMBER(XA,YA+0.7,0.5,YIELD,0.,4)
     CALL NUMBER(XA,YA+1.4,0.5,RR,0.,4)
     CALL SYMBOL(XA,YA+2.1,0.5,IAA,0.,3)
     WRITE(3,103)RR,YIELD
 103 FORMAT(7H START=,F10.6,9H : YIELD=,F10.6,2H :)
     WRITE(3,102)N,MAX,NN
     WRITE(3,103)R
     IF (NY.EQ.3) CALL PLOT (30.0,-60.0,-3)
     IF (NY.NE.3) CALL PLOT (0.0,30.0,-3)
 101 CONTINUE
 102 FORMAT(8H LENGTH(,I3,1H/,I3,4H)IN ,I3)
     XA =XXC(NX+1)-11
     CALL NUMBER(XA,0.,0.5,R,0.,4)
     CALL PLOT(0.,0.,999)
     STOP
     END
```


START
SET UP STARTING VALUES
TEST NEXT CHIP
Good
Prevent reaccess of this chip. Add it to the chain. Increment chain length. Reset counter IC Store chip co-ordinates and LAUNCH.
Bad
Yes
* at failed chip. Reposition pen at previous chip. Reset LAUNCH direction.
TEST CHAIN LENGTH
≥128
CHAIN COMPLETE
STOP (Success)
< 128
Take pen down to new chip position.
IS CHIP ON SLICE
No
Increment counter IC
Take next LAUNCH direction. Construct co-ords for next chip
CHECK COUNTER IC
≥ 6
DECREMENT CHAIN LENGTH
< 6
+ at chip I,J. Pick up co-ordinates of previous good chip. Take pen back there. Recall last LAUNCH direction. Set IC=6.
>0
TEST CHAIN LENGTH
≤0
NO ROUTE VIA THIS INPUT CHIP
STOP (Failure)

FLOWCHART FOR HEXAGONAL ARRAY SPIRAL GENERATION PROGRAM.