Design of a CMOS Colour Palette Integrated Circuit

for a Telidon Graphics Display

by

Gordon Cheng

B.A.Sc., The University of British Columbia, 1981

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Electrical Engineering

We accept this thesis as conforming to
the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1983

Department of ⋅ Electrical Engineering

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date ___ August 5, 1983 _____

ABSTRACT

The practicality of a simplified design method and algorithmic design tools in supporting large scale integrated circuit design was evaluated through the design of a custom CMOS integrated circuit (IC). The custom IC contained the colour map and digital-to-analog (DAC) conversion functions for a Telidon graphics display terminal. The digital data in the colour map determined the colour of the graphics display and was converted to analog signals to control the intensity of CRT electron guns. A scaled down prototype was designed, fabricated and tested. The prototype was functional except for a small number of errors in the layout. Subsequently a full scale prototype was designed. Simulation with the circuit simulator SPICE showed the circuit performance to be well within specification.

The simplified design method and the algorithmic layout tools provided an affordable and workable means to design ICs. Moderately large scale integrated circuits could be designed following this approach. The advantages of algorithmic layout could be combined with interactive graphics on microcomputers to provide a greatly enhanced IC design support at affordable costs in the immediate future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

I wish to thank Professors L. Young, G. Schrack and P. Lawrence for their guidance and support in my studies. I owe special thanks to Prof. Young for arranging some very insightful visits to major microelectronics companies, as well as to Dr. Colton of Northern Telecom and Mr. D. Klett of Mitel for their hospitality during these visits.

I am grateful for the ideas and cooperation of my colleague R. Mielcarski both in the thesis project and for his support as a personal friend without whom the timely completion of the thesis project would have been impossible.

I wish to express thanks to Mr. P. Thiel of Microtel Pacific Research for taking the major steps to secure my scholarship funding, and also for providing me with the opportunity to learn at MPR. I wish to thank Messrs. G. Schmiing and M. Pejskar for their kindness in providing much technical help in the IC project.

Finally, I would like to thank the Research Secretariat of B.C. for my scholarship award and the University of British Columbia for providing a teaching assistantship to support my studies in this program.

1.0 INTRODUCTION

Custom integration of electronic circuit functions on integrated circuits (IC) has for many years been practised by manufacturers of high-volume, specialized electronics products to reduce production cost and to achieve a competitive position beyond that possible by using off-the-shelf ICs. Most electronics product manufacturers cannot exploit large scale integration (LSI) technology for two main reasons: the cost of IC design and fabrication.

The cost to acquire a large scale integrated circuit fabrication technology and facility, estimated today at about $25 million US dollars, is prohibitively high for most companies. Alternatively the use of commercial IC manufacturing facilities are available but has traditionally been difficult to arrange due to reasons of propriety and cost. In addition the fabrication technology of commercial IC houses differ substantially such that an IC designed for a particular processing facility is in general not portable, hence the custom IC is in effect single-sourced. The single-sourcing of the chip which is usually a critical component of the product is economically and logistically unacceptable.

The availability of custom ICs to the common electronic firm improved greatly in 1979 when Mead and Conway [Mead80] introduced the "silicon foundry" concept [Hon79], a job-shop type fabrication facility for custom ICs. The process technology offered is standardized as is the interface to the foundries. Processing cost for each IC "run" is reduced by the "multi-project chip" concept where several IC designs are incorporated

on one die to share the mask making and fabrication costs. Further cost distribution is possible by including several dies on each wafer. Today IC prototypes can be fabricated at a cost of several thousand dollars in small quantities.

The second major problem to the small company is in IC circuit design. Although the processing cost of custom ICs is within reach, IC design cost still remains a major obstacle to fully exploiting custom integration. Several problems are encountered in custom IC design. The first problem is that the IC designer must learn to build electronic devices such as transistors and logic gates directly in the new medium as provided by the fabrication technology. A basic understanding of digital solid state circuit and IC fabrication technology is essential for most digital logic IC designs and is quite easily learned.

In contrast to the ease in gaining a workable knowledge of the process and circuits, managing circuit complexity is the major design problem [Spec83, Mead80]. Custom ICs usually encapsulate a significant amount of circuits and system functions on chip to maximize the cost effectiveness of the custom IC. As such the IC contains a large number of circuit elements ranging from thousands to hundred thousands of transistors. The IC design process is further complicated by the fact that IC design requires more steps than for digital design with off-the-shelf components. A complete IC design procedure and the corresponding software and hardware support is shown in Figure 1.1. Note the many design steps as well as the many tools needed. The last step in the IC design where the circuit
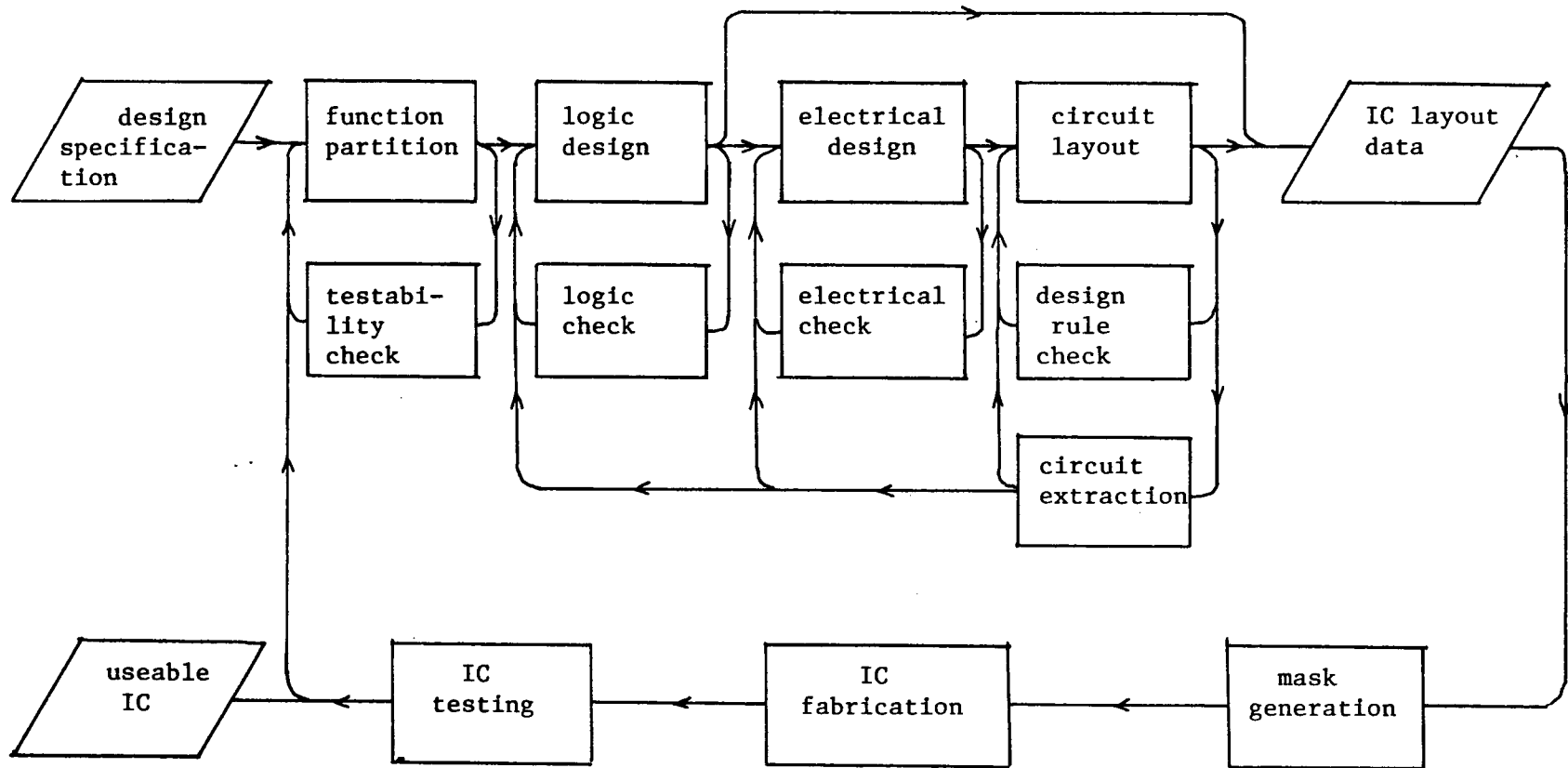
Figure 1.1   Custom IC design procedure

elements are "laid out" geometrically to build the physical devices is the most error-prone and most time consuming design step. Sophisticated and expensive computer-aided design (CAD) tools are essential to assist with the logic, electrical and geometric layout steps in IC design. The cost of commercial software circuit simulators is typically $50,000 to $500,000 US dollars. Although software simulators and geometric design rule checking programs help to eliminate most design errors, several iterations of design refinement and fabrication are commonly needed to obtain a useable IC. The availability of good IC design tools to support efficient and correct design is critical to custom integration considering the cost of each design and the turn-around time of fabrication. The CAD tools used in dedicated IC design houses are capable of supporting IC design extremely well such that very complicated ICs can be correctly designed with relatively little effort. However the cost of such a sophiscated IC design station ranges from $30,000 to $500,000 US dollars and is usually too expensive for the smaller company.

Semicustom IC design provides a solution to the design complexity problem by providing pre-designed circuits that the IC designer can connect together on a single chip to realize the desired circuit function. The mainstream semicustom design approaches are the gate-array and the library cell approaches. The semicustom IC design contracting work is currently a very competitive business; as a result excellent CAD and circuit design consultation support are available to the client.

1.1 Simplified Custom IC Design

While it is possible to contract custom IC design, the economics of total dependence on full custom or semicustom design houses may be unacceptable to the manufacturer who competes in fast evolving products such as office automation or communication equipment. For the electronic product manufacturer who wishes to acquire in-house custom or semicustom IC design support and expertise in reasonable time and cost, the so-called simplified design approach appears to be a viable alternative.

The simplified design approach was proposed by Mead in 1979 [Mead80]. In essence the simplified design approach applies simplifications and restrictions to IC electrical and layout design rules such that digital IC design could be learned quickly. Some circuit density is sacrificed in making the simplifications, but the IC designer's productivity would be greatly increased to allow functional, although usually suboptimal ICs to be designed quickly by designers with little expertise in IC design. It is projected that the continuous advances in processing technology would encourage large scale integration of complex and diversifying circuit functions on ICs needed in low to medium volume for highly specialized applications. In view of the design cost of custom LSI circuits, it would be well justified to make small sacrifices in circuit density to reduce design cost. Following a simplified design approach, system designers may be involved directly in system planning as well as in IC design thus reducing the time and the possibility of error in interfacing with special IC design groups.

The proposal of simplified IC design is similar to the definition of a programming language where the primitive statements are appropriately defined to facilitate data processing, and that slight sacrifices in code efficiency is accepted in return for greater increase in programming productivity. In the case of simplified IC design, the primitives are simplified layout rules and electrical rules for circuit design and layout.

1.2 Algorithmic Design Tools

In addition to defining the primitives, it was still necessary to decide how best to design and lay out the integrated circuit. To this end Mead proposed that digital LSI circuits and systems should be designed with an algorithmic approach. The algorithmic approach extends the ideas of program structure and algorithm to IC design. The idea of structure in IC design means that the design of a complex system such as an LSI circuit should proceed in a hierarchical manner. Although much practised in other areas of engineering, structured design was not explicitly or rigorously advocated in LSI layout and design until Mead's work in 1979. In structured design circuit functions are modularized such that complex circuits could be constructed hierarchically from succeedingly simpler circuits. The idea of algorithm involves the application of software programming concepts to IC layout by writing programs to place the geometric figures of the layout. Furthermore, programs can be written to automatically generate certain classes of circuits

subject to the application. A programmed logic array generator is a typical example of algorithmic IC design.

Algorithmic design is facilitated naturally by embedding in a high-level programming language the layout primitives such as polygons and wires to combine the geometric and the algorithmic aspects of algorithmic IC design. The program flow control and procedural features of a structured, high-level programming language would further enhance algorithmic IC design. A further advantage in embedding algorithmic layout in a high-level programming language was that an inexpensive but workable IC layout system could be obtained with a minimum of hardware required beyond conventional program development tools except for the addition of a colour plotter to generate check plots. The UBC PLAP software package is such an algorithmic design tools that the author programmed to support simplified IC design at UBC. The package consists of a layout program PLAP and a plotting program CIFP. The IC design procedure using PLAP for layout and SPICE for circuit simulation is shown in Figure 1.2. A typical design begins with a transistor level circuit schematic shown in Figure 1.3a. From the schematic the layout topology is explored by drawing a "stick" diagram as shown in Figure 1.3b. The stick diagram preserves the topological and physical processing information but eliminates geometric and design rule information to allow concentration upon global and topological layout planning. From the stick diagram, a grid layout of the circuit such as that shown in Figure 1.3c is drawn to specify the geometric features of the layout exactly. After the grid layout is checked for layout design rule violations,

function partition		logic design		electrical design		stick diagram		grid layout		PLAP program		PLAP
SPICE simulation		CIFP
design spicificat'n		CIF data
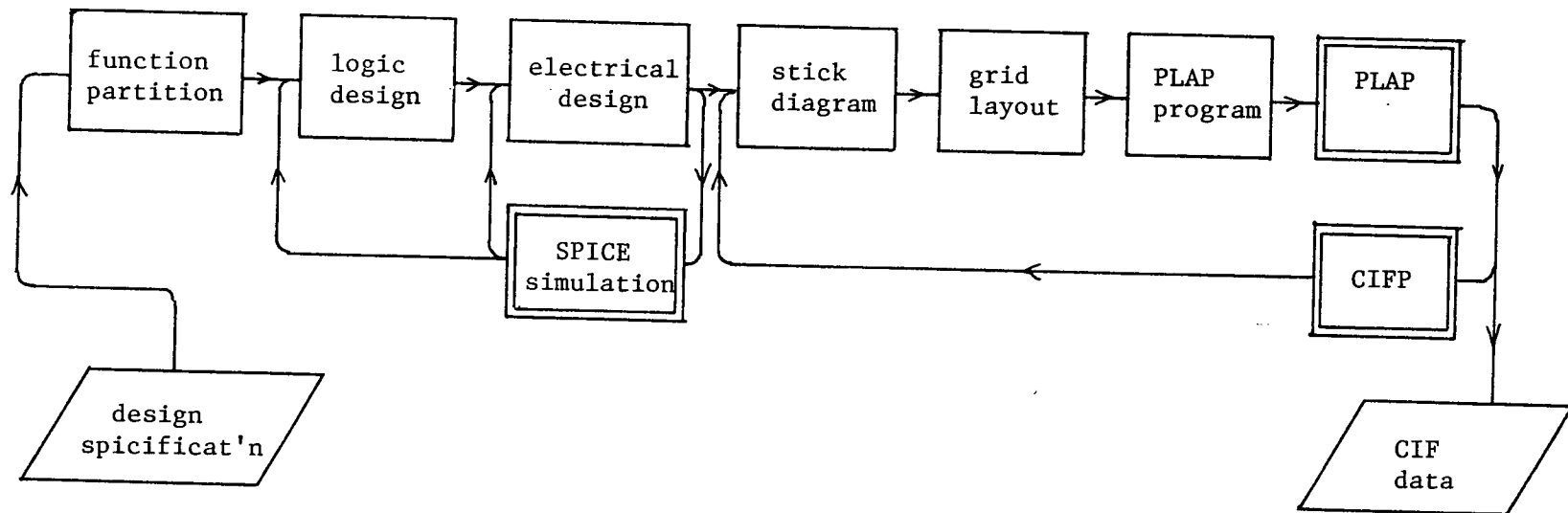
Figure 1.2   Custom IC design procedure with UBC PLAP
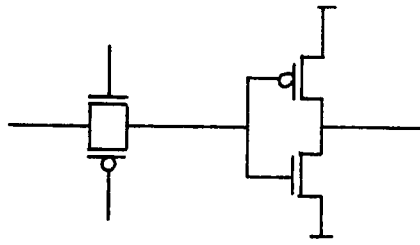
Figure 1.3a Transistor level schematic diagram
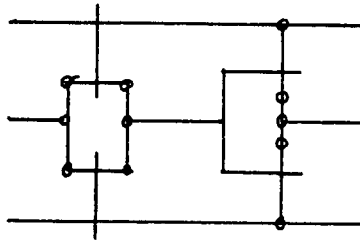
Figure 1.3b Stick diagram
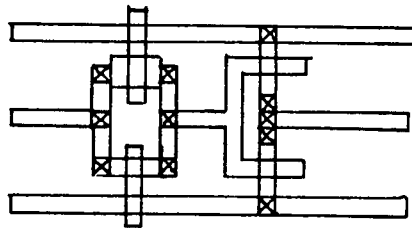
Figure 1.3c Layout on grid paper

```
Defc('shiftreg');
Technology(cmos);
Layer(metal);
wire(2,x0,y0); x(x1);

 .  .  .  .

Box(20,30,68,114);
Enddef;
```

Figure 1.3d Representation in a PLAP program

the designer writes a PLAP program to describe the layout as shown in Figure 1.3d. The PLAP program is then compiled as a PASCAL program and executed to generate a representation of the circuit in the Caltech Intermediate Form (CIF) for storage [Mead80, Hon79]. The IC is designed hierarchically: simple circuits (cells) are first laid out, then more complicated circuit modules are constructed from previously designed cells until the entire IC is obtained. The CIF code for each cell may be plotted using CIFP to check for layout rule violations and for documentation. If errors are discovered then the original PLAP program is modified, recompiled and executed to produce a new CIF file for the updated cell. When the entire chip is designed the CIF database is sent to a mask house for mask making, and fabrication of the IC then follows.

