PLA Design in NAND Structure

Chong Ming Lin

Semiconductor Engineering Group

Digital Equipment Corporation
75 Reed Road
Hudson, MA 01749 (617) 568-4888

ABSTRACT--A NAND (serial gating) structure PLA of the MOS poly-silicon gate process has been developed for high density and medium fast speed VLSI application. Dynamic clocking is used for minimum power dissipation and elimination of the ratio problem associated with static NAND gate. Ion-implantation for memory cell programming and the elimination of contact in the memory area drastically reduces the cell size, and reliability is improved. A simple but effective self-timed clocking scheme is employed for better operating margins against process variations; the overhead chip area for the clock generation is sufficiently small. The advantages of allowing metal signal and power lines to cross the PLA memory area is discussed. Some measured data from a 3.5μm NMOS Si-gate process with regard to gate height and transistor sizes are also described.

INTRODUCTION

In MOS circuit design, the NAND circuit, due to its inherited electrical characteristics, has been restricted in application for only a limited number of inputs. In the past, ion implant programmed NAND structures had only been used for very slow speed ROM applications [1] [2]. However, with the fast progress in process technology, a properly designed NAND structure PLA is becoming more attractive for some of the existing applications. New product of a new structure with a new process is always impressive for its performance, but its cost effectiveness is not guaranteed on production level. This question was better stated by G. Moore in his lecture, in the 1st Caltech VLSI Conference, January 1980:

> ". . . the semiconductor industry is not now process--technology limited for non-memory product. How to best make use of the processing technology is really what the problem is."

PLA Design in NAND Structure

The experiment discussed in this report was done primarily as an answer to a request, in November 1978, for a high density and medium fast speed PLA design. In order to make best use of a then newly developed 3.5 μm process and fully utilize the given timing spec of the speed requirement, a NAND structure PLA was proposed for better performance which was also cost effective for an existing application and beyond.

In order to achieve overall low chip size and a high operating margin; a dynamic, self-timed clocking scheme was proposed for most of the circuitries. Knowledge from measured data are used to construct circuits with better performance than the original approaches used on a circuit test chip [3].

Since the NAND and the NOR structures are the two basic building blocks and complementary to each other in MOS circuit design, the author feels that the study is also useful for understanding the general NOR type PLA design as an expected by-product.

In the following sections, process background, NAND circuit model, cell layout, reliability and design consideration will be described. Further improvement will also be discussed.

PROCESS BACKGROUND

Although the process performance is crucial in the evaluation of a new circuit structure, this information was not available in the previous papers [1] [2]. In order to test out some of the assumptions and limitations of the NAND structure PLA, a test chip was designed and manufactured in 1979 with a 3.5 μm NMOS process, P400. This process uses ion implantation for Source/Drain, plasma dry etching, silicon doped aluminum; plus 4 types of devices, as shown in Table I, which provide more flexibility in circuit design and better powerXspeedXdensity product than many of the processes in the previous generation.

The performance of the process was measured through a ring-oscillator built on the test chip. As shown in Fig. 1(c), the powerXspeed product curves indicate that in a typical case, the performance of that ring-oscillator is around 0.35 pj at 5V VCC, -3V VBB, and RT. Those curves also indicate that the process parameters are being optimized against those skewed process corners. With this kind of performance, the NAND PLA does have better potential for some applications which were not feasible by processes of the previous generations.

TABLE I

Device threshold voltages and ion-implantation definition of the P400 process.

	Boron	Arsenic	Vth (typical)
Enhancement	x	--	0.7v
Intrinsic	--	--	0.0v
Depletion-1	x	--	-1.2v
Depletion-2	--	x	-2.5v

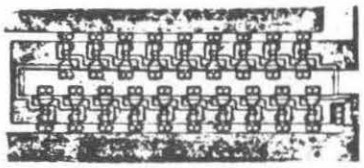


Fig.1a - Photomicrograph of a ring-oscillator with 19 stages and fan-in and fan-out of 1.

Vert. 1v/Div. Hori. 10ns/Div.

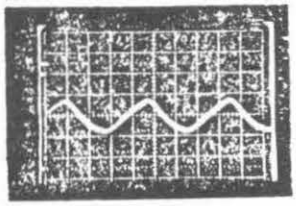


Fig. 1b - The wave form of the ring-oscillator, measured at 23 C, 5V VCC, and -3V VBB.

