Generation of Analog Voltages to Improve Flash Memory Read Speed

by

Michelle Ying-Wai Eng

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science in Electrical Science and Engineering

and Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

January 15, 1998

Author ____________________________

Department of Electrical Engineering and Computer Science
January 15, 1998

Certified by ____________________________

Dr. Christopher Terman
Thesis Supervisor

Accepted by ____________________________

Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Eng.

Generation of Analog Voltages to Improve Flash Memory Read Speed
by
Michelle Ying-Wai Eng

Submitted to the
Department of Electrical Engineering and Computer Science

January 15, 1998

In Partial Fulfillment of the Requirements for the Degrees of
Bachelor of Science in Electrical Science and Engineering
and Master of Engineering in Electrical Engineering and Computer Science

Abstract

A method to improve Flash memory read speed is discussed. This methodology entails applying a direct voltage to the wordline of the Flash cells. The direct voltage is applied using "sample and hold" at a supply voltage of 3v. A positive voltage reference of 2v is generated using a Precision Voltage Reference circuit (PVR). The voltage is then held and recharged in a sampling capacitor. The output to the wordline is held constant by adjusting the Active and Positive Pump regulation associated with a Read.
In addition, a study of possible low-voltage techniques is included which may extend 3v "sample and hold" to 1v supply voltage. An op-amp is designed which can handle a low VCC supply (1V < VCC < 1.5V) and a large input voltage swing on the reference voltage (0.05V < Vref < 0.95V). The comparator can be used for Flash memory sensing circuitry and pump regulation circuitry. The input stage uses a complementary design combined with a cascoded input to maximize the input common mode voltage range (CMR). The output inverters are designed using low threshold voltage devices in order to provide a larger output voltage swing and to minimize propagation delay.

Thesis Supervisor: Dr. Christopher Terman
Title: Senior Lecturer

Acknowledgements

I would first like to thank my Intel supervisor Michelle Chuaprasert for her untiring support and positive attitude. Without Michelle, this thesis would never have been published. Also, I would like to thank my MIT thesis supervisor Dr. Christopher Terman for his helpful comments and unbounded sense of humor. Thanks to Sherif Sweha for giving me the chance to work with some of the best engineers at Intel.

Thanks also go to my wonderful friends and family. Without their love and support, I would not be where I am today. Humor me and let me name them: Alice, Ken, Jenny, Kathy, gramps1 and 2. Special thanks to Walter von Koch, my better half.

I would also like to acknowledge Bob B., Marc L., Jeff E., Thomas T., and Kerry T. for providing support and technical expertise. Thank to Lester V., Charles Y., and Janice B. of the Legal department for their speediness in allowing me to publish. Finally, thanks to Mario A., Isidro L., Karthi R., Tom B., and Nelson C. for keeping me sane and saving me from the lions and ants.

Table of Contents

List of Figures

List of Tables

1. General Intro

1.1 Purpose

One of the challenges facing a circuit designer today is optimizing for the worst case VCC, since a chip's power supply can vary within a certain percentage of precision. Thus, if VCC = 3v, circuits must be optimized to be operational at both 2.7v and 3.6v.

The ability to read information from memory quickly is a necessary feature of Flash Memory. If the wordline to the desired Flash cells can be selected faster, one would see a speedup in reads. This thesis will concentrate on a method to improve read speed.

Currently, wordlines in Flash memory are driven high for a read to a value dependent upon VCC. Since the input voltage range can vary, the circuit designer must design for the minimum possible VCC. It is hypothesized that driving the wordline high with a direct voltage instead of referencing VCC will increase the read speed.

The initial design and simulations will be produced for a 3v technology. The methodology will also be used to extend to a 1v differential amplifier. Recreating circuitry for 1v technology introduces many complications, since the threshold voltage of a typical P or N device is 0.8v. The 1v design must conform to a new Intel Process. Many complications arise for low power design. This 1v diff amp can be used in conjunction with future innovation to extend the completed 3v design to 1v.

1.2 Flash memory

In this new age of shrinking laptops, handheld PCs, digital computers, and handheld recorders, memory storage which is safe, reliable, and low-power is becoming a high priority. Flash memory has many advantages over traditional memory media such as ROM, SRAM, EPROM, EEPROM, and DRAM. These advantages include non-volatility (retains memory after power off), updateability (rewrites are possible), high density, ruggedness, and re-write ability within a host (no UV light needed for an erase).

The Flash transistor allows its threshold voltage to be changed electrically. The modified voltage remains even after the power supply is turned off. Alteration of the threshold voltage is made possible by the Flash cell's floating gate. The polysilicon floating gate sits, or "floats," between the gate and the channel.

When the Flash threshold voltage is set greater than or equal to 5.3v, the Flash cell is considered programmed. Programming, or the "0" logic state, is accomplished by hot electron injection. The high voltages applied between gate, drain and source of the Flash transistors create an electric field, which allows some electrons to become "hot" and jump onto the floating gate. Eventually, no more charge can be accepted on the floating gate. When the programming voltages are removed, the negative charge on the floating gate remains, resulting in a higher threshold voltage for the Flash cell.

The Flash cell is "erased," or "1," if it has a threshold voltage less than or equal to 3.1 v. Erase is done using Fowler-Nordheim tunneling. Tunneling creates a field which removes electrons from the floating gate and thus lowers the Flash cell's threshold voltage.

A write to Flash memory is performed by erasing entire blocks in a "flash" and then programming the desired bytes or words. Erasing blocks instead of single devices ensures reliable threshold voltage values and avoids possible problems with device characteristics.

During a read, one must set the gate-to-source voltage of the Flash cell to 5v. If the drain current of the Flash cell flows, then the cell is a "1" (erased). If no current flows, the cell is a "0" (programmed). [1] The threshold voltages for the flash cell and a read are depicted in Figure 1.

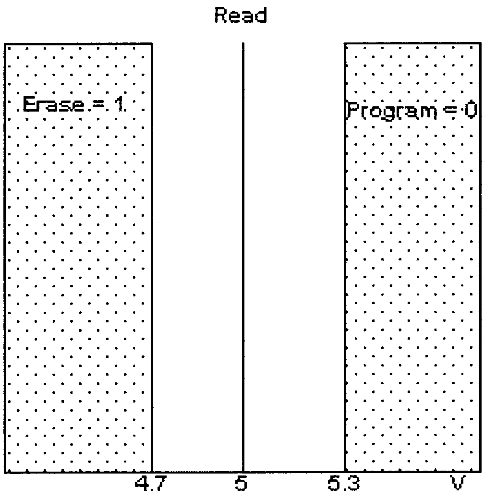


Figure 1: Threshold voltages

[1] Rabaey, Jan M. Digital Integrated Circuits. New Jersey: Prentice Hall Electronics and VLSI Series, 1996, pp. 573-577.

The necessary gate, drain and source voltages for program , erase and read of Flash cells are pictured in Figure 2.

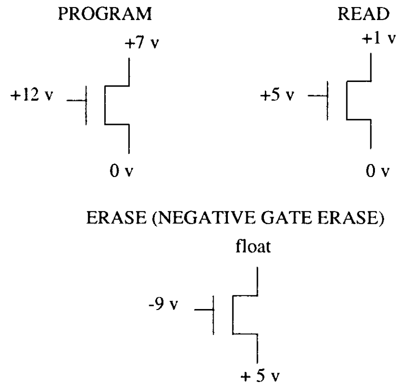


Figure 2: Required voltage levels

1.3 Pumps

Since the input power supply is 3v, one must determine how to generate 5v on the gate of the Flash cell. Voltages higher than VCC are produced by using pumps.[2] The pump concept is based on the bi-polar voltage multiplier. Switching the voltages on the ends of a capacitor can increase the resulting reference voltage. For example, if a capacitor originally has a 3v differential and the bottom node is switched to 3v, the top node will switch to 6v to maintain the 3v differential. By using this switching concept with clocked coupling capacitors, one can increase the input voltage through stages. Diodes must also be placed after the input supply and between capacitors to prevent

[2] Wu, Jieh-Tsomg, et. al. “1.2V CMOS Switched-Capacitor Circuits,” IEEE Solid-State Circuits Conference, Vol. 39, pp.388-389, February, 1996.

reverse bias current. These diodes can be implemented in CMOS using s-devices. S-devices have a lower threshold voltage than NMOS devices. Bootstrapping is also used in pumps to avoid threshold drops in voltage across devices. In other words, bootstrapping "boosts" the gate voltage on the diode to avoid a threshold voltage drop across the diode. An example pump is shown in Figure 3.