The structure of the PLAP package is shown in Figure 1.4. PLAP consists of four external modules built on the PASCAL programming language system.
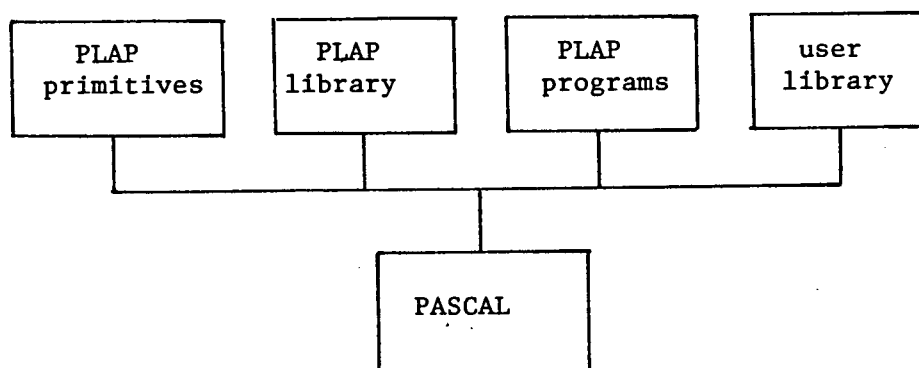
Figure 1.4 The structure of the PLAP software

The PLAP primitive module contains all primitive geometric features such as wires and polygons commonly used in IC layout.

The PLAP library contains pre-designed cells such as the input-output pad driver and the programmed logic array needed commonly in all IC designs. The PLAP source module contains the PLAP source codes to describe the layout which, when executed, causes the CIF code of the cells to be generated. The user library contains cells designed for a specific design project. Currently many companies and universities are experimenting with algorithmic design tools.

1.3   The Colour Palette IC Design Project

The practicality of the simplified design approach and algorithmic layout tools are investigated through the design of the colour palette chip, a custom integrated circuit for Telidon graphics display. The design of the colour palette chip may be viewed as a typical example for establishing in-house IC design expertise to design LSI circuits. The sponsor company is involved in the R/D of communication systems and office automation products. The potential consumer market for Telidon terminals aroused interest sufficiently to initiate a R/D program in this area. It was recognized that custom integration of graphics display control function was critical to making the product competitive in the market of low-cost graphics display hardware. Work began in 1981 to establish in-house design expertise and design support. The silicon foundry and multi-project chip concepts provided immense help to this end and access to an ISO-CMOS fabrication process was arranged. Design support was equally quickly established by installing the CMOS

Design System (CDS), a set of BASIC-based algorithmic design tool which was used for all MPC project designs within the company.

The design of the colour palette chip spearheaded the development of a proprietary graphics controller chip set for graphics display. This project is motivated by the implementation of new communication information and database systems to communicate textual and graphical data popularly available to the public. The definition of Videotex, a terminal-to-terminal text/graphics communication service was the first major step taken to establish a public database/information network. In Videotex, a network of host computers are linked through telephone lines to alphanumeric/graphics terminals. The Telidon system is a Canadian implementation of Videotex intended to support sophisticated text and graphical information display. The success of Telidon depends in part upon the availability of low-cost network terminals. However, the development of these terminals has been slowed by the lack of a clear interface and protocol standard thus creating problems in interfacing equipment from different manufacturers. The North American Presentation Level Protocol Syntax (NAPLPS) defined in 1982 [NAPL81,Flem83] attempts to solve the interface problem by providing a standard for the interchange of text and graphical data for Telidon-like systems. Original equipment manufacturers now has a clear guideline for the development of graphics terminals suitable for use in Telidon-like systems.

The raster graphics data terminal provides the most convenient and cost-effective means to display text and

graphical data. Graphical data can be output quickly by storing the picture image in a digital memory called the "bit map", and then output by scanning, thus raster graphics is often synonymous with bit-mapped graphics. The memory required in good resolution bit-mapped graphics is large, usually on the order of megabytes. The amount of memory, however, does not determine production costs because high density memory is available at very reasonable costs. The product cost determinant is the component count of the graphics control function circuitry and this count must be minimized to lower production cost. Circuits may be integratd on hybrid carriers or on silicon ICs to reduce component count. Currently only a few graphics controller ICs are commercially available. The NEC 7220 and the Intel 82720 are best known. In addition, several graphics controller chip sets have been announced by AMD and Texas Instruments but these components are either too expensive or will not be available in time. Display control function and high speed digital-to-analog converters integrated on thick film hybrids are currently available but the cost of approximately $100 per unit is not acceptable for the development of a high volume, low-cost intelligent graphics data terminal suitable for use in the market for Telidon or other consumer graphics terminals. Thus it was decided to custom design the IC chip set to meet the need for Telidon terminal production and also to serve as a vehicle to gain expertise in designing LSI within the company.

1.4 Thesis Work

The colour palette funcion of a bit-mapped graphics displayer was integrated on an ISO-CMOS silicon integrated circuit. The colour palette includes a digital memory for display colour control and a digital-to-analog converter to transform the colour code to analog signals to control the CRT electron gun intensities at video rate.

The design followed the simplified design method and was supported by a set of algorithmic layout tools. Te quality of the design, circuit performance, as well as additional circuit refinement was evaluated. Finally, the practicality of the simplified design method and the algorithmic layout tools was evaluated based on the experience gained in the colour palette IC design and in IC design work done at UBC using the PLAP software.

2.0 THE TELIDON TERMINAL

The Telidon terminal hardware may be partitioned into five major subsystems as shown in Figure 2.1. The major modules are the terminal processor, the communication module, the CRT controller, the frame buffer, and the colour palette module. An overview of system hardware and system operation is presented in Section 2.
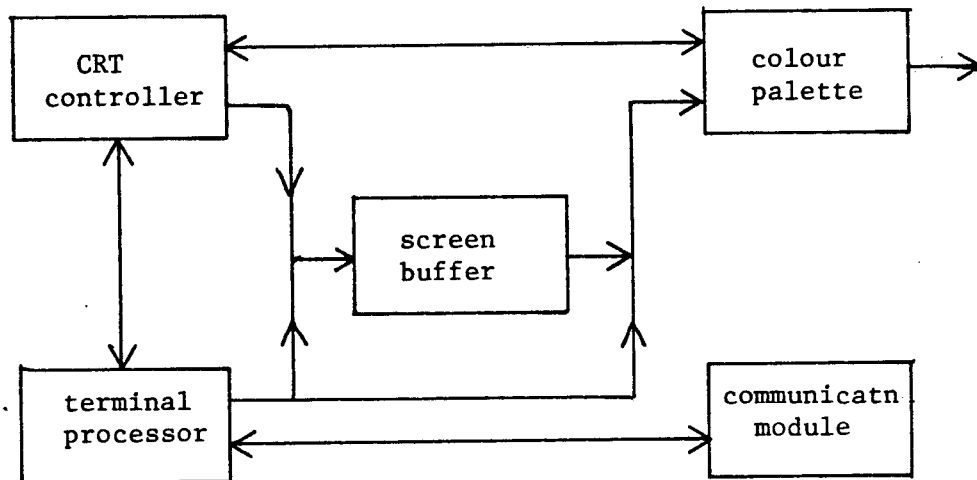
Figure 2.1

Major control functions of a Telidon terminal

2.1 Telidon Terminal Hardware and Operation Overview

The terminal is linked to the Telidon network via a phone line and communicates text, graphical data and control signals .following the NAPLPS standard [Flem83, NAPL81].

The terminal processor initializes the system, decodes NAPLPS protocol, communicates with terminal computer, and executes the Telidon data processing functions.

The communication module provides a standard interface

between the Telidon terminal and the network. The communication module is also responsible for port driving and data conversion to support serial or parallel busing of internal or external signals.

The CRT controller supervises the generation of signals to control the position and intensity of the CRT electron guns in order to display graphical data and text. Horizontal and vertical sweep is generated by dividing the master clock signal. In addition, certain video display functions such as reverse video, blinking, and various cursor types are also generated by the CRTC.

The colour of each pixel on the CRT screen is stored in a digital memory called the screen buffer. A pixel may be composed of several phosphor dots on the screen depending on the resolution of the graphics display and of the CRT.

The colour of a pixel is stored as a pointer to a word in the colour palette which contains the digital code to control the CRT guns' intensity. The terminal processor may modify the content of the screen buffer in accordance with the data to be displayed. The reading of the screen buffer to generate video output is controlled by the CRTC. Each colour word in the palette has three fields to code the intensity of the red, green, and blue electron guns. In the Telidon colour palette, each primary colour has 16 intensity levels. A 4-bit binary code is chosen to store the intensity code. For 16 colours to be displayed, 16 colour words are needed. Thus a 16 by 12 bit digital memory needs to be built which allows any 16 of 4096 different colours to be displayed simultaneously on the CRT.

2.2 Functional Description for the Colour
Palette Integrated Circuit

The colour palette integrated circuit combines the colour map and the digital-to-analog functions of the Telidon terminal. The colour palette is a dual-port register file of size 16 x 12 bits, and is accessed by both the terminal processor and the CRT controller. The terminal processor has read and write access to the palette via a 6-bit address bus and an 8-bit data bus. When the terminal processor receives a NAPLPS select colour command, it writes the appropriate colour word into the colour palette. Read access is required to determined system status and for self-testing. A colour word is written or read nibble-wise as a 4-bit pixel field. Data must be valid within 150 ns at maximum.

When a colour word stored in the palette is read for display, each of the three 4-bit intensity codes is converted to an analog signal. Three digital-to-analog converters are needed to perform this function in parallel. A pixel data rate of 186 ns (about 5.38 MHz) is required. The read access time of the colour palette register file is expected to be approximately 50 to 100 ns. The colour word read out and conversion are pipelined together by latching the data into a pipeline register for digital-to-analog conversion while the next colour word read access proceeds. The PCLK signal controls the latching of data into the pipeline register. The PCLK signal, generated either by the CRTC or by dividing the master clock directly, is synchronized both to the video address and to the horizontal and vertical sweep signals.

The speed of the digital-to-analog conversion critically

determines the display resolution. In order to avoid blurring during transition between pixel boundaries, the DAC should settle in about 20 ns. High speed DACs are traditionally fabricated in bipolar technology. Recently MOS technology gained popularity for the integration of low power, medium speed digital and analog circuit functions. In particular, D/As and A/Ds with 8-bits or higher resolution have been integrated on many communication controller and filter-codec chips [Dool80], thus it is relatively easy to design a 4-bit resolution DAC with a linearity of +-(1/2) LSB.

2.3 The Colour Palette Integrated Circuit Development Plan

It was decided to develop the colour palette IC in three phases: (1) design of a scaled prototype; (2) design of a full size prototype; and (3) production refinements.

The scaled protoytpe was intended as a vehicle to test the basic circuits needed to build the colour palette IC. Circuit functionality and performance would be validated to provide a collection of building blocks for the design in phase two.

The full-scale prototype builds on the results of phase one. Phase two had two objectives. The first objective was to test the circuit performance of a full scale palette IC built mostly from the cell library designed in phase one. The second objective is to test the performance of the IC in situ of the Telidon terminal hardware.

In phase three the colour palette IC would be perfected for production. In particular, the IC's functionality, performance, and reliability will be characterized before the IC would be released for production.

At the time of writing of this thesis the scaled prototype was designed, fabricated and tested with excellent results. The full scale prototype had also been designed and was being fabricated.

The architecture, logic, circuit designs and layout of the scaled prototype and the full scale colour palette are be described in Sections 4 and 5 respectively.

3.0 DESIGN OF THE COLOUR PALETTE SCALED PROTOTYPE

The colour palette prototype was fabricated through a multi-project chip vehicle. The area alloted to each design was 1500 by 1500 lambdas with lambda equal 2.5 um. The main objective of phase one was to design and test all circuit functions of the colour palette. In order to stay within the area allotment and also to reduce the work, a scaled down version of the entire circuit was designed which contained all major circuit functions of the final colour palette.

3.1 Architecture

The architecture of the full 12 bit by 16 word colour palette is shown in Figure 3.1. The circuit can be partitioned into modules which perform data path functions and control functions. Data path modules perform all data storage, buffering and data conversion functions. The colour palette data path register file consists of three identical 4-bit wide data paths, each of which contains the following circuits:

      (1)   digital to analog converter

      (2)   four-bit pipeline register

      (3)   four-bit by 16-word memory array

      (4)   bidirectional buffer for processor to memory access

The control modules generate the signals needed to control memory access. These include the four-bit wide and write signals for processor to memory access, and the twelve-bit (word) wide read signals for the video colour read-out.
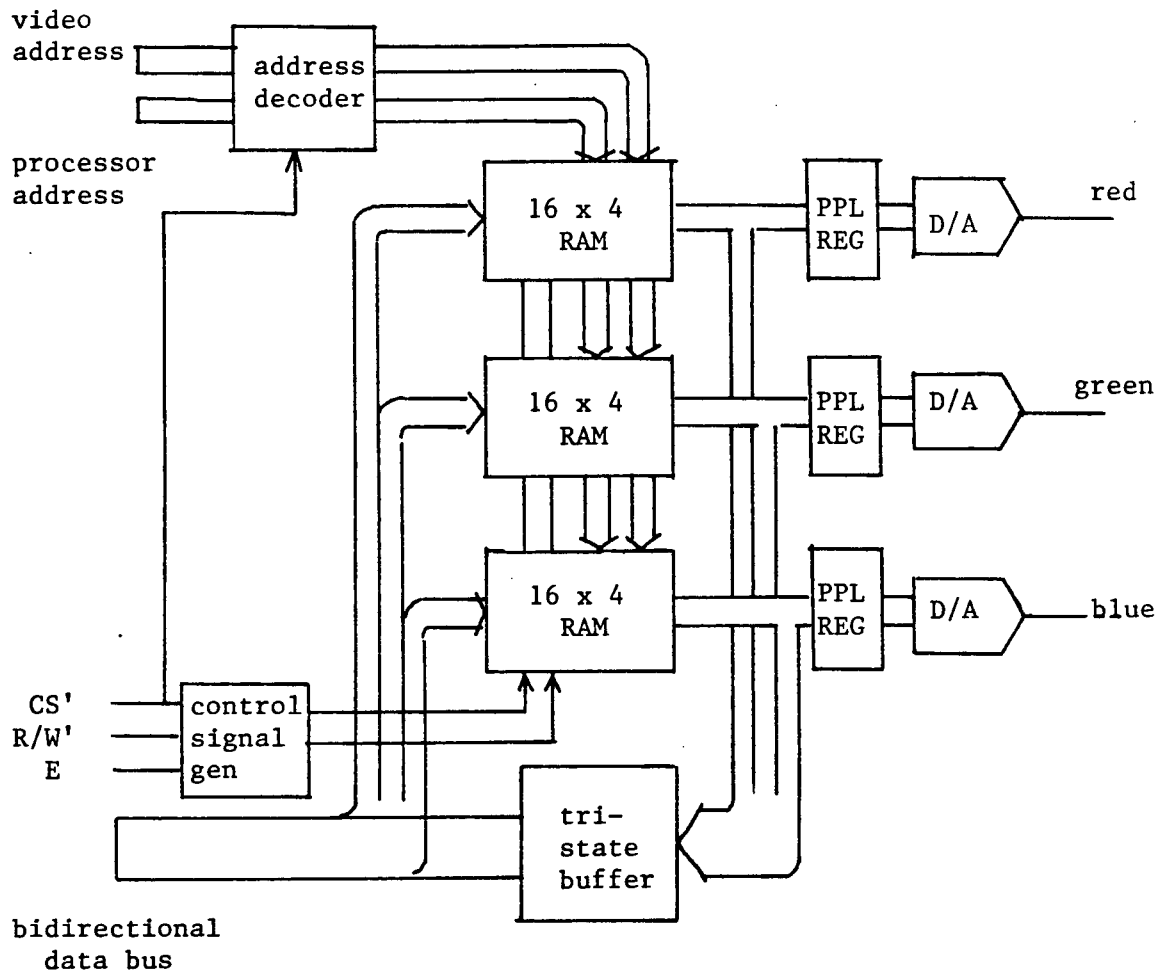
Figure 3.1

The architecture of the colour palette IC

The control modules are:

        (1)   video y-decoder

        (2)   processor y-decoder

        (3)   processor x-decoder

        (4)   control signal generator

The operation, circuit and layout of each of the major modules are described in the succeeding sections.