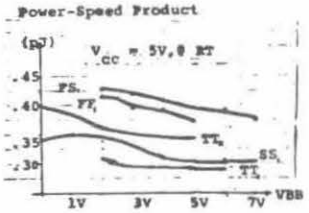


Fig. 1c - Performance of the ring-oscillator at 23 C, 5V VCC.

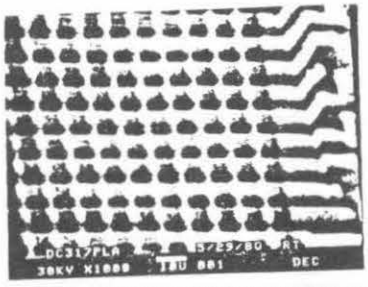


Fig. 2a SEM microphotograph of the NAND PLA array

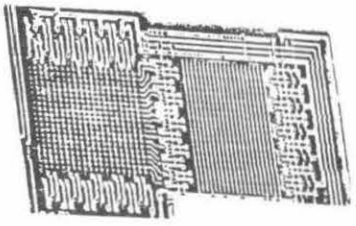


Fig.2d - A 20x20x20 NAND PLA with 8μm/4μm devices.

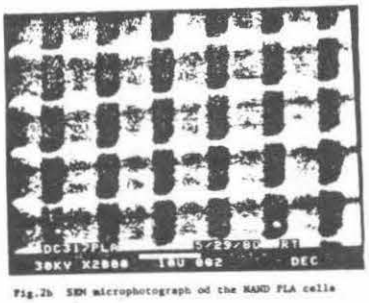


Fig.2b SEM microphotograph od the NAND PLA cells

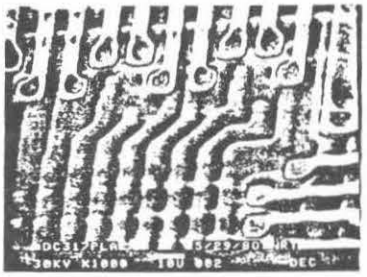


Fig. 2c - An example of the difficulty in interfacing to a small cell pitch of the NAND PLA

TABLE II

Basic Characteristics of NAND Circuit in Comparison with NOR Circuit [4]

Electrical Parameters	NOR	NAND
Logic Threshold *	$\frac{V_{DD}}{\sqrt{\frac{L_{pu}/W_{pu}}{L_{pd}/W_{pd}}}}$	$\frac{V_{DD}}{\sqrt{\frac{L_{pu}/W_{pu}}{n \cdot L_{pd}/W_{pd}}}}$
Delay Time	$T_{NOR} \approx T_{INV}$	$T_{NAND} \geq n \cdot T_{INV}$
Pull-Down Device Width (to achieve same 'VOL')	$W_{NOR} = W_{INV}$ **	$W_{NAND} = n \cdot W_{INV}$

* Assume same size for all pull-down devices.

** Punch-thru is not a problem.

An Illustration for Minimum NAND Device Channel Width Needed for an Acceptable VOL

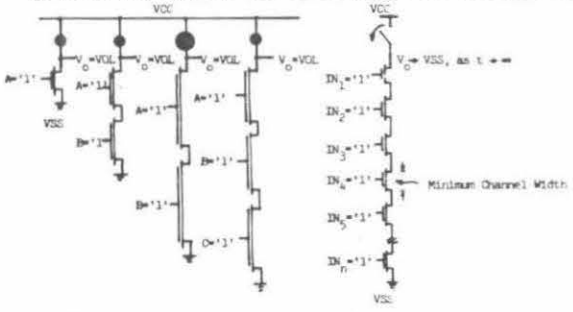
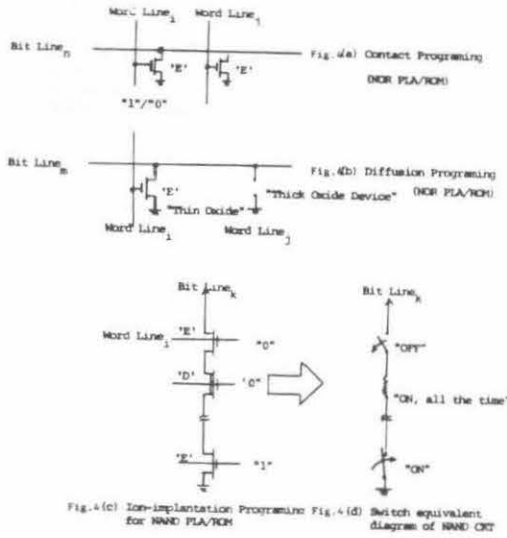
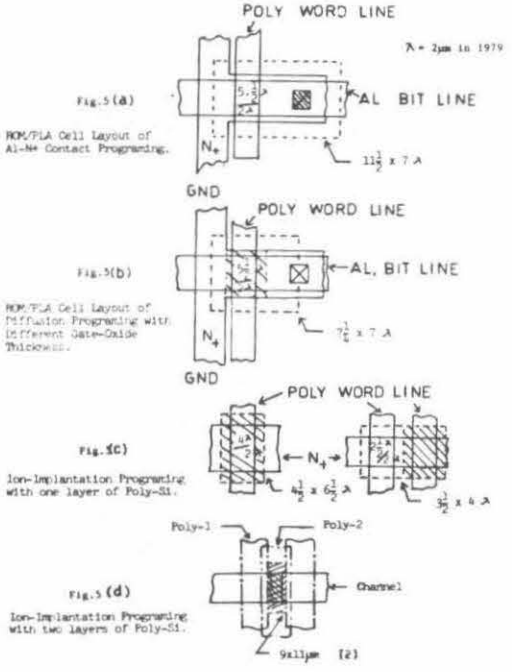