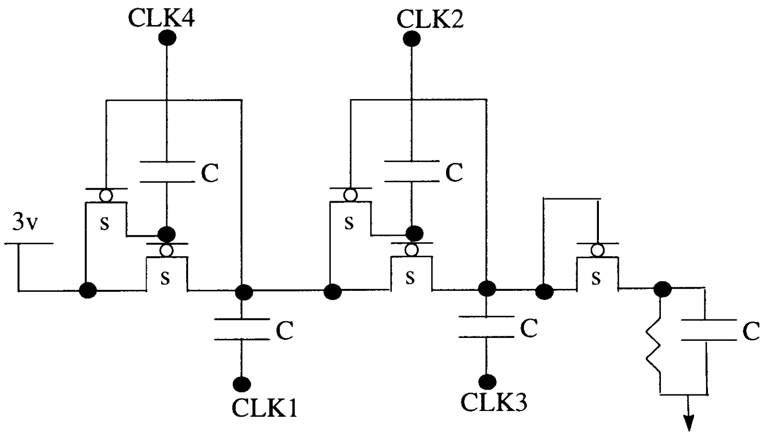


Figure 3: Charge pump "cell"[3]

An actual pump combines many of these s-device pump "cells." Clock drivers supply the clock signals to the pump. To keep the output voltage constant, regulation is used. Regulation is achieved using a differential amp, voltage divider, and oscillator as shown in Figure 4. If the inputs to the diff amp are not equal, the output of the diff amp will rise or fall. A fall in voltage at the output of the diff amp will speed up the oscillator.

This rise in oscillator frequency will change the clock drivers, which will alter the output pump voltage. As the output of the entire pump reaches its desired regulated value, the oscillator will slow down.

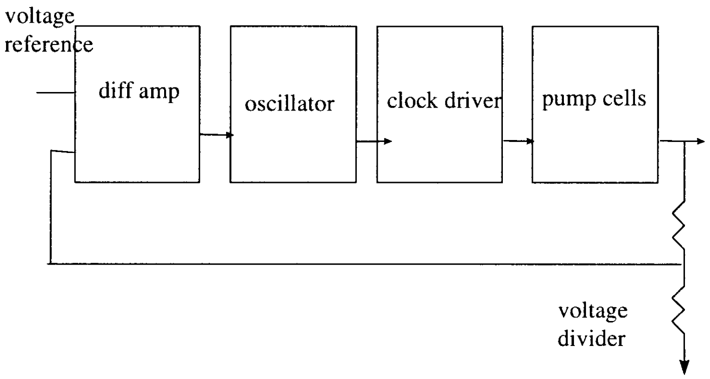


Figure 4: Pump operation with regulation

1.4 Read - wordline

The gates of the flash cells of the memory array are attached to a global wordline. In order to place 5v on the gate of the flash cells for a read, the p-device driver for the global wordline must be driven high. The decode path, including this driver, is shown in Figure 5.

[3] Tedrow, Kerry, Johnny Javanifard, et. al. "System having multiple phase boosted charge pump with a plurality of stages," U.S. Patent No. 5524266. Issued: June 4, 1996.

The positive pump output (HSRCDRV) is connected to the source of the p-device driver. The gate of the p-device driver (HGTDRV) is connected to the negative pump output. The voltages applied to the driver during a read have historically been 1.5 * VCC and -1.5 * VCC respectively. Since VCC can vary from 2.7v - 3.6v, the voltages applied range from 3.9-5.5v.

An address change at the pads will send the requested address to the required decoder and wordline. Initially, HNGTDRV=HNSELWVK=HSRCDRV=HSRCDIV = 1.5* VCC and HADDR = 0V for a read. The gate of the p-driver, HGTDRV is always kept on at -1.5*VCC. The address change will enable PREDECRA, PREDECRB, and DECR. This change will cause node A to switch from 1 to 0 and output 1.5*VCC onto PSRC. A read in the decode path is depicted in Figure 6.

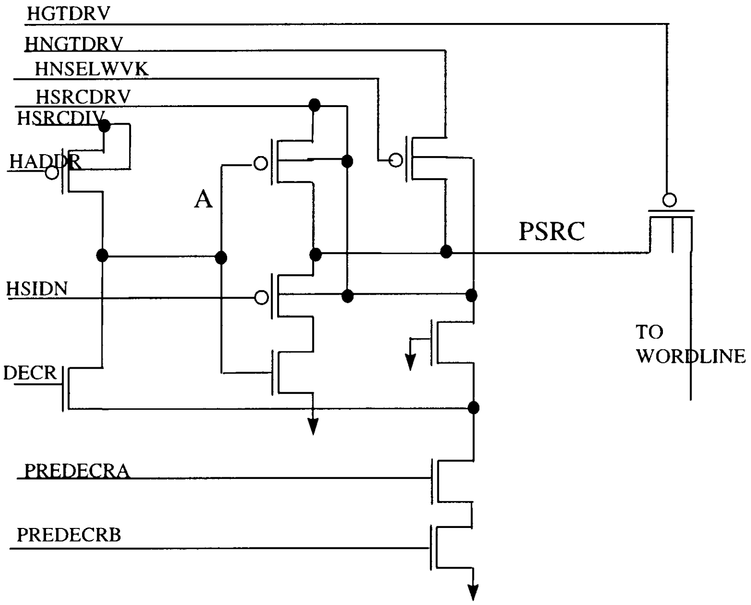


Figure 5: Decode path to wordline

It is hypothesized that there are two methods of improving read speed. First, the p-device driver can be sized as large as possible without largely affecting die size. Increasing the size directly increases speed; however, doubling the size of the devices would entail a doubling in area.

Secondly, one can directly apply up to 5.5v onto PSRC instead of referencing VCC and designing for VCC=2.7v.

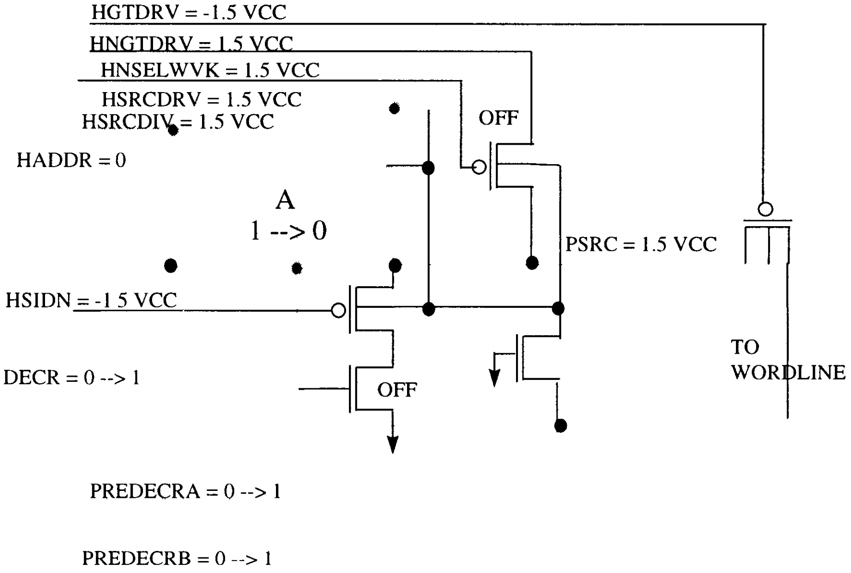


Figure 6: Path to WL for a read

The second method would require a constant voltage reference which cannot fluctuate with VCC. References of 4v (REF4) and 2v (REF2) have already been invented. The Precision Voltage Reference circuit (PVR), which generates REF4 and

REF2, uses the difference in threshold voltages between two flash cells to generate a constant voltage.[4] The PVR will be discussed in depth in the next chapter.

1.5 Overview

Implementation of 3v "sample and hold" involves first examining the path to the flash cell wordline. Simulations were run to determine the before and after effects of applying a direct voltage to the wordline. After verifying a speed improvement, one then needed to design the additional circuitry. A sampling capacitor was added to the PVR circuit to hold 2v constant. Other circuitry was also added for functionality. Two signals, PVREN and SAMPLE, are periodic and originate from the oscillator. These signals are used in combination to enable and refresh the 2v held in the sampling capacitor. Other simulations were run to verify the size of the capacitor chosen, leakage effects, coupling of other circuits, and decay off of the capacitor. Powerup and warmup sims were run to assure that these timings were not changed. Also, the standby current was measured and found to increase to 5.4uA with the new changes to the PVR. Two implementations of the PVR are discussed in the thesis, one in metal options, and the other a CAM option.