3.2 The Colour Palette Prototype

The colour palette prototype included all circuit functions necessary for the full scale palette chip. The circuits included were a 4-bit by 4-word memory array, a y-decoder for four words, and an x-decoder for processor access. Also included were the video circuits which contained a 4-bit resolution digital-to-analog converter and pipeline register. A control signal generator was included as well to generate signals to control memory read/write or storage.

The architecture of the colour palette prototype is shown in Figure 3.2. The collection of functions in the colour palette prototype is exactly the circuitry needed to control the intensity of one CRT electron gun, hence it is possible to test the dynamic operation of the system. To design full-scale colour palette, the single colour circuits three times with the memory depth expanded to sixteen words.

All cell in the colour palette prototype were designed to drive load to be found in the full colour palette chip, hence

these cells can be used directly in the design of the full scale colour palette chip. This approach minimizes the errors in proceeding from phase one to phase two design.

3.3 Circuit Operation

The symbolic logic diagram of the prototype is found in Figure 3.3. Chip select CS, chip enable E and processor to memory access mode R/W' perform conventional control functions. The processor provides six address bits A5 to A0. Address bits A5,A4 choose one of four nibbles from each colour word, and bits A3 to A0 choose one of sixteen colour words. Only one nibble per word and four words are designed for the prototype. Word select is performed by the y-decoder while nibble selection is the the task of the x-decoder. All address bits are buffered to generate the true and complement signals needed to drive true complementary gates. The chip control signals likewise are buffered and bussed to the control signal generator. An internal read signal R and write signal W are generated to control memory read, write and storage for each nibble. The R and W signal also controls the direction of data propagation through the bidirectional buffer between the memory array and the output pads.

The processor to memory access timing is shown in Figure 3.4. The E signal initiates the execution of the memory access operation after the chip is selected and after the access mode is set properly. Although only one colour palette is expected to be used in the Telidon terminal, CS is included to anticipate
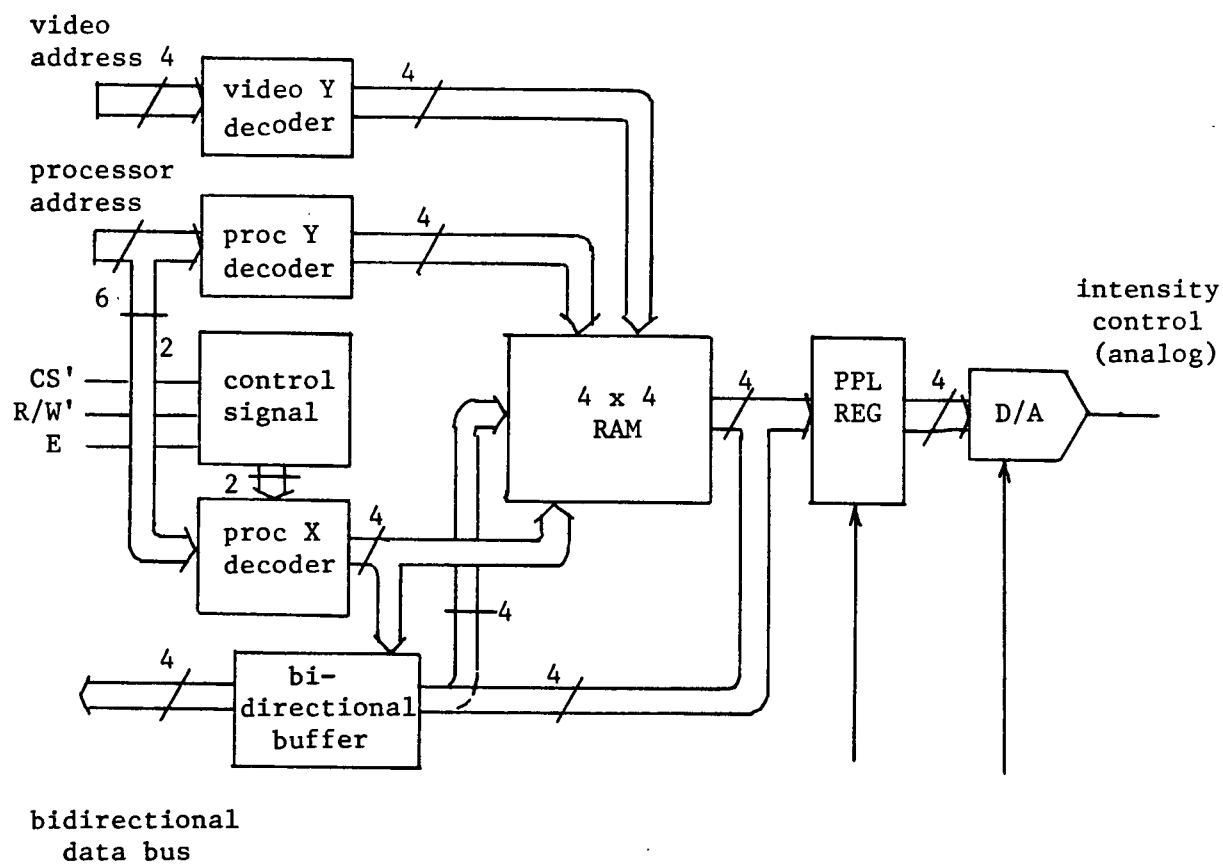
Figure 3.2

The architecture of the colour palette prototype
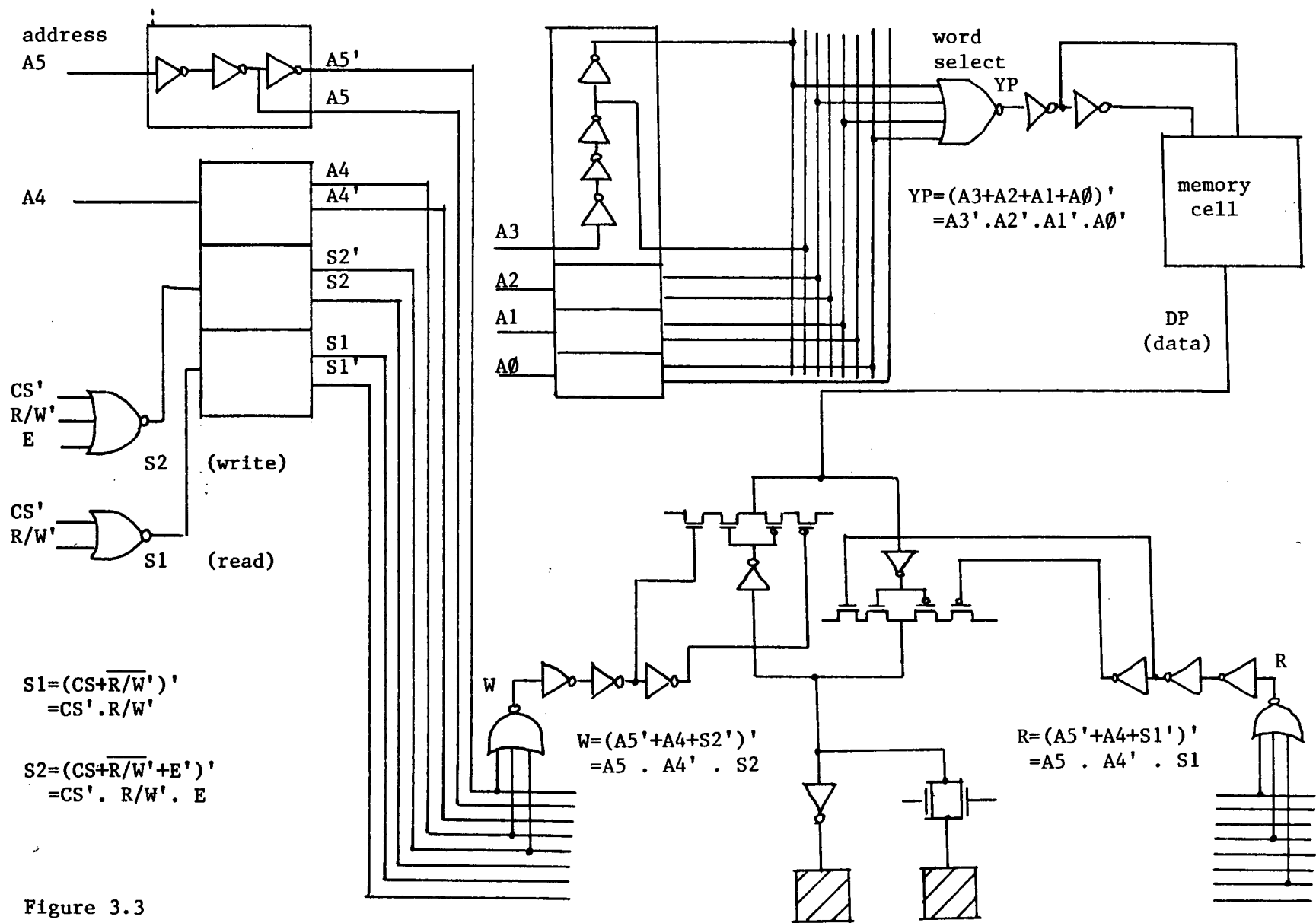
Figure 3.3

Colour palette symbolic logic diagram

address
A5

A5'

A5

A4

A4

A4'

S2'

S2

S1

S1'

CS'
R/W'
E

S2 (write)

CS'
R/W'

S1 (read)

$S1=(CS+\overline{R/W'})'$
$\quad =CS'.R/W'$

$S2=(CS+\overline{R/W'}+E')'$
$\quad =CS'. R/W'. E$

A3

A2

A1

A0

word
select

YP

$YP=(A3+A2+A1+A0)'$
$\quad =A3'.A2'.A1'.A0'$

memory
cell

DP
(data)

W

$W=(A5'+A4+S2')'$
$\quad =A5 . A4' . S2$

R

$R=(A5'+A4+S1')'$
$\quad =A5 . A4' . S1$
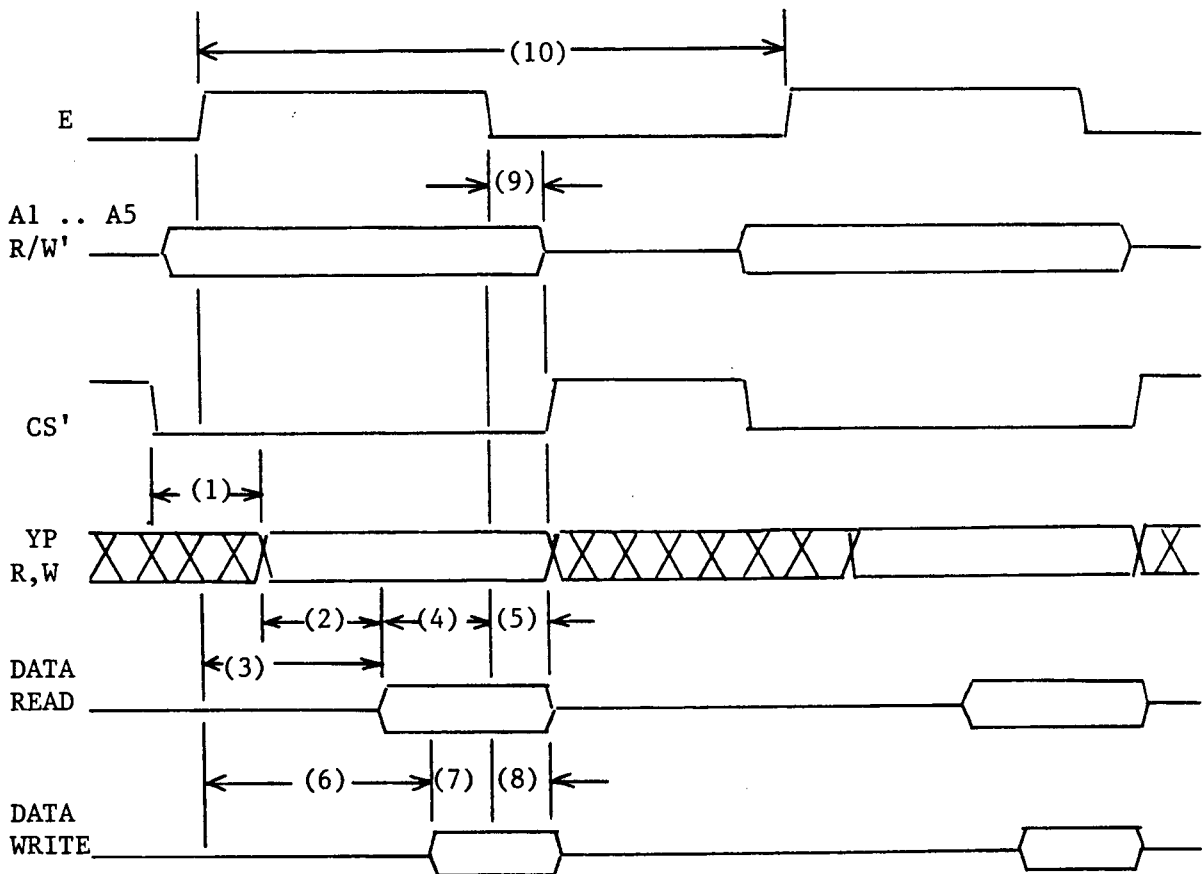
no.	time	min (ns)	max (ns)
1	Address decoding	---	---
2	Internal memory read time	---	---
3	Memory access time from E valid	0	150
4	Read data valid time	100	---
5	Read data hold time	20	50
6	Memory write data delay from E valid	---	---
7	Write data set-up time	20	---
8	Write data hold time	20	50
9	Addressm R/W', CS hold time	20	50
10	E clock cycle time	500	500

Figure 3.4

Processor-to-colour palette memory access timing diagram
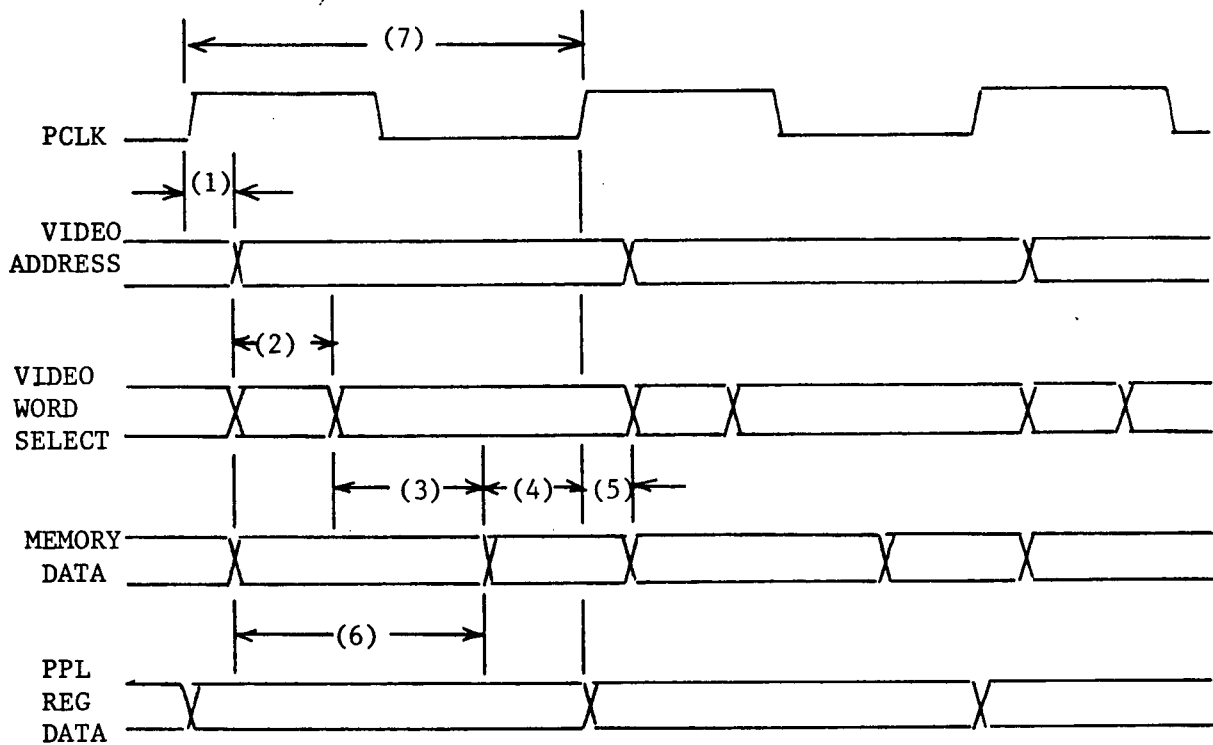
no.	time	min (ns)	max (ns)
1	Video address delay after PCLK	---	20
2	Video address decoding delay	---	---
3	Memory data read access time (internal)	---	---
4	Memory data valid time (internal)	---	---
6	Video data read access time from PCLK valid	---	120
5	Memory data hold time (internal)	---	20
7	PCLK cycle time	186	186

Figure 3.5

CRT controller (video)-to-colour palette access timing diagram

the possibility of bank switching to allow fast exchanges of palettes for future graphics capability enhancements. Address decoding begins immediately after all inputs CS,E,R/W' become valid. The y- and x-decode operation proceed in parallel.