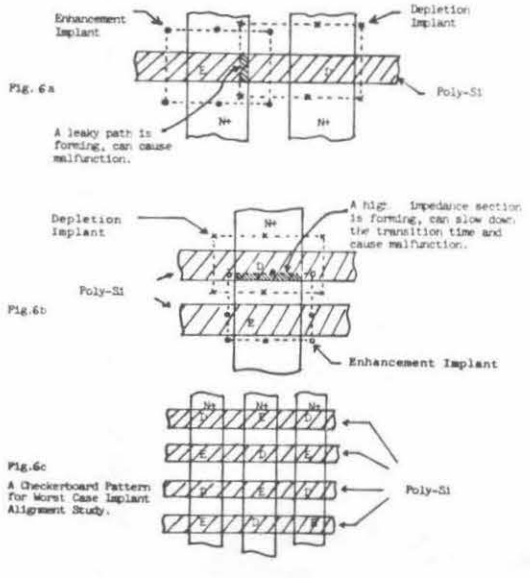
Fig. 3a NAND Circuits with Static Pull-up

Fig. 3b NAND Circuit with Dynamic Pull-up





Reliability Study of the "Implant Programing" Cell



PLA Design in NAND Structure

MODEL OF NAND CIRCUIT

As shown in Table II, the major reasons that the NAND circuit has not been used as popularly as the NOR structure are due to its inferior electrical properties in comparison with NOR circuit [4]. Although NAND circuit's d.c. characteristics, such as its logic threshold and pull-down device channel width, shown in Fig. 3(a), can be improved by using dynamic pull-up device, shown in Fig. 3(b). However, the delay time of a NAND circuit is slow and proportional to its number of transistors in series. Even so, a NAND circuit can use ion implantation as a programming method for its memory bits, as shown in Fig. 4, and this programming approach leads to the smallest PLA/ROM cell size that can be achieved by the current MOS technology.

CELL LAYOUT

For a conventional NOR Structure PLA/ROM Cell, as shown in Fig. 5(a), the size of a cell is defined and limited by its components--word line (poly), bit line (aluminum), contact between drain (N+) and bit line, memory transistor gate area, drain, and source (N+). Furthermore, due to process requirement, minimum area and space for each element must be used in the implementation of the memory cell. Thus, within the limitation of the present process technology, selection and/or elimination a certain part of the memory cell leads to different structure variations and cell sizes in PLA/ROM cell design. Table III shows the size ratio and then the basic properties of the four major types of cells which are well known to the public with straight forward layout techniques. From Fig. 5, it is obvious that the NAND structured PLA/ROM can be made up to a quarter of the size of a 'contact programming' cell, and the elimination of the contact also enhances the circuit reliability.

The size of the NAND Structure PLA/ROM cell can be made smaller if the process can provide smaller poly and N+ lines without causing electrical problems [3]. Furthermore, the alignment and resolution of the ion implant process hold the last barrier on the minimum size this approach can be.