To implement 3v "sample and hold," changes were also necessary to the active pump, which outputs the voltage needed to the wordline during a read. The original operation of the pump was first simulated and tested. Then new logic and regulation was added to take the constant REF2 (2v) input to the pump instead of the 1.5VCC input. The new regulation for the pump needed to be resized. Simulations were run to find and

[4] Tedrow, Kerry, et. al. "Precision Voltage Reference, " U.S. Patent No: 5339272. Issued: Aug. 16, 1994.

verify these sizes. Powerup was also run to verify that the timing was not changed. Standby current was also measured to ensure that it was not increased from previous numbers. Finally, the active pump was implemented and tested in both metal options and CAM options.

Since the PVR is essentially a differential amplifier, a 1v op-amp is developed within this thesis. The 1v op-amp is crucial in the search for a working 1v PVR and eventually a working 1v "sample and hold." Techniques to lower the supply voltage to 1v are discussed. These range from changing the process to using "new" devices which dynamically alter the threshold voltage of transistors. A 1v op-amp is then developed which attempts to avoid a change of process or "new" devices. A simple differential amplifier is first simulated. These initial sims fail operation, since the V_t of a device (0.8v) is obviously too high. Next, a complementary input stage comparator which is operational at 1.5v is used as a starting point for the 1v op-amp. This comparator is resized to function at 1v with an Ibias of 15uA. However, the op-amp is too slow and the current is too small to generate. Therefore, the op-amp is resized to operate with an Ibias of 30uA.

To improve the propagation delay, prime devices which have low threshold voltages are used in the output inverters. This allows the output of the amplifier to swing rail-to-rail. Finally, the input stage is altered to a cascoded complementary input stage. This allows the common mode range (CMR) to swing rail-to-rail at a low voltage. The final op-amp with cascode and prime devices is simulated and verified to be operational at 1v. However, the propagation delays in the resulting op-amp can still be improved and optimized. Final simulations are done to verify the op-amp's transfer curve and operation as a unity gain buffer.

2. 3v sample and hold

2.1 Precision voltage reference (PVR)

2.1.1 Introduction

2.1.1.1 Purpose and functionality

The precision voltage reference circuit generates voltage references REF4 (4v) and REF2 (2v). These 4v and 2v references are used as inputs to various parts of the chip, such as the positive and negative pumps, the vpp detectors, and the Y-path series regulated loadline. All of these circuits need a reliable voltage reference which does not vary as VCC can. The precision voltage reference is particularly important for the positive and active pumps. During chip turn on, the pumps must use REF2 instead of VCC regulation (since it takes a while for VCC to ramp up). The block diagram of the use of REF2 in the pumps is shown in Figure 7. The active pump logic chooses between REF2 and VCC regulation. The active pump and changes will be explained in greater detail in the next section.

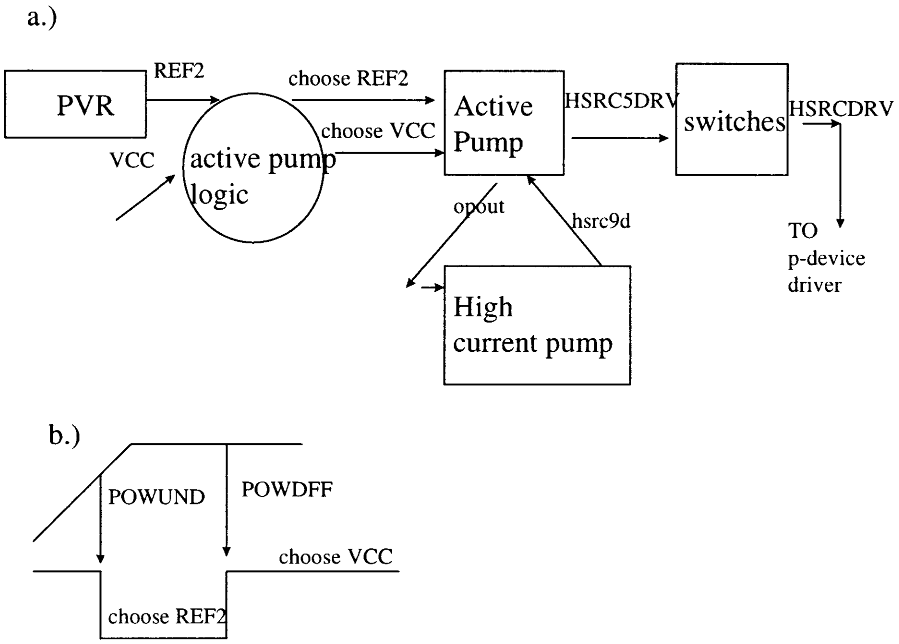


Figure 7: (a) Block diagram of pumps using REF2. (b) Traditional waveforms for choosing REF2

The precision voltage reference circuit uses the difference in threshold voltages between two flash cells to generate a constant voltage.[5] The theoretical and CMOS precision voltage reference is pictured in Figure 8. Resistors R and 2R form a resistor

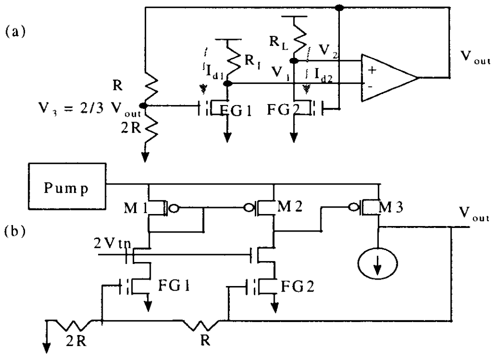


Figure 8: Precision voltage reference: (a) Theoretical, (b) CMOS

divider relationship so that $V_3 = 2/3\ V_{out}$. In order to minimize the fluctuations of the op-amp at equilibrium, the current I_{d1} and I_{d2} must be the same, and V_1 must be close toV_2.

The equations to determine V_{t1} and V_{t2} are as follows:

$$I_d \sim V_{gs} - V_t$$

$$V_{t1} = V_{gs1} - I_d$$

$$V_{t2} = V_{gs2} - I_d$$

$$V_{t2} - V_{t1} = V_{gs2} - V_{gs1}$$

$$V_{gs1} = V_3 = 2/3V_{out}\ ;\ V_{gs2} = V_{out}$$

$$V_{t2} - V_{t1} = V_{out} - 2/3V_{out}$$

$$V_{out} = 3(V_{t2} - V_{t1})$$

For $V_{out} = 4v$, $V_{t2} = 4$, $V_{t1} = 2.667$.

Figure 8(b) depicts the CMOS version of the precision voltage reference. The flash cells FG1 and FG2 are different from conventional flash cells because they are used to store specific charge values (4 and 2.667) instead of continually being programmed and erased with various threshold voltages. The two NMOS transistors connected to the drain of the flash cells are cascode devices. These transistors maintain the voltage at the drain of the flash cells at a threshold voltage below the gate voltage of the cascode device. The cascode devices prevent drain disturb, making sure the flash cell drain voltages are equal in equilibrium. In other words, the drain of FG1 = drain of FG2 = Vtn. Gates M1 and M2 act as a current mirror. Thus, if the current through FG1 is decreased in comparison to FG2, the current through M1 will decrease by decreasing the drain voltage.

[5] Tedrow, Kerry, et. al. "Precision Voltage Reference, " U.S. Patent No: 5339272. Issued: Aug. 16, 1994.

The lowered drain voltage of M1 is reflected and lowers the Vgs of M2. Subsequently, M2 will want to decrease current. Since FG2 is still maintaining the same amount of current as before, it attempts to transfer its current to M2. This higher current, in conjunction with the lowering of the Vgs of M2, causes the Vds of M2 to increase. Vds of M2 increasing implies a lowering in the drain voltage of M2. M3 is an s-device which acts as a source follower (change in voltage at its gate is reflected at its source -Vt). Thus the voltage on the gate of M3 is lowered, which lowers the output current and V_{out}. The drop in V_{out} will lower the gate voltage at FG2, which will lower the current through FG2 and put I_{d2} and I_{d1} into equilibrium again.