Once a memory cell is selected, data can be read out or written into memory. The write and read access are relatively fast in comparison to the address decode delay. The address decode and word select delay is expected to dominate the memory access time because of the load presented by many memory cells to the address decode circuits.

The time that the memory output data is held after the address and control signals become invalid is obtained indirectly from the delay arising from address decode since no change in the output is possible until the new values of CS, R/W' and E have propagated through the control signal generator to change the internal R and W signals. This delay is identical to the address decode and select delay, which is expected to be at least 20 ns.

In the operation of the video section, the video addresses P3 to P0 are decoded by the video y-decoder to select the colour word to output to the digital-to-analog conversion circuit. The colour word is latched into the pipeline register by the PCLK signal. The data is held in the pipeline register for the digital-to-analog converter for a full PCLK cycle.

Notice that the a memory access conflict can arise when a processor writes into a memory cell simultaneously as the cell is read by the video circuit. This conflict cannot be avoided because the PCLK and the terminal processor clocks are not

synchronized and no handshaking is used to coordinate their memory accesses. Fortunately such a conflict will occurs only rarely because processor writes occur seldomly once the palette colours are set. If a conflict does occur, a screen pixel will display an unpredictable colour for the duration of one PCLK cycle of 186 ns. This conflict therefore does not seriously affect the quality of the graphics display.

3.4 The Memory Cell

The memory cell stores one bit of information to code the intensity of the CRT electron gun. The memory could be designed in static or dynamic circuits [Elma81, Howe82]. The static memory cell is basically a bistable circuit. The advantages of the static memory are in its low stand-by power consumption and in the simplicity of circuit design [Klei69]. The disadvantage of static memory is the large circuit area relative to dynamic memory. The dynamic memory stores information as a small charge on MOS capacitors. Extremely high circuit density is achieved by the use of very small capacitors for data storage [Abot73]. The charge stored on the MOS capacitor is not permanent, however, because charge is lost through transistor leakage. Data loss is avoided by periodically replenishing the charge stored on the MOS capacitor. This operation, called memory refresh, reqires additional off-chip memory support circuitry. The so called "pseudo-static" RAMs integrate the refresh signal generator onto the memory chip to relieve the user of the refresh task. On-chip refresh generation, though convenient to the user, adds

complexity to the IC design.

In view of our modest IC design experience in ISO-CMOS and of time constraint, it was decided to implement the colour palette memory array as a static memory. Two kinds of static circuits, completely static or "quasi-static" circuits may be used to relize the static memory cell. The most common static memory cell is the six-transistor cell shown in Figure 3.6. The memory element is a bistable circuit built by a cross-coupled inverter pair. The output nodes to the data buses are gated by a pair of transmission gates. Data is loaded into the memory cell simply by enabling the transmission gates and then forcing the bistable into the desired state. Data read-out may be either in the single-ended or the complementary mode. Despite its apparent simplicity, the design of the completely static memory cell requires careful consideration of the noise injected into the cell during data read by the data bus sense current and of the data line transient noise coupled into cells sharing the same data line. The presence of this noise could cause parasitic state change in the memory and cause soft errors. A detailed discussion of the static memory cell is given in Section 5.

The quasi-static memory cell, shown in Figure 3.7 provides a compromise between simplicity in the memory cell circuit and in memory access control. The quasi-static memory cell needs no refresh in the continuous storage mode, but the bistable inverter pair cross-coupling is switched during memory write. Bistable cross-coupling switching adds circuit to the conventional memory cell where only the transmission gates are switched during memory access.
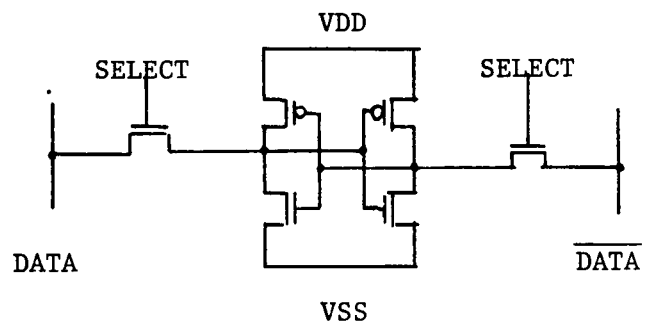
Figure 3.6

The 6-transistor CMOS static memory cell
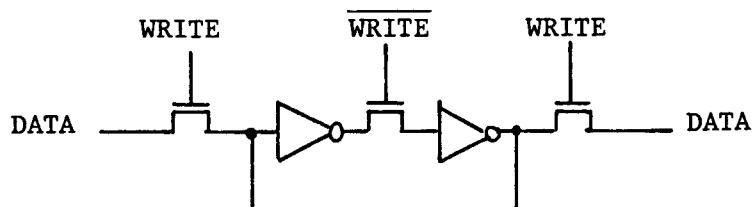
Figure 3.7

The quasi-static memory cell

The colour palette memory cell design basically followed the circuit of Figure 3.7. The gated inverter combines the CMOS complementary transmission gate and the CMOS inverter to reduce circuit area [Susu73]. Additional inverters are inserted to buffer the data output such that sense amplifiers will not be needed.

In the layout of the memory cell and memory array, data lines were wired in the metal layer and control lines were wired in the polysilicon layer to minimize propagation delay in the data path. The power busses were all wired in the metal layer for simplicity although the diffusion layer could be used [Isob81, Howe81].

3.5 Memory Peripheral Circuits

Memory peripheral circuits are the address decoders and data read out circuitry. Address decoding can be dynamic or static. Static memory decoding is achieved by using static CMOS logic gates. The simplicity in static decoding is obtained at the cost of circuit area because a fully symmetric CMOS logic gate requires that both N and P channel transistors to have equal current sourcing/sinking capability to ensure equal rise and fall times. In the case of address selects, equal select rise and fall times are necessary to minimize the interaction between the selected and unselected memory cells sharing the same bit lines [Frie68]. The static CMOS logic gate consumes much for two reasons. First, the P transistor channel width-to-length ratio must be 2 to 2.5 times that of the N transistor

because the process parameter for the P channel transistor is smaller than that of the n channel transistor by the same factor [Burn64, Cher69]. Second, both the true and the complement of all control signals are needed for the operation of truely complementary circuits. In terms of circuit area, the true CMOS static address decoder is practical only for decoding a small number of address inputs.

Dynamic decoding is used in LSI memory where a large number of address inputs are decoded. By trading circuit complexity for layout area, the dynamic decoder consumes less circuit area relative to the static decoder. A precharge technique is used to charge up the output node dynamically thereby making load devices unnecessary. Precharging saves more than half the circuit area because the P channel pullups which occupy about twice the N channel transistor area are now unnecessary. Complementary input signals are still needed of course. The disadvantage in dynamic decoding is that the precharge control must be generated from an externally supplied signal to control the precharge operation. Obviously the precharge control signal cannot be derived from the CS,E, or R/W' inputs because a chip may be continuously placed in the read or write mode while addresses change. Alternatively, the precharge control can be generated on chip by detecting an address change [Stew77].

For the colour palette, only four bits need to be decoded to select one of sixteen colour words. Since the number of address inputs was manageably small, the static decoder was the most suitable decoding method. In the layout of the x and y-decoders, the complementary address signals were bussed between

the N and P channel transistors forming a cross-point grid with the N to P channel bridges. The decoder logic is programmed simply by placing contacts at appropriate cross-points.

3.6 Video Output Circuits

The video output section consisted of the digital-to-analog converter and the pipeline register. The pipeline register was a quasi-static master slave register circuit containing two quasi-static registers such as that described in Section 3.4. The pipeline register was clocked by the video read clock PCLK to perform the sample and hold functions in order that the colour word data presented to the digital-to-analog converter was valid at all times except for a brief instant when new data was latched into the pipeline register slave stage. In the sampling interval, data was clocked into the master register on PCLK'. This data was passed to the slave register on PCLK and held for the digital-to-analog converter, hence the pipeline register output was valid for almost all of the PCLK period.

The digital-to-analog converter transforms a four-bit colour intensity code to an analog signal to control the CRT electron guns. The primary requirements of the DAC are high speed and low resolution.

The two conventional approaches [Greb72] to build DACs are voltage and current switching circuits. The current switching DAC consists of a collection of binary weighted current sources. The current sources are either switched to ground or to a current summer depending on the data input which control the

switching network. The current sources are switched on at all times to minimize signal transients when a current source is switched in to the summing node.

In the conventional voltage switching DAC, binary weighted resistances are driven by a reference voltage to generate a binary weighted set of currents. The currents are then switched and summed as in the current switched DAC.

The R-2R resistor ladder is the most common approach to generate binary weighted current sources. The entire range of binary weighted currents can be generated by using only two values of resistances by using a single voltage source or a single current source as reference. The conversion accuracy for a R-2R DAC is superior to the binary current source or binary resistance DAC because the errors associated with a wide range of source value or resistance values are absent. Further, the circuit area is reduced since the large resistances (and large area) needed to build high resolution DACs are not needed.

The DAC used in the colour palette protoytpe, available as an MPR library cell, was a simplification of the voltage switched R-2R DAC. The colour palette DAC circuit uses a resistor ladder of sixteen equal value diffused resistors constructed by inserting taps at equal distance along the length of a diffusion strip. The sixteen taps provide sixteen analog voltage levels. The taps were switched to the output point by a pass transistor controlled by a binary decoding network which selected one of sixteen analog voltage taps to output depending on the four-bit colour intensity code. The DAC's accuracy is maintained if no current is diverted from the resistor ladder

into the output load. The output node therefore must be terminated by a high-impedance voltage follower with a sufficiently wide bandwidth to track the output signal at a video data rate of 5.4 MHz. An output voltage swing of 1 V was needed to control the electron gun intensity circuit. Sixteen diffused resistors of 66 ohms each were driven by an externally supplied 1 volt voltage reference to give a static dissipation current of about 1.5 mA.

3.7 The Multi-Project Test Chip

The colour palette prototype, along with a collection of test cells were fabricated on a multi-project chip. Static testing only was possible with the test cells because limited area did not allow the inclusion of output pads to appropriately drive the off-chip load. Nonetheless, these individual test cells allowed the palette IC to be tested in a modular way such that problems in circuit function could be isolated and verified easily. The test cells included on chip were a memory cell, an x-decoder, a y-decoder, a bidirectional buffer, a pipeline register, a control signal generator and a simple process monitor. A plot of the colour palette chip is shown in Figure 3.8.

3.8 Test Results

The MPC die was bonded and returned for testing in January, 1983. Static testing verified that all testable test modules
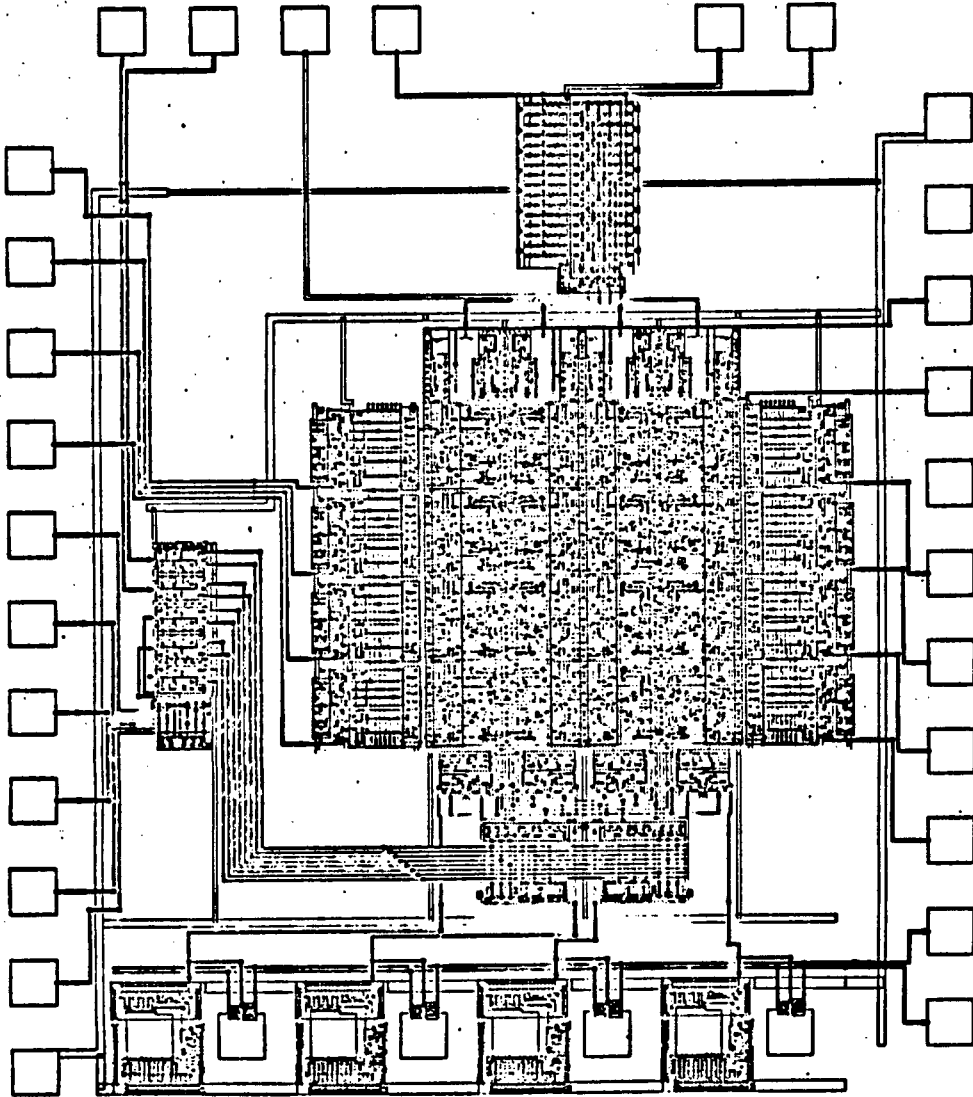
Figure 3.8

Layout of the colour palette prototype IC

worked properly. The dynamic testing of the colour palette prototype revealed several design errors:

(1) The power lines of the processor to memory access output pads were connected backward. The power line reversal forward-biased all device-to-substrate junctions and caused current on the order of several hundred milliamps to be drawn. The high current rendered the palette IC untestable for fear of damaging the IC.

The problem was solved by cutting the power lines to the output pads by the use of an ultrasound excited probe. The cut, however, disabled the processor output circuit such that dynamic testing of processor read (where memory data output must be driven by the output pads) is no longer possible.

(2) The P-well of the processor input pad transmission gate was not connected to VDD. Consequently, charge accumulated in the P-well due to substrate leakage current. Since the charged P-well, now probably at or near VDD, was the substrate for N channel transistors, the input pad N channel transistor threshold voltage was raised to above 3 V. As a result, the N channel transistor did not turn fully on, and it was impossible to discharge the internal data bus. Since the internal data bus must be pulled to VSS to write a logic '0' into memory, it was impossible to write in a logic '0'. Fortunately the memory was initialized to the logic '0' state upon power up, so that

at least a partial dynamic check of memory write and storage was still possible.