When the NAND PLA/ROM cells are placed closer to each other, the enhancement and depletion implant can overlap into each other. Out of the four possible overlapping cases, there is one fatal case and the other three cases, although not fatal, can all cause electrical problems as shown in Fig. 6.

One way to test out the process and equipment limitations is using a checkerboard test pattern, shown in Fig. 6(c). This test pattern, if properly decoded, would be able to indicate the safty margins left in a process for implementing the NAND structure PLA/ROM. On the other hand, a NAND structured PLA/ROM is also a good test tool for process control monitoring, especially for ion implantation's definition control.

PLA Design in NAND Structure

CIRCUIT DESIGN CONSIDERATIONS

As shown in Table II, the NAND Structure PLA with static pull-up would have difficulties with ratio, pull-down device size, and slow speed in discharge against the constant conducting pull-up device. In dynamic operation, the ratio problem is eliminated, pull-down device size is minimized, and discharging time is reduced. However, the generation and implementation of the control clocks will complicate the design and require extra silicon area. As a result, in dynamic operation, the design effort and total implementation area for a PLA is more than putting two ROM arrays together. Furthermore, in NOR structure dynamic (or semi-dynamic) PLA design, an interface circuit is needed to allow precharging of the two arrays at the same time during the early cycle time of the operation. Consequently, a clock scheme of four phases generated from the system is the most common approach, and the safest way is using four non-overlapping clock phases to execute the operation at the expense of slow through-put time and zero process tracking capability.

For the proposed NAND PLA, the elimination of the interface circuit between the two arrays simplifies the layout work between the arrays and save area from implementing the interface circuit. Basically, a dynamic NAND circuit's access time is limited by its precharge and discharge time. In this proposed NAND PLA, precharge time is significantly reduced with precharge from both ends, because RC constant of the serial channel is halfed. The precharging devices of the AND array are driven by bootstrapped voltage level, which gives the AND array full VCC level that helps to speed up the discharge time in the OR array. The precharge operation can be further optimized by generating a longer pulse width with normal VCC level for the OR array, because the OR array will be enabled only after the AND array is settled. A high beta ratio input inverter in the output register with amplified positive feedback through an intrinsic device, as shown in Fig. 8(g), allows a weak logic "1" output from the OR array at VCC-VTn with no difficulty in the initial sensing and final stored level.

CLOCKING SCHEME AND GENERATION

The proposed NAND PLA uses five clock phases. Their wave forms are show in Fig. 7(a), where Cin, Cpr, Cena, and Ceno are used for the control of the PLA operation flow. While Cla is designed for the latch of the processed data against precharge and low frequency operation's leakage problem. It depends upon each design's system spec and circuit structure, Cla may be spared without effecting the performance.

TABLE III. ROM/PLA Cell Type and Sturecture

Type	No. Contact per Cell	Source Area shared by No. of Cell	Drain Area shared by No. of Cell	Programming Method	Cell Size Ratio to Type a. cel	Comments
a	1	≥ 2	≥ 1	Contact	1.0	NOR Structure Fast Turn Around Time
b	1/2	≥ 2	≥ 1	Diffusion	≈ 2/3	NOR Structure Thin Gate Oxide- Selected Thick - Dis-selected
c	#	≈ 2	≈ 2	Ion Implant	≈ 1/3	NAND Structure Enhancement Implant- Selected Depletion Implant- Don't Care
d	#	≈ 2	≈ 2	Ion Implant	≈ 1/3	NAND Structure Two layers of poly-Si, yield is less than single Poly process.