2.1.1.2 Inputs and outputs

Table 1 summarizes the inputs and outputs of the precision voltage reference. It includes the new signals added: PVREN, SAMPLE, and SELSMPHLD, which will be explained through the course of the thesis.

Inputs	
IS5V	indicates VCC=5v
POWDFF	enable PVR during powerup (on 55us after POWUND)
PGMEN	enable trimming of PVR
MF1DRAIN	connected to drain of 2.666Vt cell during trim
MF2DRAIN	connected to drain of 4.0Vt cell during trim
MFGATE	used for trimming flash cells
PVREN	PVR enable, from MFO
SAMPLE	PVR sample (refresh) from MFO
SELSMPHLD	ON = sample and hold OFF = old PVR
Outputs	
REF4	outputs 4V
REF2	outputs 2V (sampled)

Table 1: Precision voltage reference inputs/outputs

2.1.1.3 Past Work

Unfortunately, the precision voltage reference's circuitry burns too much power and current and is therefore shut-off during Standby or Deep Power Down modes. The precision voltage reference is also off during Active mode (Read mode), since 1.5*VCC is used for the wordlines. Also, the PVR takes more than 600ns to warm up when turned on. This would make it impossible to try to enable the circuit for a Read, which has to happen in under 120ns. In other words, if a Read happened after the chip was in Deep Power Down mode, there would not be enough time to warm up the precision voltage reference circuitry and to access the constant REF2 reference. One would need a way to hold the 2v reference permanently so that it could be used as a reference to the active pump. The constant REF2 reference would allow a constant voltage to be generated on the wordline.

A recent Intel project has designed a way to hold REF2. However, the project's objective was to obtain tighter regulation of their voltages. This high precision voltage regulation is needed to program the multilevel Flash cells in their project.[6] For multiple bit Flash memory, the sensing regions for threshold voltages are much smaller. Therefore, the voltages need to be accurately regulated. The multilevel cell project is also designed on a 5v technology with different circuitry, including completely different pumps, regulation, and logic. Yet the idea of holding REF2 seems possible to apply to this problem of direct voltage regulation. The 2v reference could be held in a capacitor which would be recharged every so often in a "sample and hold." The main difference would be that the "sample and hold" will be used for improving read speed instead of for regulating programming voltages. Also, the extra voltage self-regulation circuit used in the previous project for programming is not a necessary feature, since the sensing margins for the current project allow more room for error.

[6] Tedrow, Kerry, et al. "High precision Voltage Regulation Circuit for Programming Multiple bit Flash Memory, " U.S. Patent No: 5546042. Issued Aug. 13, 1996.

The main concern of using "sample and hold" is its reliability and the impact of the large capacitor on die size. REF2 and REF4 must be accurate to within 3% of their values at all times. Results from the past 5v Intel project indicate that "sample and hold" can be reliable and also consistent, since it maintains the precise voltage of REF2 and allows the precision voltage reference to only turn on during refresh of the capacitor.

2.1.2 Methods

The initial step in determining the feasibility of direct wordline regulation is to examine the path to the wordline. When an address change occurs on the pad of the chip (indicating a read request), the address is input into the decode path. The end of the decode path to the wordline is shown again in Figure 9. The correct wordline is chosen by the decode path. To set the wordline high for a read, a gate-to-source voltage of approximately 11v is placed on the p-device driver for the wordline.

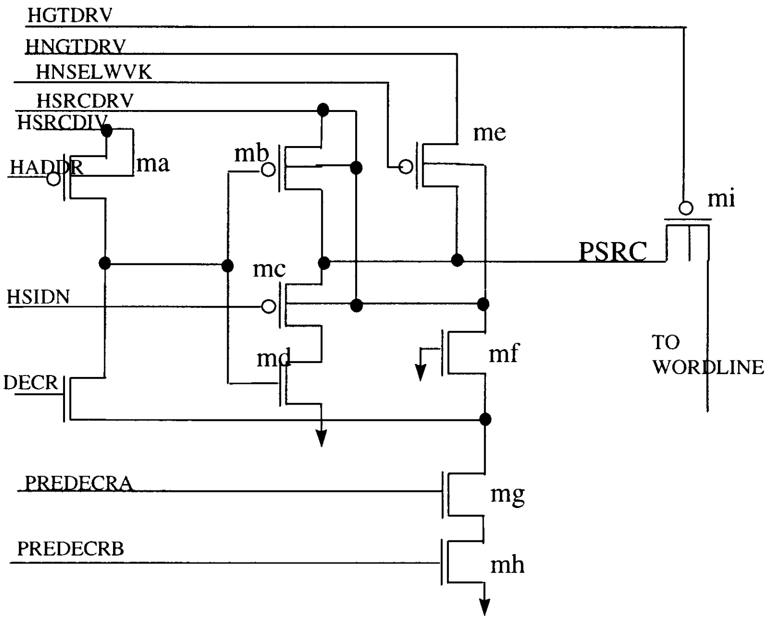


Figure 9: Decode path to wordline

Initial simulations were done to verify the wordline select and deselect timings. In these simulations, the traditional voltages applied to the p-device driver are 1.5*VCC to

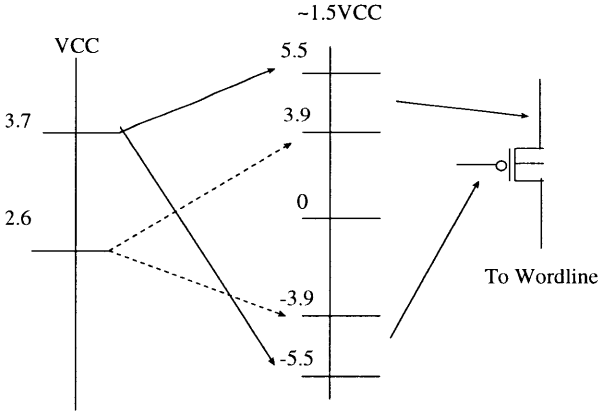


Figure 10: Map of traditional VCC to wordline voltage

the source and -1.5*VCC to the gate (see Figure 10). If the p-device driver is doubled in size, there is a significant decrease in wordline select time. However, doubling the p-device size would entail a doubling in area. The results of the original simulations of the decode path are in Appendix A.

The hypothesis of a speedup in wordline select time by applying a direct voltage is tested by simulating the decode path in HSPICE and replacing (during a read) HSRCDRV=PSRC=1.5VCC with a direct range of arbitrary voltages. Test vectors sweep the voltage HSRCDRV from 3.9-5.5v to determine the speed-up gain.

The results of direct voltage for worst case HSRCDRV at slow n slow p, T=100, VCC=2.6V are shown in Table 2.

HSRCDRV (V)	WL select time (ns)
1.5VCC	142
3.9	145.5
4.9	112.9
5.1	110.0
5.3	107.71
5.5	107.71

Table 2: Results of direct voltage on wordline (multiplied by an undisclosed constant)

Results for other skews are shown in Appendix B.

These results clearly indicate that applying a direct voltage to the source of the p-device driver will speed up the wordline select time. Since devices tend to breakdown with high gate-to-source voltages, and the maximum value of HGTDRV is -5.5v, HSRCDRV should range from 5.1-5.4v. However, more in-depth analysis needs to be made by other project members to determine if there is an issue or problem with having a low VCC = 2.6v and a high wordline line voltage = 5.4v. To test this theory, the following HSPICE formula can be used:

HSRCDRV = is5v + [(1.75 + (sign (-.25, vcc-3)) * vcc*(0.5 - sign(0.5, vcc-4))].

This formula accounts for process shifts (which can change depending on VCC) and changes the multiplier. In other words, 5.4volts/3.7v = 1.5x, while 5.4volts/2.6v = 2x. The formula was derived by using the final simulation data of the output on HSRCDRV from the pumps. Other project members will use the derived equation for future simulation and research.

In order to apply a direct voltage to the wordline driver, it is important to understand the path to the wordline upon chip powerup. The voltage HSRCDRV is generated in the following manner: a pump which is regulated to the correct value outputs to switches, which outputs HSRCDRV. There is also logic which controls the inputs to the regulation of the pump. A block diagram of the path of HSRCDRV is shown in Figure 7.

To use a direct voltage for HSRCDRV, the logic for regulation of the pump (the Active Pump) will have to be changed to always use REF2 instead of VCC as a reference. The regulation will also need to be altered to obtain the correct value desired on HSRCDRV.