(3) A metal to metal spacing violation at the y-decoder sporadically shorted the A0 bus to Gnd. As a result, only even numbered memory locations was read reliably.

(4) The lack of proper buffering between the digital and analog circuit of the DAC allowed noise to be coupled to the analog output. As a result the analog output was corrupted by high frequency glitches arising from the dynamic intensity decoding, and also by a DC offset which exceeded the one-half LSB level (63 mV in this case).

Once the proper steps were taken to avoid the above problems, dynamic testing of the video data path gave very encouraging results. Data was loaded into memory and read out successfully. The memory addresses were generated by a cyclic sixteen state counter in synchronization with the pipeline register clock PCLK. Four different bit patterns were loaded into the four colour words and the memory words were read out cyclically. The output of analog signals was valid up to at least 13 MHz. The analog output showed that for a load of 15pF, the DAC settles in less that 3 ns. The test results verified the functionality and performance of most major circuit modules of the colour palette chip and provided directions for improvement in the phase two design.

4.0 THE FULL SCALE COLOUR PALETTE INTEGRATED CIRCUIT DESIGN

Phase two of the full scale colour palette design began on April 25, 1983. In the phase two design, the colour palette was expanded to include the circuit for the red, green and blue electron guns. Layout mistakes made in the phase one design were corrected. A new digital-to-analog converter was designed to minimize the noise in the analog video output. The dynamic DAC used in the prototype was also included on the chip to ensure that D/A conversion was possible in anticipation of circuit errors in the new DAC. Separate test circuits such as those included on the phase one MPC chip were not included for the lack of area.

Selected critical paths of the full scale colour palette were simulated on SPICE. Simulation results predicted that the IC performance would be well within specification. An error margin as high as 50% could be tolerated.

The full colour palette design was completed on May 13, 1983, consuming nine man-weeks of work. The same design tools as those used for phase one design were used: the MPR CDS software package running on the HP9826 or HP9836 desk top computer. Check plots are output via the HP7220 eight pen plotter. Section plots which windowed into parts of the layout were generated on the VAX 750 and output via the HP7220 to allow careful checking of the layout for design rule errors. A great deal of time was saved by using the section plotter because only the portion of layout of interest needed to be plotted. The resolution of the output was also much improved since the output is scaled to fill the maximum platen area of the plotter.

The full scale colour palette chip contained approximately
6000 transistors and measured 0.4 cm by 0.6 cm in size. The
circuit was being fabricated at the time of writing of this
thesis.

4.1 The Static Digital-to-Analog Converter

The static digital-to-analog converter was the only new
circuit included on the full scale colour palette chip. The
dynamic decoding circuit was replaced by a static decoder. The
NOR gate loads were implemented by continuously switched-on P
channel devices. The high frequency noise caused by the dynamic
precharge operation found in the dynamic DAC was thus removed.
Additional buffering between the digital and the analog
circuitry was provided by placing large power and ground busses
between the two sides. Separate grounds for the digital and the
analog circuitry also helped to reduce noise.

The transmission gates which switched the selected tap
voltage to the output node were modified to fully complementary
transmission gates to increase the output voltage swing to about
4 V. The static DC offset found in the dynamic DAC may now be
reduced by properly setting the gain of the off-chip buffer to
scale the video output down to 1 volt voltage swing. The
reference voltage was 5V. The resistor ladder resistance was
increased appropriately to maintain a 1 mA maximum dissipation
current.

5.0 MEMORY CELL CIRCUIT REFINEMENT

The circuit area of the full-scale colour palette chip, as shown in Figure 3.8, was dominated by the static register file. Not withstanding the reduction in the number of chips that can be manufactured per wafer, the large die size could lead to low yield and increased packaging cost. Therefore the full scale colour palette chip area should be reduced through some additional refinement. Since the memory array consumed most of the die area, and since it was not optimal in design, the memory circuit is the first candidate to be redesigned. The memory unit cell in turn must be optimized because it was repeated many times in constructing the memory array. Changes in the memory cell design would necessitate the redesigning of data read-out circuits. A design refinement of the memory cell circuit is presented in Section 5.

5.1 Memory Cell Design Alternatives

The prime objective in the redesigning of the memory cell was to reduce the memory array layout area. To this end commercial random access memories (RAM) ICs provided excellent guidelines [Holl78, Brig78, Akiy79, Isob81, Abot73, Elma80]. Careful consideration was given to the two mainstream digital memory circuit techniques to determine whether the colour palette memory array should be realized as a static or a dynamic RAM.

Dynamic RAMs (DRAM) are ideal for very large scale integration. DRAM chips currently in the market have a memory

capacity of up to 128K bits. The DRAM stores the information bit as a charge on a very small MOS capacitor to allow high density integration of memory elements. The capacitor charge is replenished periodically, typically once every one to two milliseconds to replace the charge leaked through the oxide-to-substrate isolation and through the non-conducting channel of OFF-transistors. The refresh circuitry integrated on chip requires additional refresh signal generation support circuits off-chip. Hence extremely high density memory is achieved at the expense of circuit complexity.

Static RAMs (SRAM) rely on bistable circuits to store the information and thus are less dense than dynamic RAMs. Because the information is stored in a bistable circuit, no refresh operation is necessary and the memory circuit is simplified at the expense of memory density. SRAMs are available in both the NMOS and the CMOS technology. Relative to the NMOS SRAM, the CMOS SRAM had the advantage of low stand-by power dissipation. However, the CMOS memory requires about twice the circuit area of NMOS SRAMs. A typical SRAM memory capacity is about 2K to 4K bits.

The colour palette requires only a 192 bit memory organized in a word-wide configuration of 12-bits by 16-words. The modest memory size is thus best implemented as a SRAM to avoid the DRAM circuit complexity. The memory array is constructed by a new dual port CMOS static memory cell. It was decided to design the dual port memory cell in two steps. First the basic six-transistor single port CMOS static RAM cell was investigated to obtain an understanding of the circuit design problems. In the

second step, the results from step one were extended to finalize the circuit and layout of the dual port static memory cell.

5.2 The CMOS Static RAM Cell Circuit

A survey of the literature and the IC market showed that a 5-transistor and a 6-transistor variation of the basic cross-coupled inverter pair bistable circuit are used in SRAM cell design as shown in Figure 5.1. The 6-transistor cell was the first bistable used for high density memory element application [Hodg68]. The 5-transistor cell was derived from the 6-transistor cell when it was noticed that only one read path was needed to access the bistable circuit; hence the opposite transfer gate was removed. The 5-transistor cell consumes approximately 30 percent less area than the 6-transistor cell [Brig78]. However the 5-transistor cell was more susceptable than the 6-transistor cell to state reversal caused by noise injected into the memory cell during read access [Holl78,Brig78]. Extra precaution would be needed in the 5-transistor cell design to prevent soft errors caused by destructive readout.

In the literature, much debate is directed to whether the memory cell transfer gate (G1,G2 in Figure 5.2) are best implemented by an N-channel, a P-channel transistor or a complementary transmission gate [Akiy79, Isob81]. In the interest of high circuit desity, the transfer gate is commonly realized by a single channel gate instead of a complementary gate. There is much controversy in the choice of the single

VDD

SELECT

SELECT

DATA

$\overline{\text{DATA}}$

G1

(1)

(2)

(3)

G2

(4)

sense
current

Cg1

Cg2

VSS

VSS

VSS

5-transistor memory cell
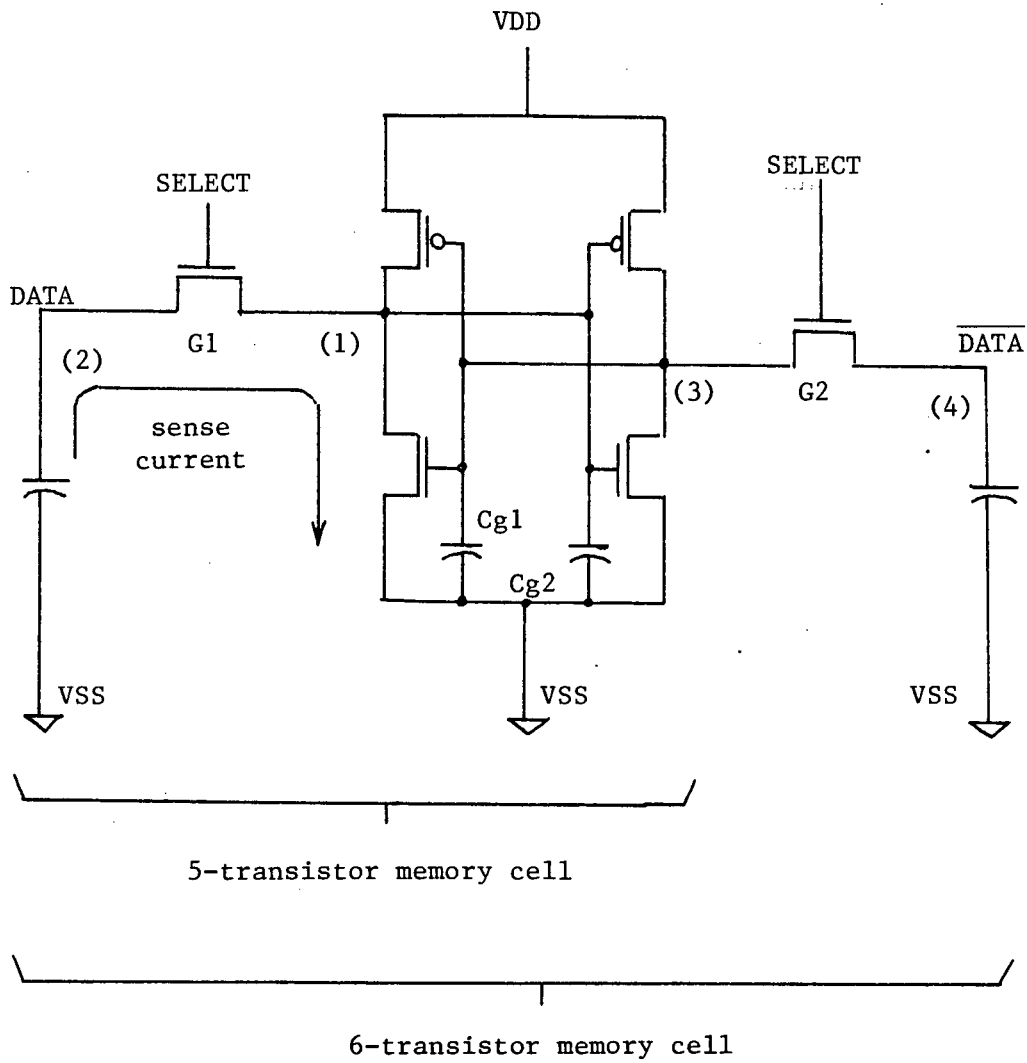
6-transistor memory cell

Figure 5.1

The five- and six- transistor CMOS static memory cell

channel transfer gate between the N and P channel pass transistor. Both the N channel transfer gate [Akiya79] and the P channel transfer gate are used in commercial memory products. The exact choice of the transfer gate type is made after consideration of circuit operation and circuit stability. It was decided that the ease of design and the greater stability of the 6-transistor cell more than compensated for the 30 percent area penalty for the small colour palette memory.

5.3 Basic Circuit Operation

The basic CMOS static 6-transistor SRAM cell is shown in Figure 5.1. The bistable storage element is made up by two inverters which are cross-coupled. The data are stored as gate charges on the internal nodes (1) and (3). In the stable state, the internal node voltage is either VSS or VDD and the voltage level of the internal nodes are complementary. The stored datum and its complement are gated to the output nodes (2) and (4) via transfer gates G1 and G2 as controlled by the SELECT signal. The output node capacitance lumps the sense amplifier input capacitance, the interconnect capacitance, the bus-line-to-substrate capacitance and other stray capacitances.

The bistable inverter consists of the P-channel pullup transistor and the N-channel pulldown transistor. The substrates are appropriately biased to set the source-to-substrate voltage to zero.

In the storage or the standby mode, the transfer gates are non-conducting and isolate the bistable storage nodes from the

bit lines. The inverters are in complementary states. For each inverter either the pullup or the pulldown transistor is ON and the other transistor is OFF, thus the internal nodes are either pulled up to VDD or pulled down to VSS. Since the transfer gates are OFF in the standby mode, only a very small current typically on the order of several picoamperes is drawn through the ON transistor to replenish the storage node charge lost via oxide insulation leakage and reverse saturation leakage through the channel of the OFF transfer gates.

During a read access the transfer gates are turned on. The bit lines, precharged to VDD or VSS, appear to the memory cell as a passive capacitance. The bit line capacitances are then discharged through the internal node connect to the opposite power supply or remain unchanged if the internal node is at the same voltage level as the precharged bit line. The current from the discharge of the bit line capacitance, shown in Figure 5.1 as Isense, is amplified by a data sensing circuit for output.

In a memory write operation, the bit lines are connected to active complementary sources and the transfer gates are turned on. The external sources charges or discharges the internal nodes via the transfer gates thus forcing the bistable circuit into the desired state.

In general, the pullup transistor is the smallest, that is, its d.c. transconductance is lower than that of the pulldown or the transfer gate since it need to supply only the leakage current for the storage node of the opposite inverter. The pulldown is the largest in order to sink the sense current adequately to provide fast data read out. The transfer gate size

is in general smaller than the pulldown to limit the sense current in order to avoid state reversal during read access. The topic of memory cell stabilty is discussed in Section 5.4.

5.4 Memory Cell Instability

The memory cell can fail to operate reliably for many reasons some of which are discussed in this section.

5.4.1 Problems Related to Processing

The native threshold voltage of the N and P channel transistor should ideally be matched, that is, be equal in magnitude [Burn64]. The logic threshold and the noise margin of the transistor circuit depends directly on the threshold matching therefore the noise susceptability of the memory is directly affected [Klei69, Elma81]. Drift of the native threshold voltage of 30 percent is not uncommon due to processing variations. A consideration of worst case native threshold voltages would therefore be necessary to ensure that the circuit is stable and meets speed specification.

During circuit operation, the transistor threshold voltage often shifts due to the back gate bias effect [Cher69, Cobb70]. This phenomenon could be viewed as an equivalent reduction in the width-to-length (W/L) ratio of the gate of the transistor. In the operation of the memory cell when the sense current sinks to ground through the pulldown transistor T1, the voltage of the internal node (1) rises and causes the transfer gate threshold

voltage to rise thus reducing the sense current and slowing down data read out.

The back gate bias effect is more severe for the N-channel transfer gate than for the P-channel transfer gate because the bulk threshold coefficient Gamma is proportional to the root of the substrate doping concentration.

$$|V_T| = |V_{T0}| + \gamma \left[\sqrt{2\phi_B + V_{S\text{-}SUB}} - \sqrt{2\phi_B} \right]$$

$$\text{where} \quad \gamma = \frac{\sqrt{2q\epsilon_0\epsilon_{Si} N_{SUB}}}{C_{ox}}$$

For a P-well CMOS process, the N-channel transistors are built in the P-well. In order to control the native threshold voltages of the N-channel and the P-channel transistors, the P-well doping concentration is typically one order of magnitude higher than the N-substrate doping concentration. Hence the magnitude of Gamma for N-channel transistors is about 3 times greater than for P-channel transistors.

Akiya [Akiy79] selected the P-channel transfer gate over the N-channel gate to reduce threshold voltage shift. However the use of P-channel transfer gates consumes more circuit area because the hole mobility is about half the electron mobility in bulk silicon. In the design of the colour palette SRAM cell, the N-channel transfer gate is chosen in order to minimize circuit area. The speed and stability of the memory circuit was simulated by SPICE to guarantee that the memory met design specification.

5.4.2 Problems Related to Sense Current

The time to discharge a capacitor through an N-channel transistor is much shorter than for charging due to back gate bias effect [Craw67, Cobb70], hence in order to reduce read access time the bit line is usually precharged for a memory read. If the memory cell storage node was at VSS, then the bit line capacitance discharges through the transfer gate, through the internal node and then through the pulldown transistor to ground. The discharge current, called the sense current, is detected by the sense amplifier and relayed to the output drivers to forward the data to the external world.