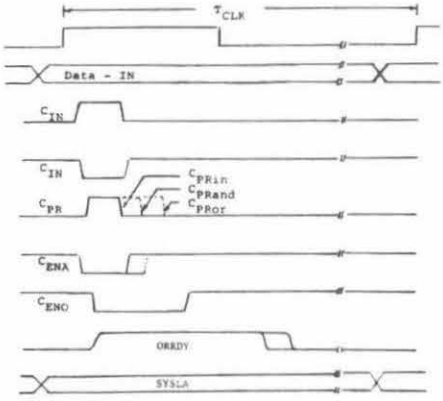


FIG. 7a CLOCK SCHEME AND WAVE FORMS FOR THE NAND PLA

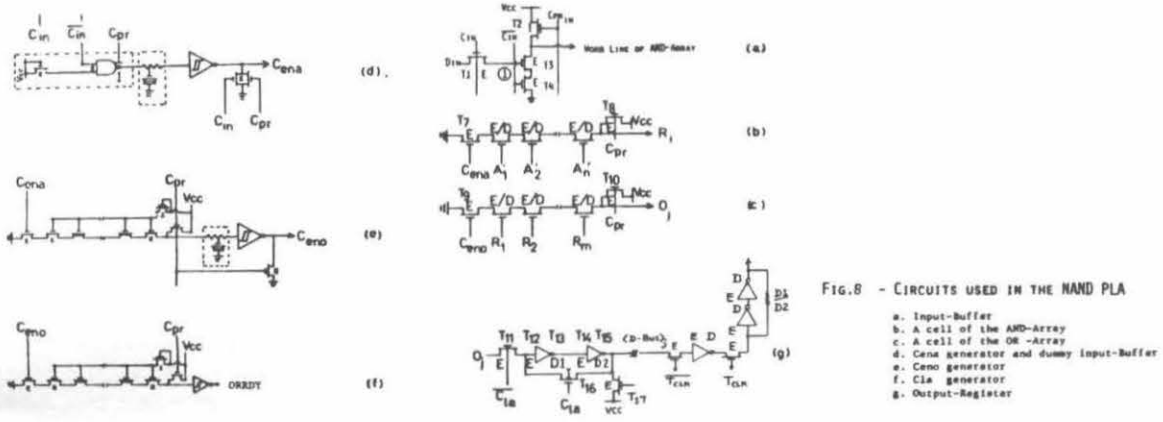


FIG.8 - CIRCUITS USED IN THE NAND PLA

a. Input-Buffer
b. A cell of the AND-Array
c. A cell of the OR -Array
d. Cena generator and dummy input-Buffer
e. Ceno generator
f. Cia generator
g. Output-Register

PLA Design in NAND Structure

All these clocks are derived from the rising edge of the input system clock, Tclk, and they are generated through dummy circuits which provide enough tracking capability against process variations and power supply changes. In order to use the same circuits or cells to achieve the best result for dummy circuits, the proper number of depletion devices are suggested to use for capacitance loading duplication, and device width effect {3} is another source to provide extra operation margin.

CIRCUIT OPERATION:

1. Array precharge--

Cpr is a self-timed clock pulse triggered by the rising edge of the system clock, Tclk. As shown in Fig. 9, both Cpr and Cin are generated through a dummy circuit which uses a row of the OR array to track the loading in OR array and uses a column of the AND array to track the completion of the precharge action. During this period, both Cena and Ceno are '0'. This allows both arrays to be fully charged to their highest possible level against couplings and the so-called "charge sharing" problem (3).

2. Data input--

New input data should be ready in the beginning of a new cycle. Those input data are strobed into the input buffer through T1, see Fig. 8(a), and temporarily stored at node ① after the Cin pulse is gone. During the precharge time, T4 is turned 'off' by $\overline{\text{Cin}}$. This guarantees the completion of the precharge to be independent of those input data--such that either T3 is 'on' or 'off', there is no d.c. path through T2 to ground due to the exclusive wave forms between Cpr and $\overline{\text{Cin}}$. The input buffer only consists of 4 transistors. Transistor T2 is an intrinsic device which gives a higher output voltage than an enhancement device when Cpr is high, but it will not conduct much current, with proper choice of channel length, when Cpr is low. This structure allows optimal output level and minimum device sizes for the pull-down devices T3 and T4, as well as for T2 itself, partly because the intrinsic device has the highest mobility among the four types of devices available in this process. Also, due to the compactness of the structure and the lack of appreciable d.c. path in this input buffer, the interface problem is helped and power dissipation is greatly reduced.

3. Enabling of the arrays--

Once the arrays are fully charged, Cpr and then Cin go down to '0'. When Cin is '1', T4 in the dummy input buffer, INdmbf, turns 'on', and the output of the INdmbf, see Fig. 8(d), start to change from a precharged '1' to '0' due to the input is VCC. With enough depletion type capacitors on the output and a Schmitt trigger to sense the change, Cena is ensured to turn 'on' only after the completion of all the inverted input data have been transferred to their outputs, or word lines of the AND array.

PLA Design in NAND Structure

Cena enables the AND array to decode its inputs through its pre-programmed memory bits. Cena also enables a dummy circuit in the AND array to discharge from its precharged level to '0' and thus generate the Ceno pulse, as shown in Fig. 8(e), by the same principle as the Cena generation. With the starting of the Ceno clock pulse, NAND PLA is ready to send out its decoded results to the output register and outside buses.