The reference of 2v (REF2) generated by the precision voltage reference circuit needs to be held constant so that it can be always be used as a reference to the active pump during a read. This methodology of "sample and hold" previously mentioned is implemented and tested for 3v functionality within this thesis.

A "sampling capacitor" is added to the output 2v node to hold the 2v reference of the PVR constant. The first task undertaken is to determine the size of the sampling capacitor. The capacitor is sized as small as possible to minimize impact on die size. It is also sized to minimize voltage change on the output node due to coupling of other circuits. Initial hand calculations are done by estimating the load on the 2v output node (REF2). By taking into account the coupling from other circuits, one can estimate the worst case variable voltage change on REF2:

$Q_s = C_sV_s \;;\; Q_L = C_LV_L$

$V_t = Q_t/C_t = (C_sV_s + C_LV_L)/(C_L + C_s)$

$V_s = 2V \;;\; V_L = 2 \pm 1v$; assume load is not fully charged

$V_t = [2C_s + C_L(2 \pm 1)] / (C_L + C_s)$ [7]

Estimates of C_L are made by first calculating the capacitance for the current Intel process. The input gate capacitance of a minimal length transistor is found using the equation:

$C_{gate} = [(\varepsilon_o * \varepsilon_{ox})/t_{ox]} * \text{area.}$ [8]

$\varepsilon_{ox} = 3.9 \;;\; \varepsilon_o = 8.854e^{-12}$

Wire or interconnect capacitance is estimated using the HSPICE model:

[7] Weste, Neil and Kamran Eshraghian. Principles of CMOS VLSI Design. 2nd edition. Reading, MA: Addison-Wesley Publishing Co., 1993, p. 241.

[8] Weste, Neil and Kamran Eshraghian. Principles of CMOS VLSI Design. 2nd edition. Reading, MA: Addison-Wesley Publishing Co., 1993, p. 181.

$CAP_{eff} = M*Scale[L_{eff}*W_{eff}*C_{ox} + 2(L_{eff} + W_{eff}) *Capsw]$. [9]

$M*Scale = 1$;

$L_{eff} = L_{drawn} - 2\ dw$; $dw = DL_{eff}$ (values from Intel process library).

$W_{eff} = W_{drawn} - 2dw$; $dw = Dw_{eff}$

C_{ox}, dw, Capsw are taken from the Intel process library.

Interconnect capacitance is estimated for L_{drawn} = 1000um and W_{drawn} = 1um.

The complete load and values for gate capacitance and interconnect combined are estimated in Table 3.

Circuits	**gate capacitance (pf)**	**interconnect cap (pf)**	**total load (pf)**
powerde	0.12	0.088	0.208
hsrcdrvreg	0.18	0.208	0.388
lcpmp	0.6	0.24	0.838
actpmp	0.07	0.387	0.46
pmpoth	0.2	0.208	0.408
hladout	0.308	0.89	1.2
negpmp	0.6	3	3.6
pmpnegreg	0.3		0.3
		total C_L =	7.4 pF

Table 3: Capacitance estimates for coupling calculations

Summing up the total load from the table, C_L = 7.4 pF.

[9] Avanti. Star-Hspice User's Manual. Vol 2. Campbell, CA: Meta-Software, Inc., 1996, p.11-9.

Therefore, $V_t = [2C_s + (7.4pF) * (2 \underline{+} 1)] / 7.4 \text{ pF} + C_s$.

A table of more current interconnect parasitics, using new and updated layout is in Appendix C.

Table 4 shows the percent coupling error for a 1v change on C_L. From the table, the size of the capacitor is chosen to be 30pF. A 1v change on C_L is an overestimation; actual test data from a test circuit indicates a very small coupling change ~0.3mV = 0.0015% error for a 30pF capacitor.

C_S	+1v	-1v	error
20pF	2.27	1.73	13%
25pF	2.23	1.77	11%
30pF	2.20	1.80	9.89%
35pF	2.17	1.82	8.7%
40pF	2.156	1.84	7.8%

Table 4: Coupling error on REF2

The load model for the node REF2 is determined by attaching models of circuitry which take REF2 as an input. FTRC is used to model interconnect, reflecting the distance from the precision voltage reference circuit to the circuit requiring REF2.

The load model is simulated with the original precision voltage reference circuit (without sample and hold) to verify operation and warmup within 600ns. An important part of the load model is the current drawn on REF4 from the negative pumps. This is reflected in the DC voltage in the negative pump model to reflect the worst case current drawn during an Erase, which is about 75uA. For this simulation, the precision voltage reference is enabled after 60ns. The flash gates are set to 2.67 and 4.0 in the HSPICE model. The results of the simulation for VCC=2.6v, T=100, mark are shown in Figure 11. As can be seen in the diagram, it takes REF4 and REF2 less than 600ns to rise. REF4 is designed to initially overshoot 4v and then drop back to 4v. This is to ensure the precision of REF4 with loading.

Decay of the sample capacitor is not a strong issue as long as the refresh capacitor time determined is short enough to avoid a large decay. The capacitor is a Poly1/Poly2 capacitor, which can be broken into squares for ease of layout. To avoid a large impact on die size, the 30 pF capacitor will be hidden under routing signals. Therefore, the layout of the sampling capacitor will be done after all other circuits have been completed.

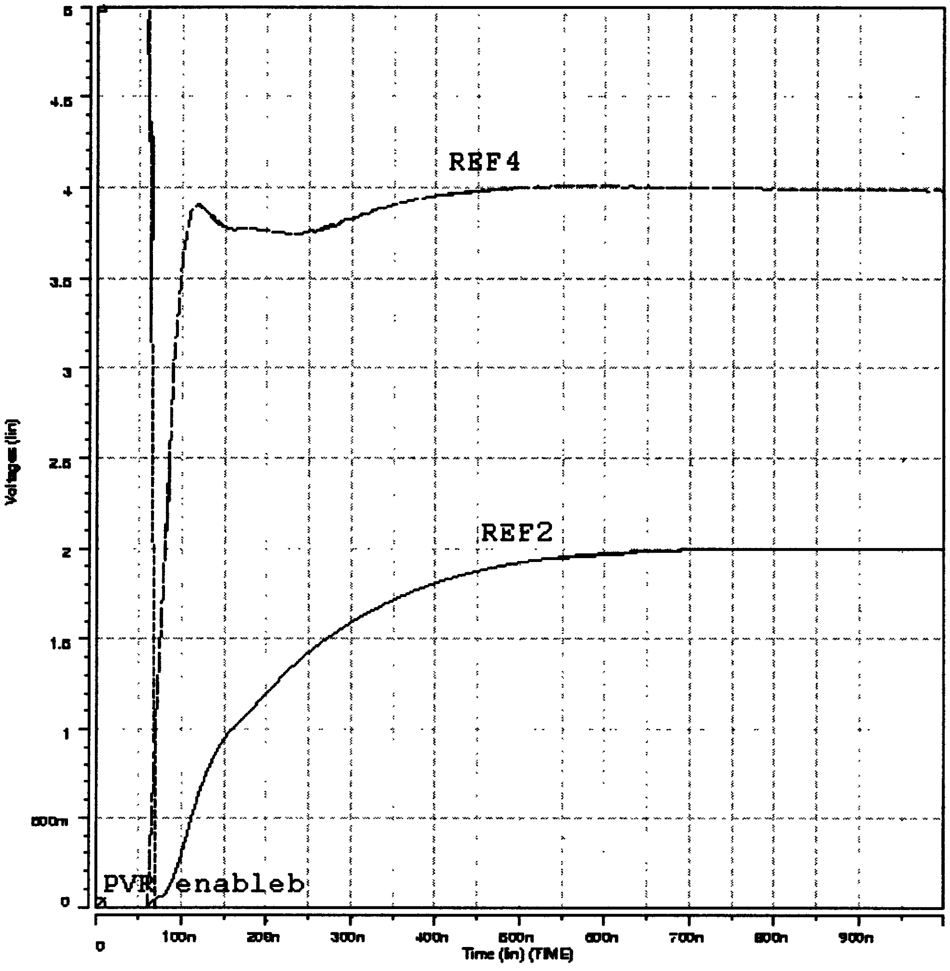
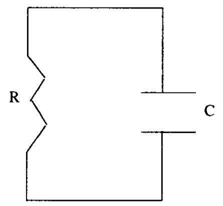


Figure 11: Original PVR with load

On a simple level, one can do an imprecise hand calculation for the decay. The capacitor and loading circuits can be modeled as an RC circuit:



The equation for the voltage over time is V= $V_0e^{-t/RC}$, where V_o = 2v. Assuming that the loading circuits are modeled as capacitive loading and resistance (in actuality, the gates also have diode leakage), we have $V = 2e^{-1.5ms/5.1Mohm(37pF)}$. The voltage level at this approximation is about V=1.96, within the 3% of 2v that is required. An HSPICE simulation gives more accuracy, proving that the capacitor does not decay much. The results of the simulation are shown in Figure 12. To test the decay of the 30pF capacitor, a schematic is built which includes the load model, the capacitor, and a 2v voltage supply. The voltage supply is attached to the capacitor and load (REF2) from 101ns to 3000ns. Then, the voltage supply is shut off, and the charge is allowed to decay over time. From simulation, the voltage stored in the 30pF capacitor decays to V=1.99v after 5ms. This voltage drop is obtained using the worst case situation, in which all load would be on.