The danger in a precharged read occurs when the sense current is large enough to raise the internal node voltage (which is the gate voltage of the other inverter in the cross-coupled flipflop) near to the point when the flipflop could reverse state [Hodg68, Frie68]. The state reversal occurs when the inverters are biased past the unity gain point into the high gain region shown in Figure 5.2.
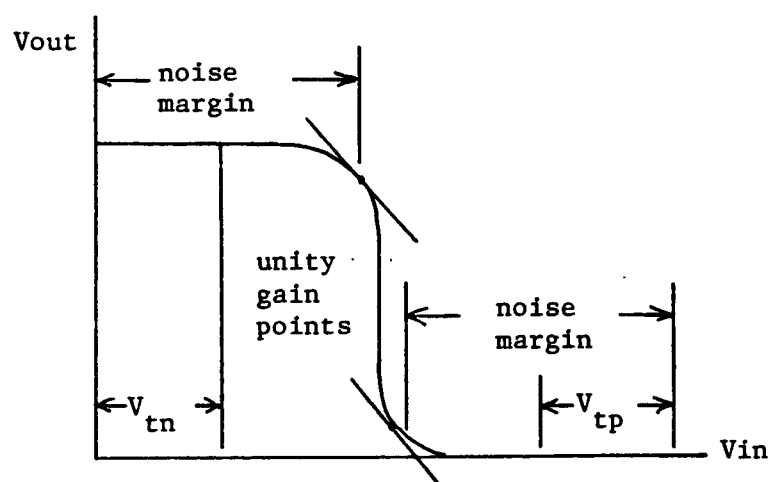
Figure 5.2 The D.C. transfer characteristics of an inverter

The internal node voltage rise is caused by (1) the VDS of the pulldown transistor which sinks the sense current, and (2) the charging of Cg2 shown in Figure 5.1 by part of the sense current. Because the bit line load capacitance is typically two orders of magnitude greater than the storage node capacitances Cg1, Cg2, the read out "noise" injected into the memory cell by the sense current presents a serious destabilizing effect. The read out noise tends to pull the memory cell to the opposite binary state, thus causing a soft error. The effect of read out noise is especially pronounced for the "single-ended" read out of the 5-transistor [Holl78] cell because the negative feedback supplied by the complementary bit line is absent.

The effect of sense current toward soft error was discussed by many authors [Hodg68, Frie68, Holl78, Akiy79]. The solution is to adjust the transconductance of the transfer gate relative to the pulldown such that the VDS of the pulldown transistor does not exceed the inverter unity gain point. The gate of the cross-coupled inverter is thus maintained at a sufficiently positive voltage to prevent the initiation of positive feedback in the bistable circuit which expedites the parasitic state reversal. A design conflict arises at this point because reduction of the transfer gate gain will increase the memory read and write time, while increase of the pulldown gain increases memory cell area.

5.4.3 Instability Related to Capacitive Cross-coupling

Noise due to capacitive cross-coupling, or crosstalk, is injected into the memory cell through the parasitic coupling capacitances Cgn, Cgb, and Cbn shown in Figure 5.3 [Frie68, Shic68].
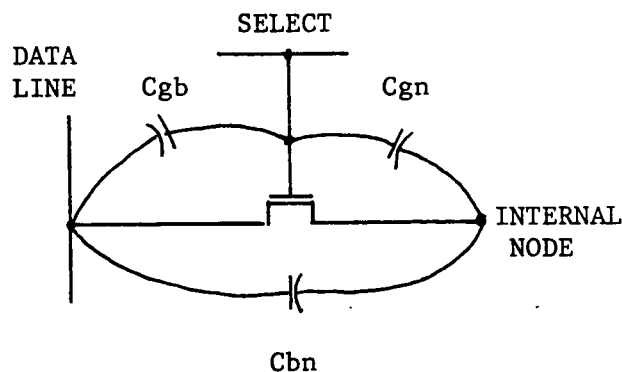
Figure 5.3

Parasitic coupling capacitances from the data
line and the select line to the internal node

The memory cell in this case is not accessed and remains deselected. When another cell in the same column which shares the same bitlines is read or when the bit line is precharged, however, the bit line voltage transient is coupled into the gate of the transfer gate and also into the internal storage node. If the signals propagating on the bit lines are opposite to those stored on the internal nodes, then both internal nodes will experience a "pull" to reverse state. To worsen matters, the crosstalk may partially turn on the transfer gates to further expose the internal nodes to bit line transient voltage changes. Care must be taken in circuit layout to reduce the stray capacitance between the select lines, the bit lines and the

internal storage nodes.

5.5 Simulation of Read Access Stability

Read access stability was simulated to investigate the effect that sense current injection had toward parasitic state reversal [Vlad80a, Vlad80b, Shic68]. The 5-transistor memory cell was simulated to help isolate the effect of the sense current without the negative feedback of the complementary read path. An N-channel transfer gate was used. The bit line was modeled by a 0.4pF load capacitance and was precharged to VDD to supply the sense current. Because the bit line was usually laid out in metal, bit line resistance was small in comparison to the transfer gate ON-resistance which was typically 10K ohms and thus was neglected in the simulation.

In the simulation, the internal read time of the memory cell was defined as the time which it requires for the precharged bit line to discharge from VDD to half VDD. All transistor channel lengths were set to the minimum value or 5 microns for the ISO-CMOS process used. Transistor transconductance was set by varying the gate width. Several practical combinations of transconductance values for the pullup, pulldown and the transfer gate were simulated.

State reversal occurs for all cases simulated where the transfer gate greatly exceeds the gain of the pulldown transistor. The circuit simulated is shown in Figure 5.4a. The circuit waveforms for a stable read and a state reversal are shown in Figures 5.4b and 5.4c respectively. The exact ratios of
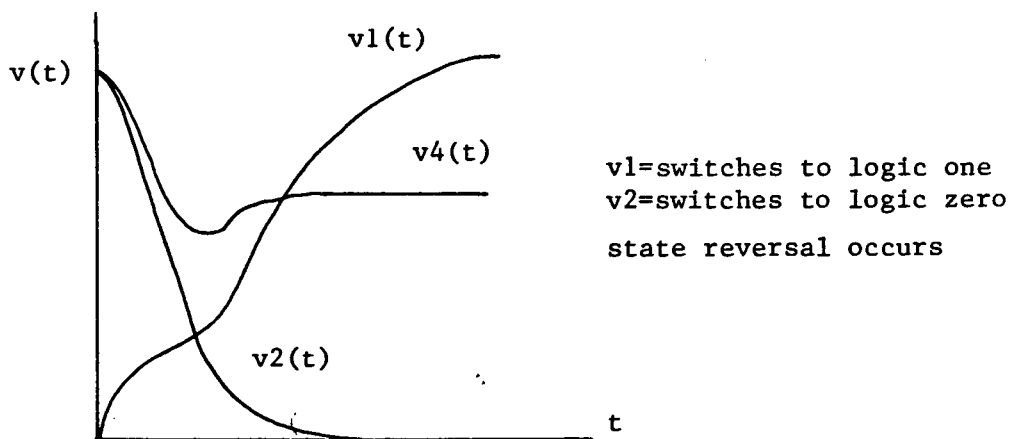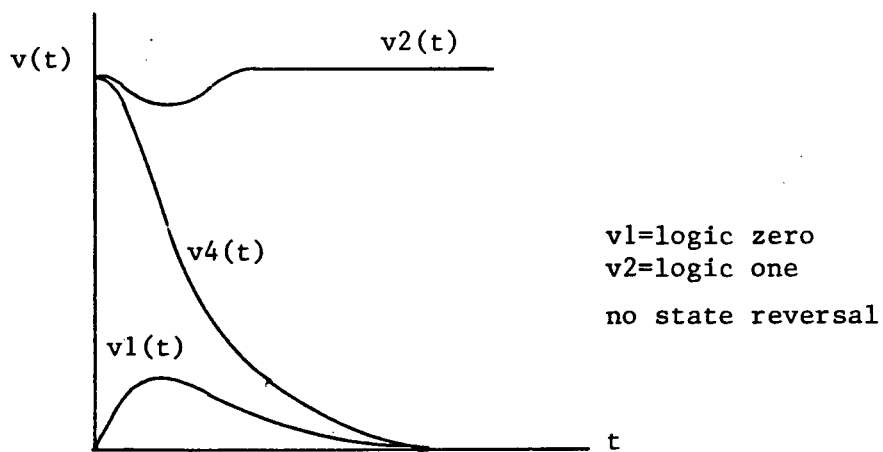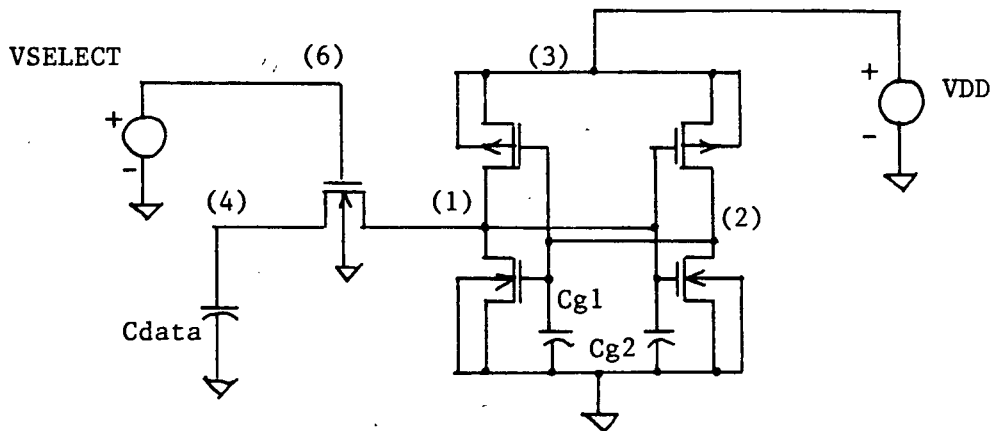
Figure 5.4

Stability simulations for the 5-transistor memory cell

transfer gate gain to pulldown gain which causes state reversal is a function of the gain of the pulldown. The maximum noise voltage for all cases was approximately at the inverter unity gain point of about 1.8 volts for VDD of 5 volts and for symmetric P and N-channel transistor native threshold voltages of magnitude of 0.5 volt. Internal read access times of less than 10 ns were predicted in all the circuits simulated.

Increases in the gain of the pulldown reduced the internal node noise voltage rise caused by the sense current as expected, the read time was also decreased. Analysis of the simulation data presented in Section 5.7 showed that transfer gate transconductance slightly smaller than the pulldown transconductance yielded an adequate safety margin for read out noise and yet maintains near minimum circuit area.

5.6 Analysis of Memory Cell Simulation

The SPICE simulation data indicated in general that read out noise susceptibility decreased with increasing area of the pulldown transistor. In order to choose the optimal circuit, the simulation data was analysed to determine the interrelation between the circuit area, access speed and noise susceptibility.

The analysis proceeded in four steps. First the sense current discharge circuit which consisted of the transfer gate and pulldown was modeled by a single transistor circuit to the first approximation using the Shichman and Hodges [Shic68, Vlad80a] model to summarize the combinations of transconductances needed to achieve circuit stability. Second,

the memory cell nominal circuit area was approximated by a function of the pulldown and transfer gate area with the assumption that the pullup transistor occupied only a negligibly small fraction of the static memory cell area. Third, a graph was constructed to unify the stability simulation data and the analysis result relating internal read time and nominal circuit area. Fourth, the optimal memory cell transistor gains were selected to minimize the nominal circuit area subject to a given internal read time.

5.6.1 Derivation of a Sense Current Discharge Model

The sense current discharge circuit is represented by the circuit of Figure 5.5b. It is desired to find the effective single transistor gate ratio (W/L)EFF which gives the same internal discharge time as the circuit of Figure 5.5a. Using the Shichman and Hodges model, the discharge time was solved in terms of the single transistor (W/L)EFF. The (W/L)EFF for each transfer gate and pulldown combination was then computed by substituting the simulated internal read access time.
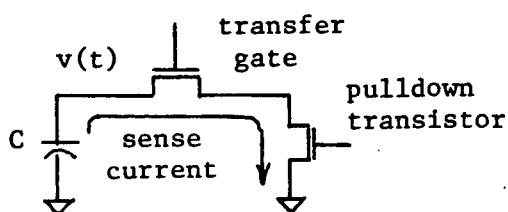
Figure 5.5a

Sense current discharge
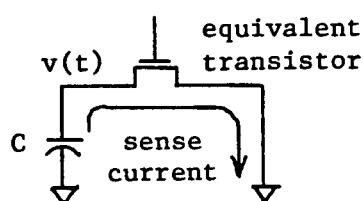
Figure 5.5b

Equivalent circuit circuit

The total time Ttot for v(t) to discharge from VDD to 0.5*VDD is

$$Ttot = Tsat + Ttriode$$

Tsat is the time when the transistor is in saturation

$$T_{sat} = \frac{\ln\left[\dfrac{\gamma V_{DD} + 1}{1 + \gamma(V_{GS} - V_T)}\right]}{\dfrac{\beta}{2 C_{load}} \gamma (V_{GS} - V_T)^2}$$

where γ is the back gate bias coefficient, β the transconductance, V_T the native threshold voltage and V_{GS} the gate-to-source voltage.

Ttriode is the time when the transistor is discharging in the triode region

$$T_{triode} = \frac{\cdot C_{load}}{\beta(V_{GS} - V_T)} \ln\left[\frac{2(V_{GS} - V_T) - \dfrac{V_{DD}}{2}}{\cdot \dfrac{V_{DD}}{2}}\right]$$

Thus (W/L)EFF is

$$(W/L)_{EFF} = \frac{C_{load}}{T_{tot} K'(V_{GS} - V_T)} \ln\left[\frac{2(V_{GS} - V_T) + \dfrac{V_{DD}}{2}}{\dfrac{V_{DD}}{2}}\right] + \frac{\ln\left[\dfrac{1 + \gamma V_{DD}}{1 + \gamma(V_{GS} - V_T)}\right]}{\dfrac{T_{tot}}{2 C_{load}} \gamma(V_{GS} - V_T)^2}$$

The discharge time Ttot is dominated by the time to discharge the bit line capacitance when the transfer gate is in the triode region. The credibility of the (W/L)EFF value was checked by representing the transfer and pulldown transistors by a naive model consisting of the transistor ON-resistances connected in series as shown in Figure 5.6.
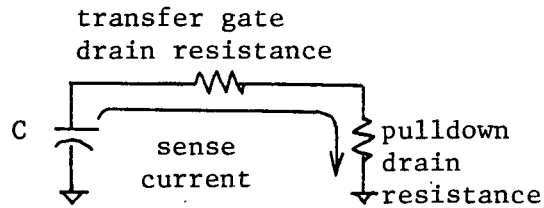
Figure 5.6   A resistance model for sense current discharge

A nominal (W/L)NOM value was defined as:

$$(W/L)NOM = \frac{(W/L)^{pulldown} * (W/L)^{transfer}}{(W/L)^{pulldown} + (W/L)^{transfer}}$$

The nominal (W/L)NOM value represented the apparent current sinking capability of the transfer gate and pulldown combination based only upon their respective transistor size. The effective value (W/L)EFF in contrast represented the effective, that is actual current sinking capability of the circuit. The (W/L)EFF is slightly smaller than the (W/L)NOM value for the same circuit because the nominal value did not account for the effect that sense current injection had on the memory circuit.