4. Output Register--

As shown in Fig. 8(g), the output buffer contains 7 transistors with static pull-ups used in the register to provide easy data storage through amplified positive feedback. The advantage of precharging the output data bus lines is incorporated into the circuit design to save power and size in the output section. Because of the precharge from T17, T15 can be made small as D1 (light depletion) device for sustaining purpose. This also speeds up the discharge on the bus line if $\overline{\text{Dout}}$ is a '0'.

Even though the output section only contains a minimum number of transistors, the layout work to interface the OR array cell pitch to the output buffer is not an easy task. Techniques like: combining two buffers together; bringing out output from both ends of the OR array; or constructing the buffers at a distance and then connect the two parts through spread out N+, Poly, or metal lines, are up to the designer's choice for the best matching between the NAND PLA and system requirement.

5. Latch of the output register--

For a dynamic PLA without any sustaining pull-up device used in the arrays, maintaining Cena = Ceno = '1' to keep ORRDY at '0' is needed against noise and coupling. Furthermore, isolating the OR array precharged outputs from their output buffers is also crucial for operation at low frequency where leakage eventually will change a precharged '1' to a '0'. In this proposed NAND PLA, a simple but effective design is used, as shown in Fig. 7(b). The dummy circuit in the OR array generates a delayed "data is ready" signal, ORRDY, which tracks after the completion of the OR array transition through narrower device width transistors and/or a Schmitt trigger. The Cla clock is also controlled by a system latch signal, SYSLA, which happens only after the transition is over, see Fig. 7(b).

PLA Design in NAND Structure

TEST RESULTS

Since most of the circuits proposed for the NAND PLA in this paper are of dynamic operation, power dissipation is being optimized to a minimum. Thus, the major concerns left for this approach are speed related device width effect, gate height effect, circuit operating range, and effect of precharge methods [3]. These measured data are considered to be useful as a reference point in related applications with this kind of circuit structure.

FURTHER IMPROVEMENT

SPEED: If the negative '0' level can be generated from an external VBB power supply, a Depletion-1/Depletion-2 pair can be used for a transistor programming purpose. A D1 device, even with the lowest carrier mobility among the four types of devices, does conduct current more strongly at the same gate voltage than the enhancement type, thus the D1/D2 pair would be faster in transition time than the regular E/D2 pair.

DENSITY: Since this structure does not need metal lines in the array, this PLA's memory area is free for metal lines of other functions on the chip, as shown in Fig. 10. If this PLA section is properly located on the chip, further area savings can be achieved through sharing the memory area with wide power lines or limited number of data/control lines. The problem of poly lines going across the N+ lines can be handled by using depletion implant at the cross section. Properly clocked poly line, buried contact, and a few more contact points to the power lines will further improve the topology and electric conditions in this special application.

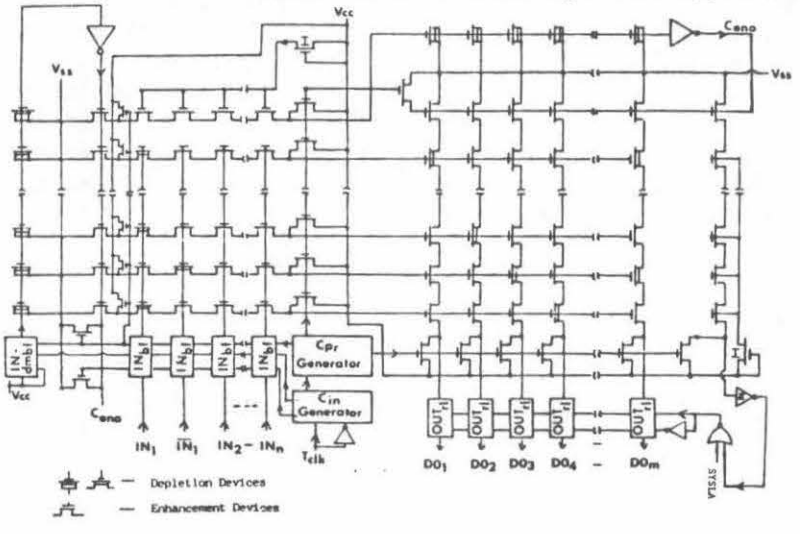


Fig.7b - The Circuit Structure of the NAND PLA.