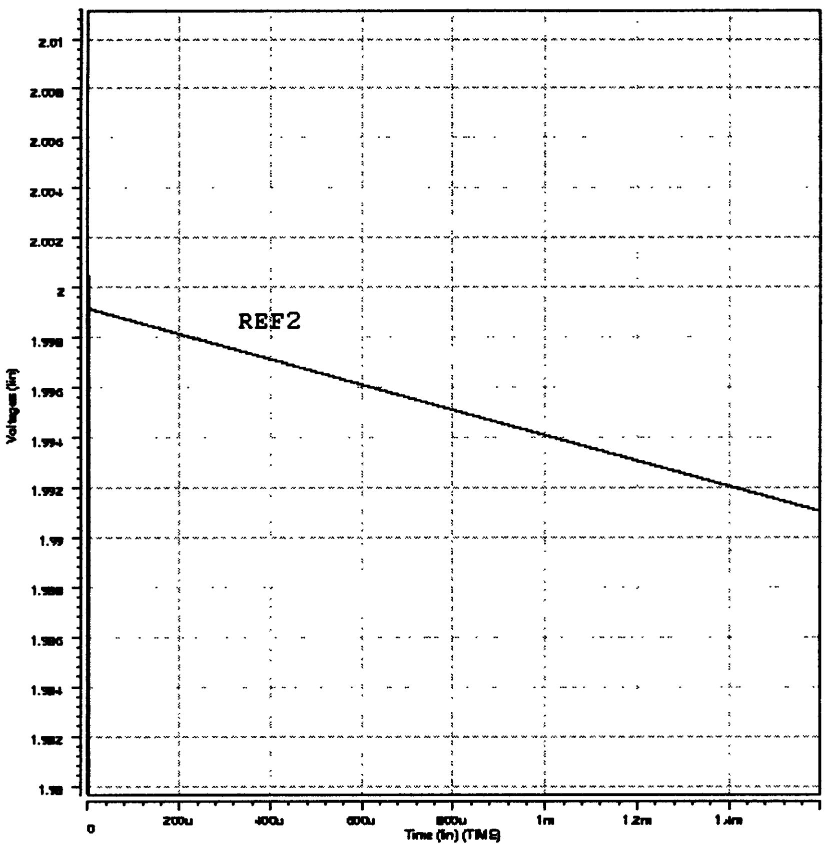


Figure 12: Decay of capacitor over time

A block diagram of "sample and hold" is shown in Figure 13. The sample

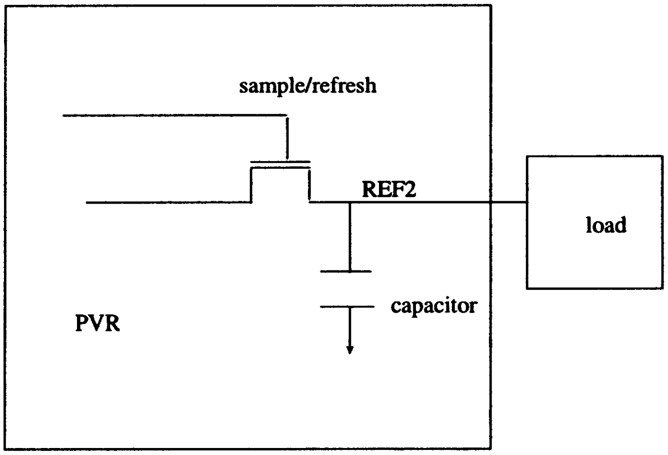


Figure 13: Concept of "sample and hold"

capacitor holds the charge of REF2 and is periodically refreshed by connecting the PVR to the REF2 node.

To initially test the sample and hold theory, a model is used for the PVR behavior. The model is shown in Figure 14.

A 2v power supply simulates the output of the Precision voltage reference. A 1.9v power supply is connected to the capacitor and load to initially charge them to 1.9v. 1.9v is well below the level required for REF2 and thus simulates below the worst case. The 1.9v supply is connected from 100ns to 2001ns. The sampling transistor is then turned on at 2.6us. As the resulting waveform in Figure 15 shows, it takes about 250ns to charge up the 1.9v load to 2v. However, these initial sims are not completely accurate, since it does not include the real precision voltage reference.

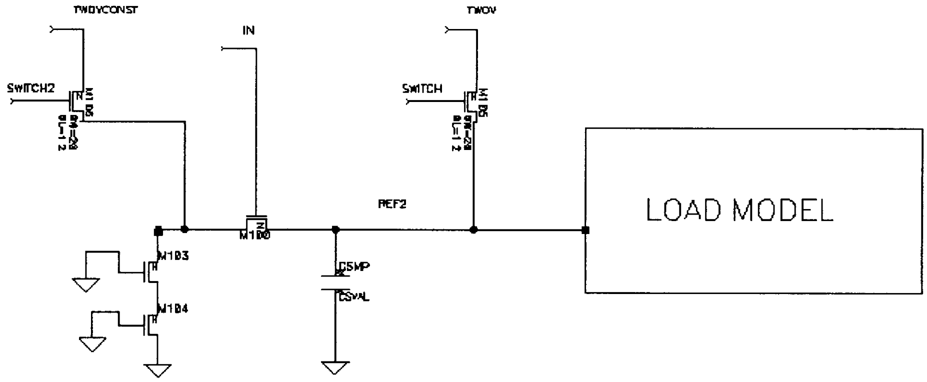


Figure 14: Model to test initial sampling changes

When attaching the actual precision voltage reference circuit, the refresh time of the capacitor and load increases, since some reverse current and the coupling of capacitors slows down the charge up process.

The modifications to the precision voltage reference to allow "sample and hold" are depicted in Figure 16 as metal options. Zoomed schematics are pictured in Appendix D-Appendix E. The additional signals are: PVREN and SAMPLE. These signals are derived from the oscillator. PVREN enables the precision voltage reference, while