A plot of (W/L)EFF and (W/L)NOM versus the internal read time is shown in Figure 5.7. Note that the (W/L)NOM value deviates from the (W/L)EFF values and the decrease of read access time bottoms out due to the effect of the sense current which tends to cause the cell to pull to the opposite state. In fact, the increase in read access time of a simulated circuit after the bottoming out point indicates that the cell was approaching a state reversal.
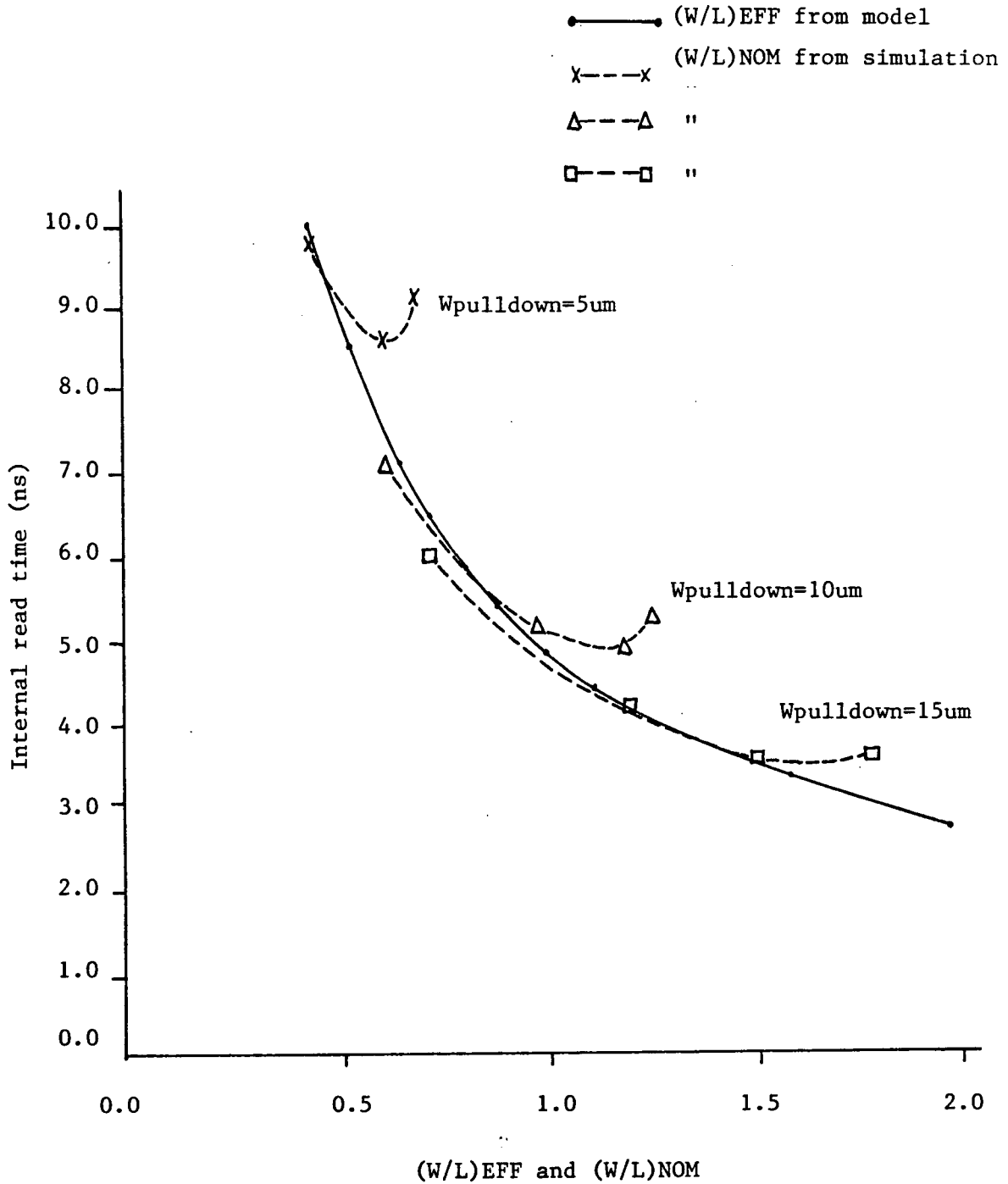
Figure 5.7

Plot of internal read time vs. (W/L)EFF and (W/L)NOM
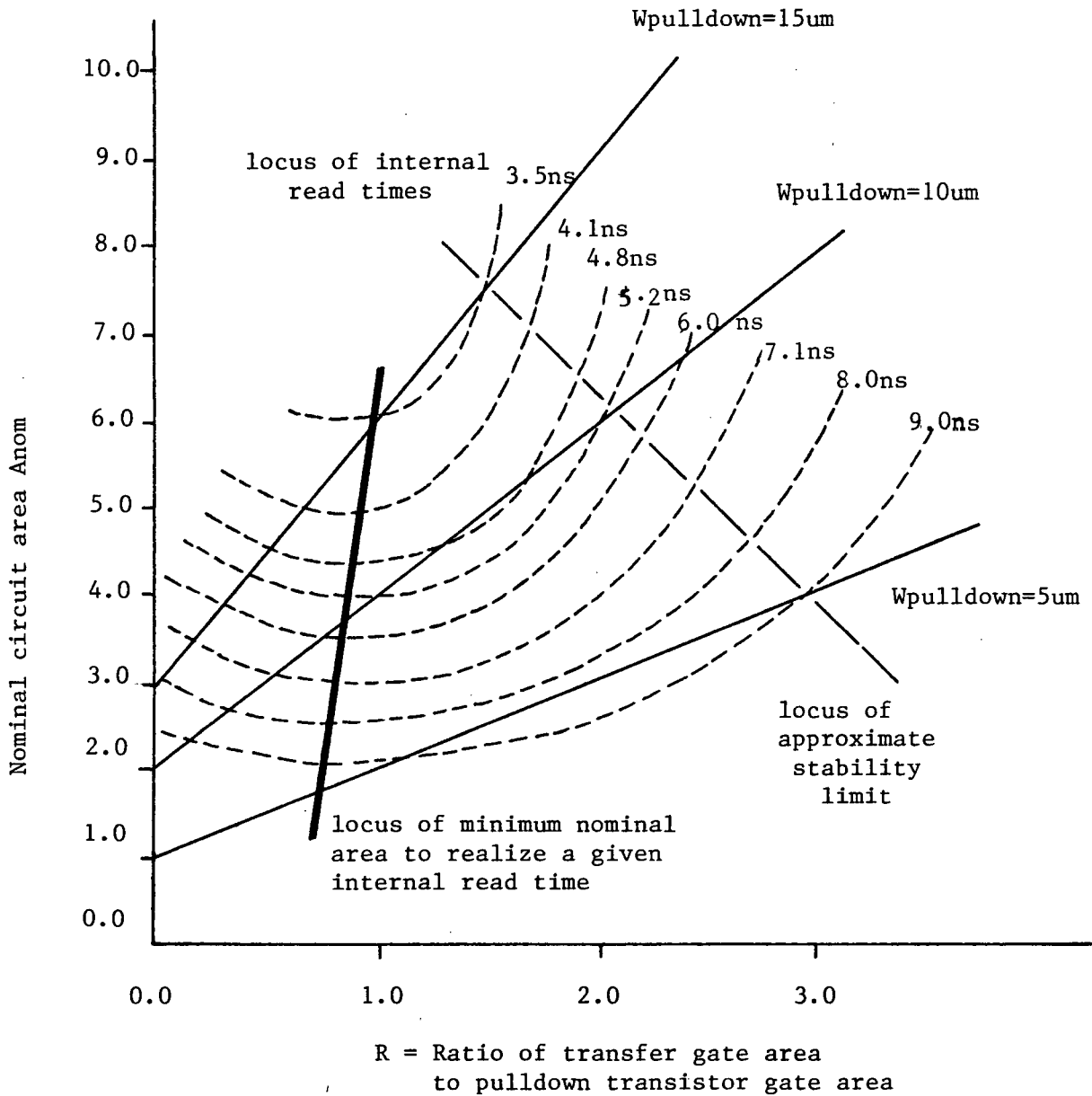
Figure 5.8

The unified design chart for memory cell design

5.6.2 Definition of Nominal Memory Circuit Area

Since the pullup transistor needs to supply only the leakage current for the internal node during storage, the pullup is small compared to the transfer gate or the pulldown. The memory cell circuit area Anom is expressed as a function of the pulldown and the transfer gate transistor size using the assumption that the area consumption of the pulldown is negligible. A further simplification of the circuit area is achieved by considering the layout area to be directly proportional to the transistor gate area. The nominal memory cell circuit area Anom is then defined as a function of the transfer gate and the pulldown gate size to obtain a nominal estimate of the memory cell circuit area needed to achieve a given internal read access time.

$$Anom = Apd * (1+R)$$

where

$$Apd = \frac{\text{pulldown gate area}}{\text{minimum gate area(5um x 5um)}}$$

$$R = \frac{\text{transfer gate area}}{\text{pulldown gate area}}$$

5.6.3 Memory Circuit Design Chart

In the chart of Figure 5.8, the results of the discharge model analysis, simulated circuit stability, the internal read time, and the nominal circuit area of the single port 6-transistor memory circuit are combined. The family of curves parametrized by the internal read time was obtained by

interpolation of the SPICE simulation data. The black data points obtained from simulation using pulldown transistor gate width of 5 um and gate widths of 5, 10, and 15 um were connected by thin black lines. It is seen that the pulldown gate size must be increased in order to achieve short internal read times. Of course the nominal circuit area increased with the size of the pulldown and the size of the transfer gate. The stability limits were established from simulation data. In terms of nominal circuit area, the most cost effective circuit was obtained when the transfer gate gain is slightly less than the pulldown transistor gate gain as shown by the dashed lines connecting the locus of minimum area for each internal read time curve.

5.6.4 Memory Circuit Selection

The 6-transistor cell internal read access time was simulated to ensure that the results obtained for the 5-transistor cell was applicable to the 6-transistor cell. Bulk threshold coefficients of 1.63 and 0.52 were used for the N and P-channel transistors to reflect the ISO-CMOS process. A slight increase in internal read time was observed due to the increased bulk threshold effects on the transfer gate, but on the whole the results were close to those obtained for the 5-transistor cell. The data read out for the 6-transistor cell was much more stable due to the presence of the second transfer gate and bit line which provided negative feedback to counter to destabilizing effects of the sense current.

The write operation of the 6-transistor memory cell was
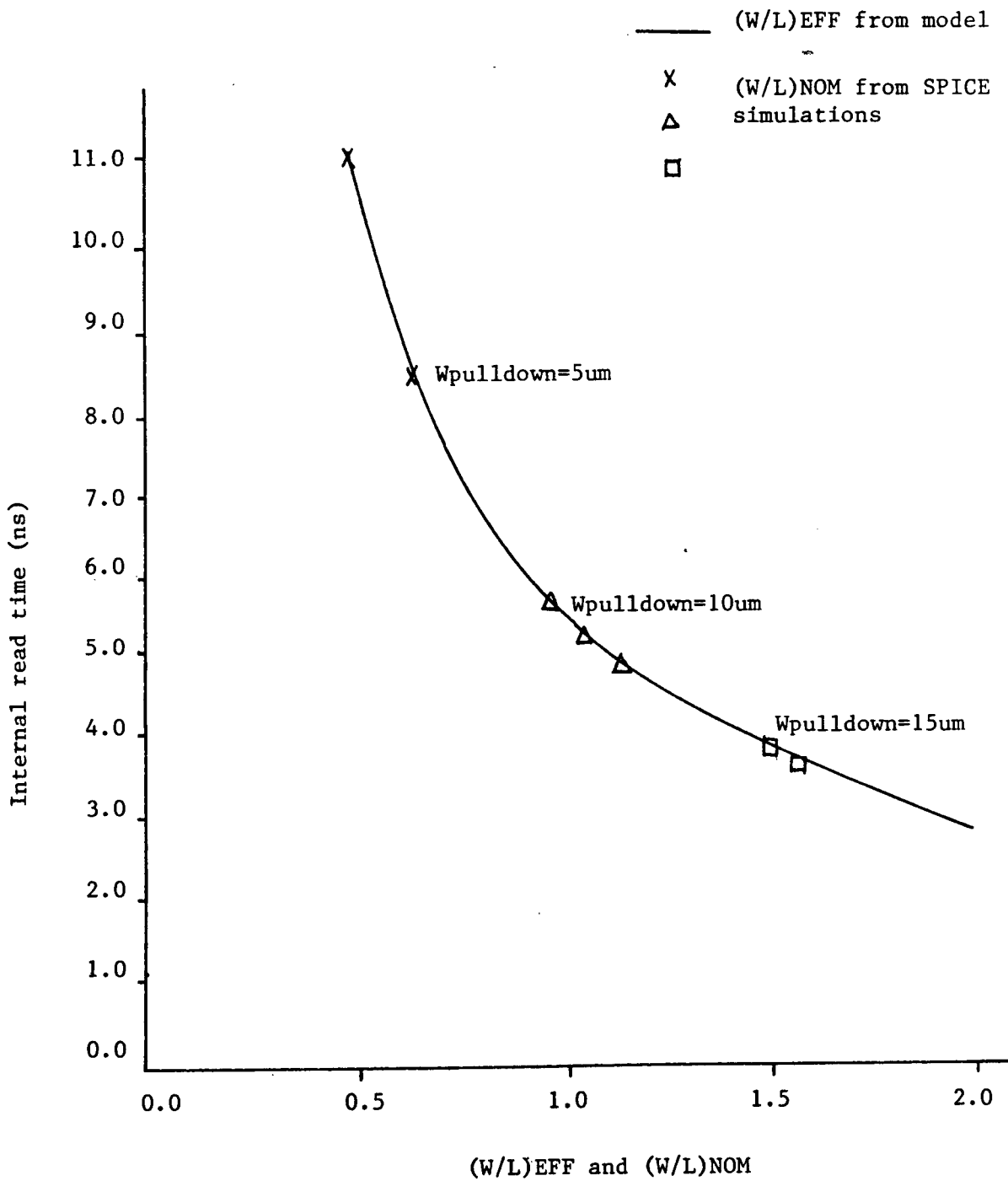
Figure 5.9

Plot of the internal read time vs. (W/L)EFF
and (W/L)NOM for the 6-transistor memory cell

simulated by applying voltage sources directly to the bit line external nodes. The memory cell was first placed into a stable state and then the worst case internal write time was simulated by forcing the memory cell into the opposite logic state. Circuits with pulldown transistor gate widths of 5, 10, and 15 um with gate length fixed at 5 um were simulated. The transfer gate sizes were maintained to be the same size as the pulldown to minimize the circuit area. The internal write time was defined as the time required for both internal nodes of the bistable circuit to settle to within 10 percent of the asymptotic voltage level. Simulation showed that the internal write operation was completed within 3 ns for all three circuits simulated. As expected, the internal write time was constrained by the time it required for the internal node to charge from VSS up to VDD. In general it is desirable to design the memory cell such that the internal read and write times are approximately equal. To this end, the circuits with pulldown and transfer gate widths of 10 to 15 microns were prefered.

It was now possible to select the optimal circuit for a 6-transistor memory circuit. Although the internal read time and write times did not necessarily reflect the external read or write times, the very short internal access times of approximately 10 ns for circuits (2) and (3) in Table 5.1 left a wide safety margin between the internal memory access times and the maximum external access times of about 100 ns for the video read out operation.

ckt	pulldown gate W	pullup gate W	transfer gate W	internal read time	internal write time	nominal ckt area
1	5 um	5 um	5 um	11.0 ns	2.7 ns	2.0
2	10 um	5 um	10 um	5.5 ns	2.2 ns	4.0
3	15 um	5 um	15 um	3.7 ns	2.1 ns	6.0

Table 5.1

Access times and nominal circuit area of
some practical 6-transistor memory cells

5.7 Design of the Dual Port Static Memory Circuit

The dual port CMOS SRAM circuit required for the colour palette IC was obtained by making some minor modifications to the optimal single port 6-transistor SRAM memory circuit. The basic circuit of the dual port memory cell is shown in Figure 5.10.
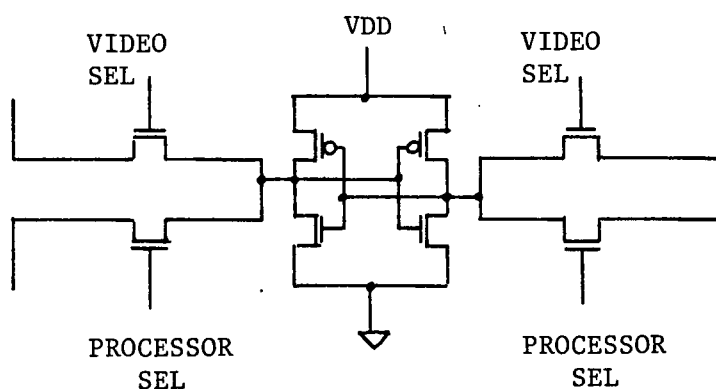
Figure 5.10

The dual port memory cell circuit

The dual port memory cell was obtained simply by adding an extra pair of transfer gates to implement the second port. The optimal transistor sizes computed for the single port 6-transistor cell could not be used directly in the dual port memory cell without considering the interaction between the two ports. In the operation of the colour palette, the processor had both read and write access to the memory array while the video circuit has only read access to the memory. Further, since the video and processor-to-memory accesses occur asynchronously, a conflict may arise when the processor writes to a memory location which is simultaneously read by the video circuit. Table 5.2 shows the possible modes of operation for the dual port memory cell.

mode	processor write	processor read	video read	comment
1	no	no	no	no operation
2	no	no	yes	video read only
3	no	yes	no	processor read only
4	no	yes	yes	dual port read
5	yes	no	no	proessor write only
6	yes	no	yes	dual port write and read
7	yes	yes	no	impossible
8	yes	yes	yes	impossible

Table 5.2

Modes of operation for the dual port memory cell

The boundary conditions for the dual port memory cell circuit occurs when (1) the processor read and video read access overlap; and (2) when the processor write and video read

accesses overlap. In case (1), the combined sense current flowing into the VSS internal node must be considered to ensure memory read stability. In case (2), the extra loading presented by relatively large video bit line capacitance must be considered to avoid an incomplete state reversal during a memory write. Extension of the single port memory cell result indicated that given a pulldown size, the size of the transfer gates of the two ports should satisfy

$$(W/L)EFF = (W/L)_{PORT1}^{EFF} + (W/L)_{PORT2}^{EFF}$$

Dual port memory circuits with various combinations pulldown, processor port transfer gate and video port transfer gate size were simulated. The P-channel pullup was fixed at minimum gate size of 5 um by 5 um. The combined video and processor transfer gate gain was set equal to the pulldown gain to maintain good stability and optimal circuit. The single read access, single write access, the dual read access and simultaneous read and write access times were in the neighborhood of 10 to 30 ns.