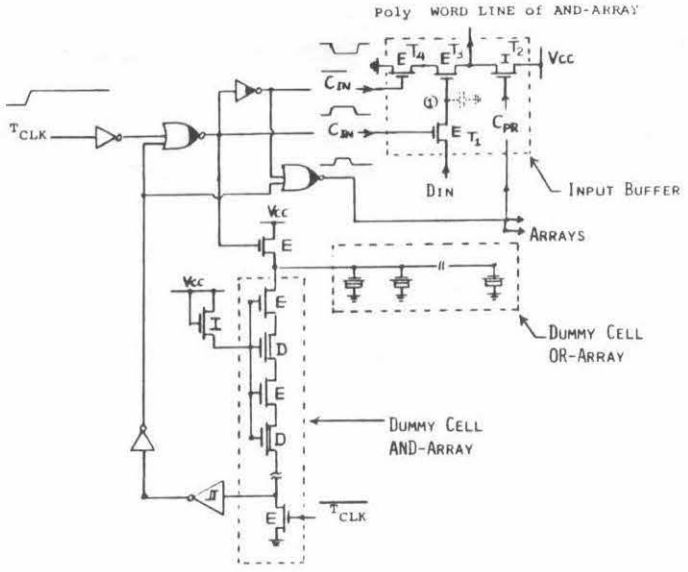


FIG. 9 Input-Buffer and the generation of the self-timed clocks C_{IN} and C_{PR}.

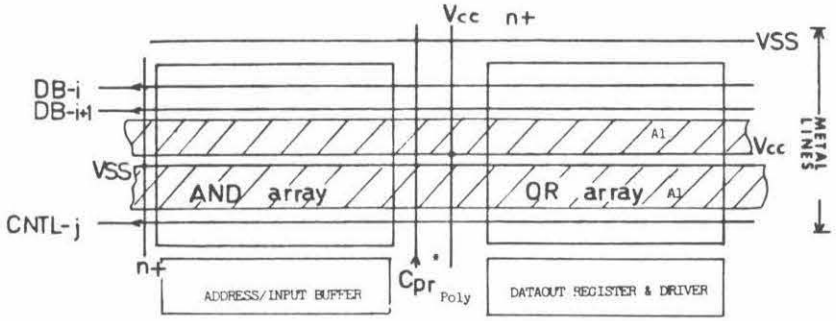
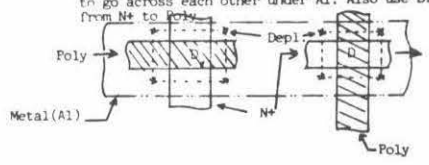


Fig. 10 PLA Memory Area is FREE for Metal Lines!!!

* Use Depletion Ion-implant to allow Poly and N+ to go across each other under Al. Also use B.C. from N+ to Poly.



ACKNOWLEDGEMENTS

The author would like to thank those whose contribution and help made this work possible.

L. Nguyen	--His request for development of an advanced VLSI Chip.
D. Morgan Tuan H.T. (BURROUGHS)	--Their technical evaluation and encouragement.
J. Zeh	--His vision and commitment on the project.
J. Schneider	--His decision, funding and continuous support.
T. Northrup, K. Slater, and F. Zereski	--Their continuous encouragement and support.

Layout	--H. Nguyen, J. McHood, M. Riley and A. Sella
Process	--L. Y. Wu
P400 Program	--R. Spencer
Test	--R. Saul and D. Cote

Review and preparation of the manuscripts, figures and typing--M. Forsyth, J. Blake, A. Flohr, J. Giles, R. Ryan, and M. Burton.

REFERENCES

1. H. Kawagoe and Nobuhior Tsuji, "Minimum size ROM structure compatible with silicon-gate E/D/MOS LSI," IEEE J. Solid-State Circuits, Vol. sc-11, No. 3, pp. 360-364, June 1976.

2. Y. Kitano, S. Kohda, H. Kiduchi and S. Sakai, "A 4Mb full wafer ROM," in ISSCC Dig. Tech. Papers, Feb. 1980, pp. 150-151.

3. Chong M. Lin, "A 4um NMOS NAND Structure PLA," IEEE J. Solid-State Circuits, April 1981.

4. Carver Mead, Lynn Conway and Charles L. Seitz, Introduction to VLSI systems, Addison-Wesley Publishing Co., 1980, pp. 15-16, and Ch. 7.