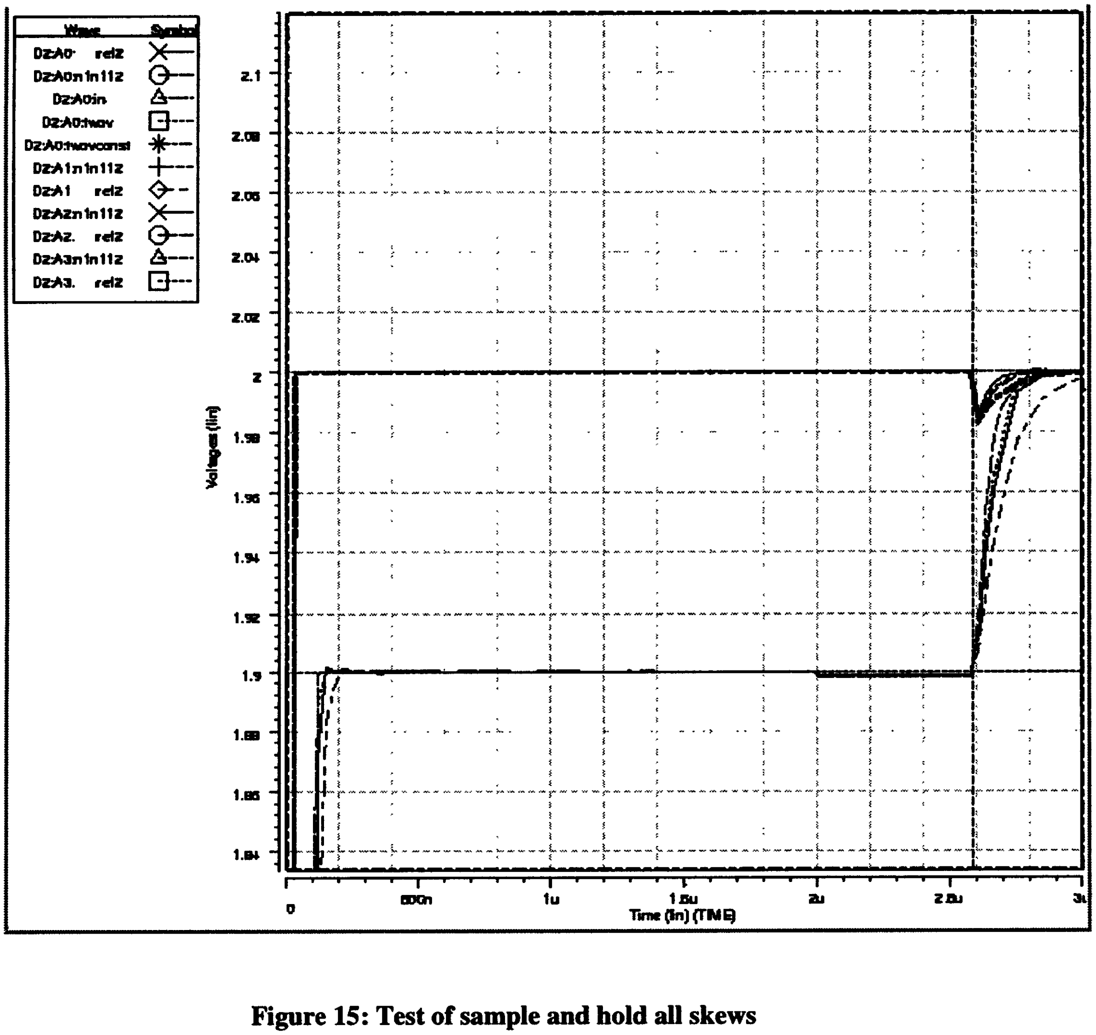


Figure 15: Test of sample and hold all skews

SAMPLE turns on the pass-gate to refresh the capacitor. POWDFF is also input and ORed with SAMPLE, since the load for REF4 and REF2 needs to be connected during chip turn on. PMPEN must be high before the capacitor can be sampled to ensure that the precision voltage reference is on and charged before refresh. CSMP is the sample 30pF capacitor which holds the 2v charge when the precision voltage reference is off. The changes are initially made with metal options.

However, due to the complexity of the circuitry, the changes are later implemented using a CAM option and test bit SELSMPHLD. The CAM option alternative is depicted in Figure 17. Zoomed versions of the CAM option are shown in Appendix F-Appendix G.

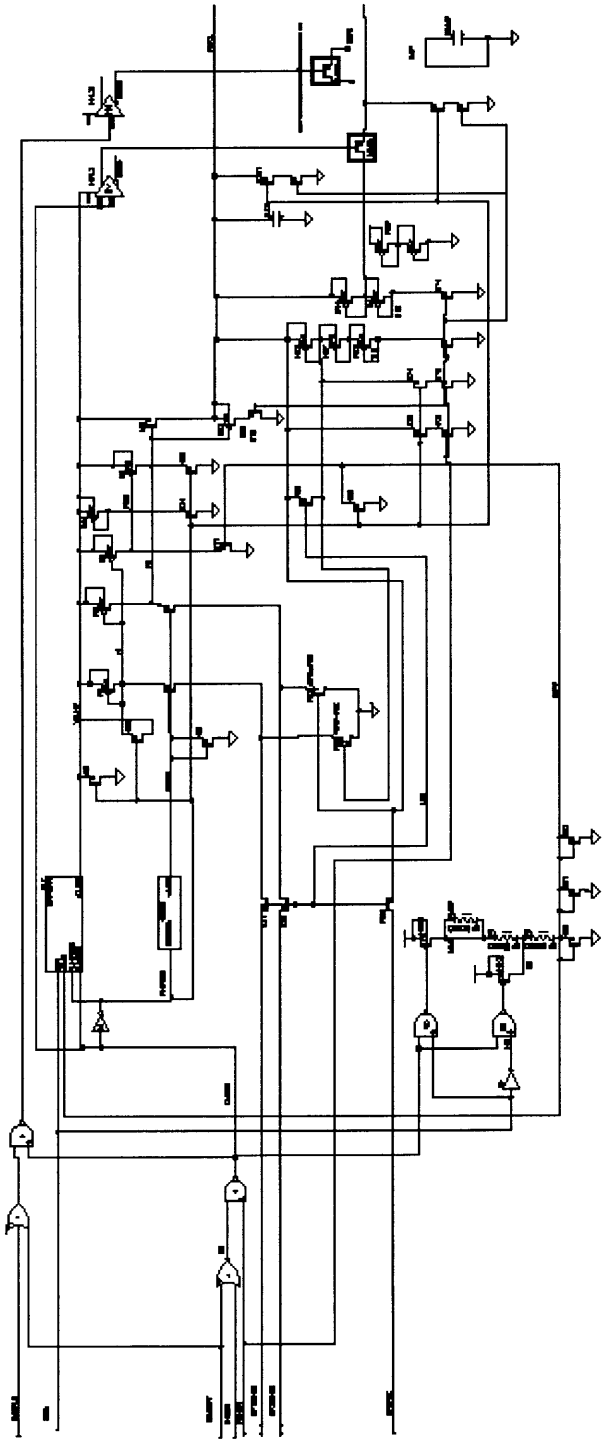


Figure 16: PVR with Metal options

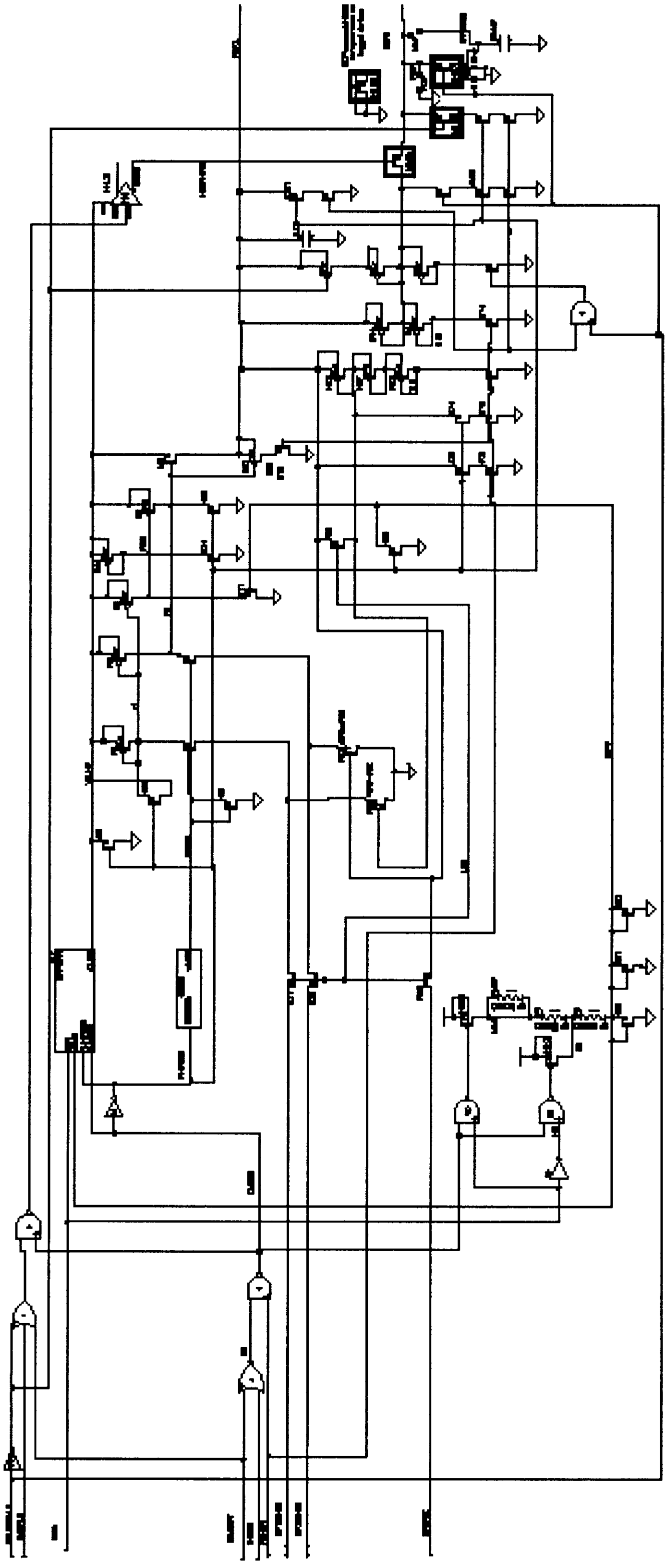


Figure 17: CAM option of PVR

The sampling waveforms for refreshing the sample capacitor are shown in Figure 18.