5.9 Layout for the Single Port Memory Circuit

The layout of a single port 6-transistor SRAM cell is shown in Figure 5.11. The power grid and data (bit) lines are laid out in metal and are run vertically to minimize voltage drop and propagation delay along the length of the memory array. The word select is wired in polysilicon and runs orthogonal to the data lines to minimize the coupling capacitance between the two

buses. Also since the select line gates the memory cell transfer gate therefore the select lines are best laid out in polysilicon. For a buld CMOS process, typical metal and polysilicon-to-bulk capacitances are 2.5e-5 and 3.5e-4 pF/sq.um. Typical sheet resistances for metal and polysilicon are 0.03 and 30.0 Ohms/square. The capacitance of metal and polysilicon with respect to substrate were comparable, however the sheet resistance of polysilicon is several orders of magnitude greater than that of metal. If the propagation delay along a bus line were represented by a lumped RC model, then the time constant of polysilicon is about 4 orders of magnitude greater than that of metal. The large polysilicon RC time constant could cause signal skew in high speed circuits where long polysilicon interconnects are used.

In the interest of reducing the polysilicon select signal bus length, the memory cell should have a large height to width ratio. In order to reduce stray capacitance, a height-to-width ratio of 1 to 2 is more appropriate. In the layout of the single port memory cell shown in Figure 5.11, the P-channel pullup transistor is 5 by 5 um; the pulldown and the transfer gate both has channel a width of 15 um and a gate length of 5 um.

The single port memory cell shown in Figure 5.11 has a circuit area of 972 square lambdas, with lambda equal 2.5 micron. In comparison, the dual port memory cell used in the colour palette prototype consumed 8580 square lambdas per bit of storage. Assuming that the new dual port memory cell circuit area was twice that of the single port memory cell, that is having a circuit area of approximately 2000 square lambdas, the

VSS   DATA   VDD $\overline{\text{DATA}}$ VSS

27
lambdas

36
lambdas

active   pwell p^+ polysilicon   contact n^+ metal
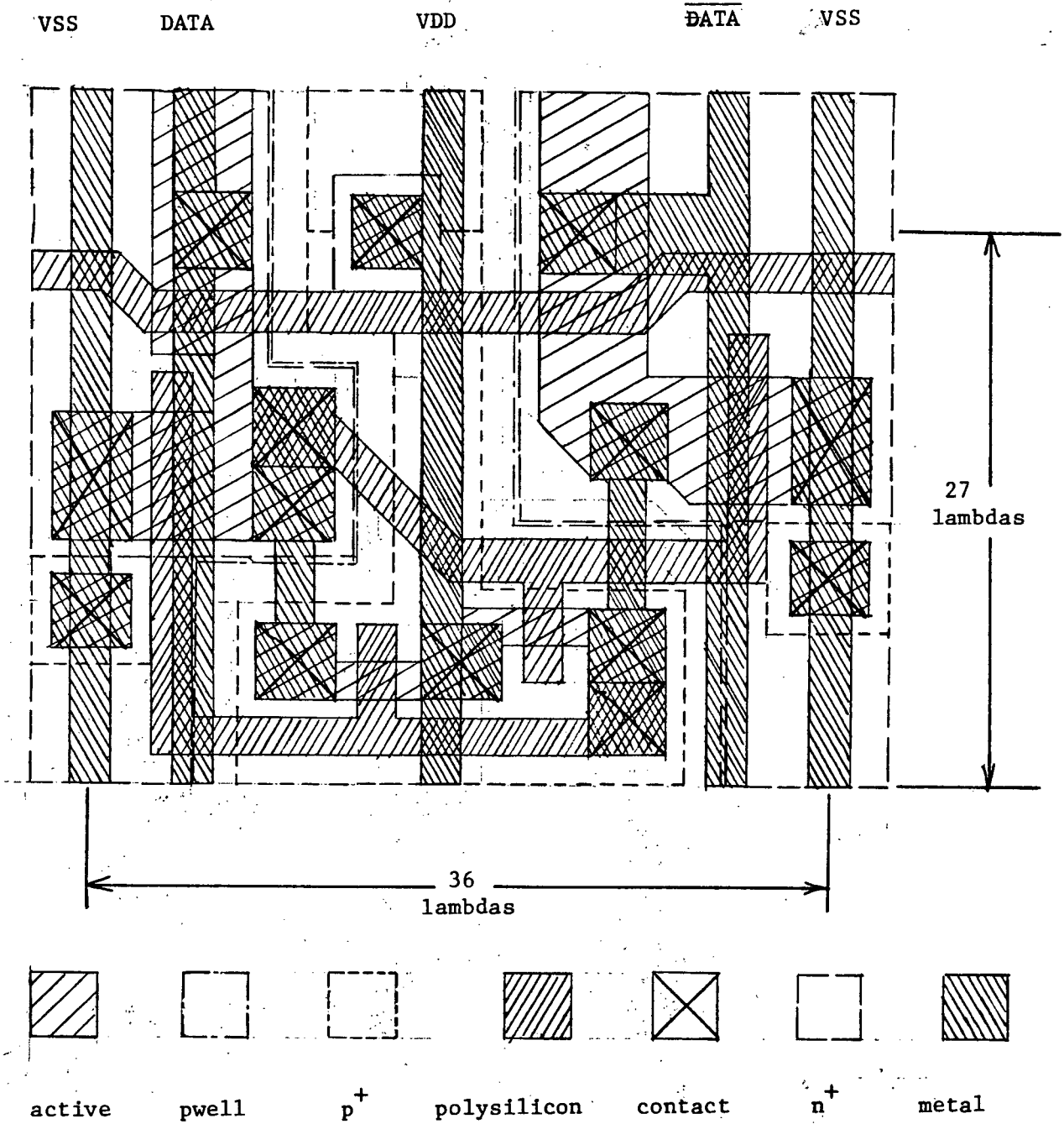
Figure 5.11

Layout of a single port CMOS static RAM cell

memory array circuit area would be reduced by about 75 percent. The full scale colour palette chip die size would then be reduced by about 50 percent.

6.0 CONCLUSION

The practicality of the simplified design method and algorithmic layout tools to support custom LSI design was evaluated in light of the complexity of the IC designed, the capability of the hardware and software design tools, and the work required to complete the designs.

Two prototypes of the colour palette custom IC were designed. The scaled down prototype contained almost all the functional circuit modules of the full scale colour palette. The prototype contained approximately 800 transistors. The full scale prototype made use of most of the pre-designed cells and had a transistor count of approximately 6000 and die size of 0.4 by 0.6 cm. Simplified electrical and layout design rules, as well as speed and and loading estimation rules were followed in the design work. With the exception the dynamic DAC used in the prototype, all circuits were static and complementary. The choice of static and complementary circuits simplified the design task at the expense of layout area. In light of our modest IC design experience and the time available, this simplification was necessary to maximize the probability of obtaining a functional IC. As a result, the prototype circuits were indeed functional but they consumed layout area extravagantly. The circuits, especially those which critically determine the overall processing speed of the IC should be designed carefully with the help of circuit simulation to achieve accurate performance characteristics and minimal circuit area. The colour map memory cell was redesigned which was projected to reduce the IC area by about 50 percent.

The set of Microtel Pacific Research algorithmic layout tools used in the design provided the basic geometric primitives for layout work. Cell instantiation, node naming and node finding functions were provided to assist with the hierarchical construction of the IC. The algorithmic layout tools were programmed in HP-BASIC and executed on the HP9826 and HP9836 desk-top computers. Check plots were obtained through low resolution B/W screen displays as well as from the HP7220 eight colour pen plotter. SPICE was used to simulate critical paths in the circuit but no automatic design rule checking program or any other verification tools were available.

The colour palette prototype required two man-months to design and the full scale prototype about 2.5 man-months. Indeed it was possible to design fairly complicated ICs with relatively little effort. It was found that the algorithmic layout tools were very useful for the design of small "leaf" cells. Little time was spent in generating a cell once its circuit was laid out in detail on grid paper. The IC and its layout, therefore, must be planned carefully so that the circuit can be constructed by a small collection of simple cells in a hierarchical manner, and such that the layout of complicated circuits on unmanageably large sheets of paper is avoided. A hierarchical design approach helped greatly to reduce design rule errors. This design strategy is similar to the concept of structured or modular computer programming. The analogy, however, is incomplete because unlike the logical integration of program procedures, the assembly of cells of an IC layout required the interconnection of cells globally over a large circuit area. No

provision is available in the tools used to assist in cell placement and routing to minimize layout area and layout rule violations, which are both important toward increasing the cost-effectiveness of the IC. More than half the total design effort was put toward global placement and interconnection as the result of the lack of such tools. Despite its shortcomings, the set of algorithmic layout programs did provide affordable and workable tools in the design of the colour palette prototype. It may be concluded from this work that large scale integrated circuits may be efficiently designed following a simplified and algorithmic design approach. There is, however, some possibility of improvement in design tools.

The rapid increase in cost-effectiveness of microcomputers and high resolution colour graphics display system today opens the possibility of integrating both algorithmic and interactive graphics on microcomputer-based workstations for IC design. Further, the increasingly sophisticated microcomputer operating systems could be used to multi-task circit design, layout, verification and documentation simultaneously. Numerically intensive simulations could be off-loaded to a mainframe computer to avoid compute-bounding the response of the microcomputer. IC design can thus be supported effectively by a mainframe computer serving a network of satellite workstations each with some significant local processing capability. The problem of how best to combine the advantages of algorithmic and interactive graphics for IC design still exists. From a technical standpoint, this is a problem in user and device interface. Alternatively, from a conceptual standpoint, the

interface of the graphics and algorithmic design approach points to the many levels of data representation and translation involved in IC design. This complication touches upon spatial and logical data processing in particular, and upon information representation and manipulation in general.

Bibliography

[Abot73]   R.A. Abbott, W.A. Regitz, and J.A. Karp,
           A 4K MOS dynamic random-access memory,
           IEEE Journal of Solid State Circuits, SC-8,
           October 1973, 292-310.

[Akiy79]   M. Akiya, and M. Ohara,
           New input/output designs for high-speed
           static CMOS RAM,
           IEEE Journal of Solid State Circuits, SC-14,
           October 1979, 823-828.

[Boll73]   H.J. Boll, and W.T. Lynch,
           Design of a high performance 1024-bit switched
           capacitor p-channel IGFET memory chip,
           IEEE Journal of Solid State Circuits, SC-8,
           October 1973, 310-318.

[Brig78]   G.R. Briggs, et al.,
           40-MHz CMOS-on-sapphire microprocessor,
           IEEE Transactions on Electron Devices, ED-25,
           August 1978, 952-958.

[Burn64]   J.R. Burns,
           Switching response of complementary-symmetric
           MOS transistor logic circuits,
           RCA Review, 25, December 1964, 627-661.

[Carr72]   W.N. Carr, and J.P. Mize,
           MOS/LSI Design and Application,
           New York, McGraw-Hill Co., 1972.

[Chat79]   P.K. Chatterjee, G.W. Taylor,
           A.F. Tasch,Jr., and H.S. Fri,
           Leakage studies in high-density dynamic
           MOS memory devices,
           IEEE Electron Devices, ED-26, April 1979, 564-575.

[Cher69]   G. Cheroff, D.L. Critchlow,
           R.H. Dennard, and L.M. Terman,
           IGFET circuit performance - n-channel versus
           p-channel,
           IEEE Journal of Solid State Circuits, SC-4,
           October 1969, 267-271.

[Cobb70]   R.S.C. Cobbold,
           Theory and Application of Field-Effect Transistors,
           New York, Wiley-Interscience, 1970.

[Craw67]   R.H. Crawford,
           MOSFET in Circuit Design,
           New York, McGraw-Hill Co., 1967.

[Dool80]   D.J. Dooley, Ed.,
           Data Conversion Integrated Circuits,
           New York, IEEE Press, 1980.

[Elma81]   M.I. Elmasry, Ed.,
           Digital MOS Integrated Circuits,
           New York, IEEE Press, 1981.

[Flem83]   J. Flemming, and W. Frezza,
           NAPLPS: a new standard for text and graphics,
           BYTE, 8, February 1983, 203-254.

[Frie68]   J.H. Friedrich,
           A coincident-select MOS storage array,
           IEEE Journal of Solid State Circuits, SC-3,
           September, 1968, 280-285.

[Garr81]   P.H. Garrett,
           Analog I/O Desgin - Acquisition, Conversion, Recovery,
           New York, Reston Publishing Company, 1981.

[Greb72]   A.B. Grebene,
           Analog Integrated Circuit Design,
           New York, Van Nostrand Reihhold Co., 1972.

[Hodg68]   D.A. Hodges,
           Large capacity semiconductor memory,
           Proc. IEEE, 56, July 1968, 1148-1162.

[Holl78]   R.J. Hollingsworth, A.C. Ipri, and C.S. Kim,
           A CMOS/SOS 4K static RAM,
           IEEE Journal of Solid State Circuits, SC-13,
           October 1978, 664-668.

[Hon79]   R.W. Hon, and C.H. Sequin,
           A Guide to LSI Implementation,
           Tech. Rep., SSL-79-7, Xerox Palo Alto
           Research Centre, 1979.

[Howe81]   M.J. Howes, and D.V. Morgan, Eds.,
            Large Scale Integration, Devices, Circuits, and
            Systems,
            New York, John Wiley and Sons Publishing Co., 1981.

[Isob81]   M. Isobe, et al.,
            An 18-ns CMOS/SOS 4K static RAM,
            IEEE Journal of Solid State Circuits, SC-16,
            October 1981, 460-465.

[Klei69]   T. Klein,
            Technology and performance of integrated complementary
            MOS circuits,
            IEEE Journal of Solid State Circuits, SC-4,
            June 1969, 122-130.

[Mead80]   C. Mead, and L. Conway,
            Introduction to VLSI Systems,
            New York, Addison-Wesley Publishing Co., 1980.

[NAPL81]   North American Presentation Level Protocol Syntax,
            American National Standards Institute and
            Canadian Standards Association,
            Document No. BSRX 3.110-198X, 1981.

[Shic68]   H. Shichman, and D.A. Hodges,
            Modeling and simulation of insulated-gate
            field-effect transistors switching crcuits,
            IEEE Journal of Solid State Circuits, SC-3,
            September 1968, 285-289.

[Spec83]   Special Issue on VLSI Design: Problems and Tools,
            Proceedings of the IEEE, 71, January 1983.

[Stew77]   R. Stewart,
            High density CMOS ROM,
            IEEE Journal of Solid State Circuits, SC-12,
            October 1977, 503-506.

[Susu73]   Y. Susuki, K. Odagawa, and T. Abe,
            Clocked CMOS calculator circuitry,
            IEEE Journal of Solid State Circuits, SC-8,
            December 1973, 462-469.

[Taru69]   Y. Tarui, et al.,
            A 40-ns 144-bit n-channel MOS-LSI memory,
            IEEE Journal of Solid State Circuits, SC-4,

October 1969, 271-279.

[Tori78]   Y. Torimaru, K. Miyano, and H. Tokeuchi,
           DSA 4K static RAM,
           IEEE Journal of Solid State Circuits, SC-13,
           October 1978, 647-650.

[Vlad80a] A. Vladimirescu, and S. Liu,
           The simulation of MOS integrated circuits using
           SPICE2,
           Tech.Rep., UCB/ERL M80/7, Department of Electrical
           Engineering, University of California at Berkeley,
           February 1980.

[Vlad80b] A. Vladimirescu, A.R. Newton, and P.O. Paderson,
           SPICE Version 2G.1 User's Guide,
           Department of Electrical Engineering,
           University of California at Berkeley, October 1980.