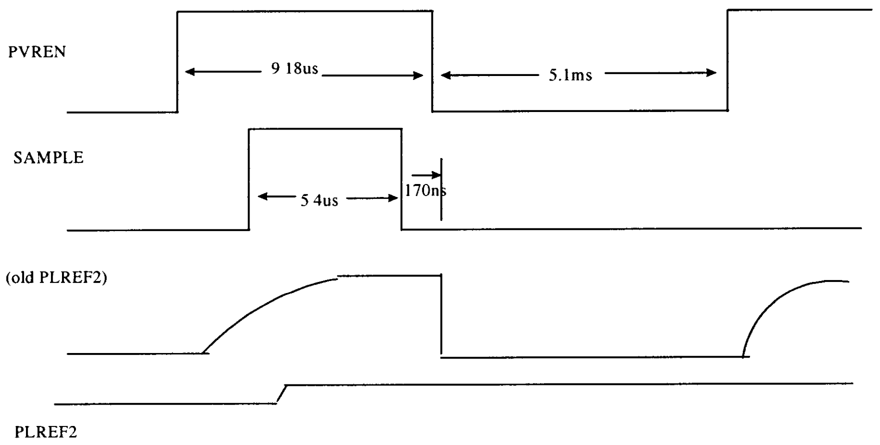


Figure 18: Sampling waveforms for precision voltage reference - Multiplied by an undisclosed constant

The period of the enable waveform is 5.1ms to minimize standby current and capacitor decay. Previously, the PVR was off during Standby. Therefore, enabling the PVR every 5.1 ms will increase Standby current (Icc). Standby current is increased to 5.4uA with these sampling waveforms.

The sampling waveform is high long enough for the worst case REF2 to refresh to a proper value. PVREN must remain high for 50ns more than SAMPLE to avoid the

REF2 node being pulled down when the precision voltage reference is disabled. This timing takes into account the propagation delay from the oscillator to the precision voltage reference. The simulation to determine the propagation delay from the oscillator to the precision voltage reference is shown in Figure 19. FTRCs are used to model the interconnect from the output of the oscillator to the precision voltage reference sample pass-gate transistor. The delay of 170ns is measured from 50% rise of SAMPLE to 50% rise of the input to the pass-gate (M100). Internal gate delays inside the precision voltage reference on both SAMPLE and PVREN also add buffer times.

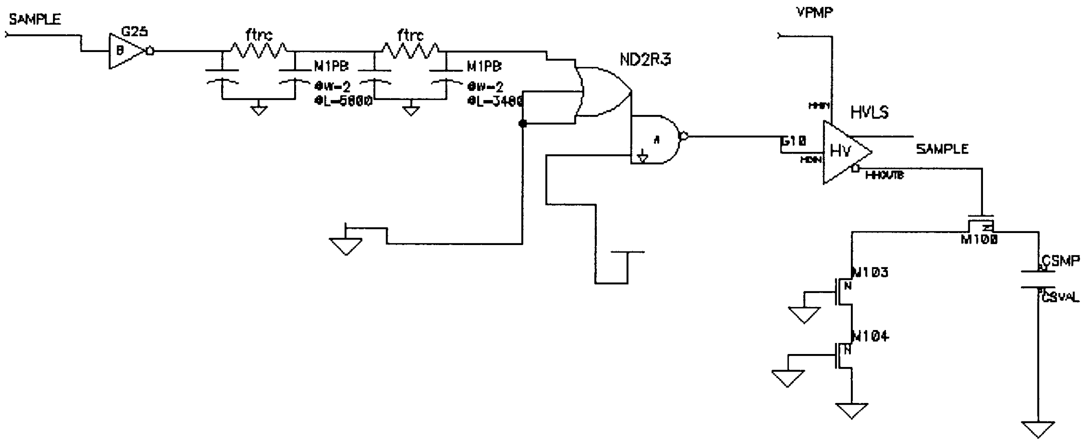


Figure 19: Model to calculate prop. delay from oscillator to PVR

Reverse bias leakage, or leakage current, between diffusion regions and substrate is a concern for the precision voltage reference.[10] Transistor source and drain diffusions

[10] Weste, Neil and Kamran Eshraghian. Principles of CMOS VLSI Design, p.231.

and n-well diffusions form parasitic diodes which can become reversed biased. A double guard ring has been added around the pass-gate transistor to minimize leakage. P+ and N+ guard rings act as "dummy collectors" for the reverse biased current, attracting the hole/electron current.[11] The leakage current for a 30pF capacitor has been measured on a test circuit. The results in Table 5 assure that leakage with a guard ring will be minimal, below 1mv/ms. Leakage may increase slightly due to the fact that the actual pass-gate used is slightly larger than that in the test circuit.

Temperature (C)	Leakage current	Leakage rate
25	$3X10^{-15}A$	0.1 uv/ms
85	$7.5X10^{-13}A$	25 uv/ms

Table 5: Leakage on passgate transistor, nominal discharge

The oscillator is always on; therefore the precision voltage reference is sampled in all modes (Active, Standby and Deep Power Down). The PVR could remain on at all times during Active mode. Instead, it is sampled for a more accurate REF2 and for simplicity, since it must be sampled during Deep Power Down and Standby.

The voltage divider gates from REF4 to REF2 are sized larger to allow a faster conversion during powerup.

Since I_{ds} α ß, increasing W/L should increase the current and speed up the division of REF4 to REF2.

[11] Weste, Neil and Kamran Eshraghian. Principles of CMOS VLSI Design, p.162.

$(I_{ds} = ß\ [(V_{gs} - V_t)V_{ds} - V_{ds}^2/2]$ linear region

$ß = \mu\varepsilon/t_{ox}(W/L))$

2.1.3 Results

The skews used in simulation are shown in Table 6.

<table>
<tr><th>VCC</th><th>Temperature</th><th>Process File Skew</th></tr>
<tr><td>3.3 v</td><td>25</td><td>mark</td></tr>
<tr><td>2.9</td><td>140</td><td>slow n slow p</td></tr>
<tr><td>2.6</td><td>100</td><td>slow n slow p</td></tr>
<tr><td>3.7</td><td>-40</td><td>fast n fast p</td></tr>
<tr><td colspan="3">Below are skews for pump simulations. They are needed for circuits which can vary if the N and P speed ratio is changed.</td></tr>
<tr><td>2.9</td><td>140</td><td>slow n fast p</td></tr>
<tr><td>2.6</td><td>100</td><td>slow n fast p</td></tr>
<tr><td>2.9</td><td>100</td><td>fast n slow p</td></tr>
<tr><td>2.6</td><td>100</td><td>fast n slow p</td></tr>
</table>

Table 6: Skews for simulations

Simulations were done to verify the correct logic for the precision voltage reference. At first, the precision voltage reference was simulated without initializing the load and capacitor. However, it would have taken a long time for the simulation to run and charge up all of the load and the capacitor. In true operation, the load will already be charged up when the PVR is turned on to refresh the sampling capacitor. Therefore, the load model and sampling capacitor were initialized to a set voltage. This was done by using a power supply attached to the load and capacitor through a switch. The switch was turned on for about 300ns, long enough to charge the load.

All of the load in the model was attached in order to simulate the worst case, when all other circuits are requiring REF2. The negative pumps also draw current from REF4, which is simulated in the model by setting a voltage on the pump model to draw a DC current.

The input PVREN was set high to enable the precision voltage reference. After 2us, SAMPLE was enabled to sample and refresh the capacitor. The output waveforms were examined to identify the time needed to enable the precision voltage reference for warmup and the time needed to sample the capacitor and recharge back up to 2v. The voltages of 1.96v, 2.06v, and 2v were used as the initial charge on the load. These reflect the worst case of 3% maximum variation on REF2. All skews are in Figure 20. The actual sampling waveforms were determined from these simulations.

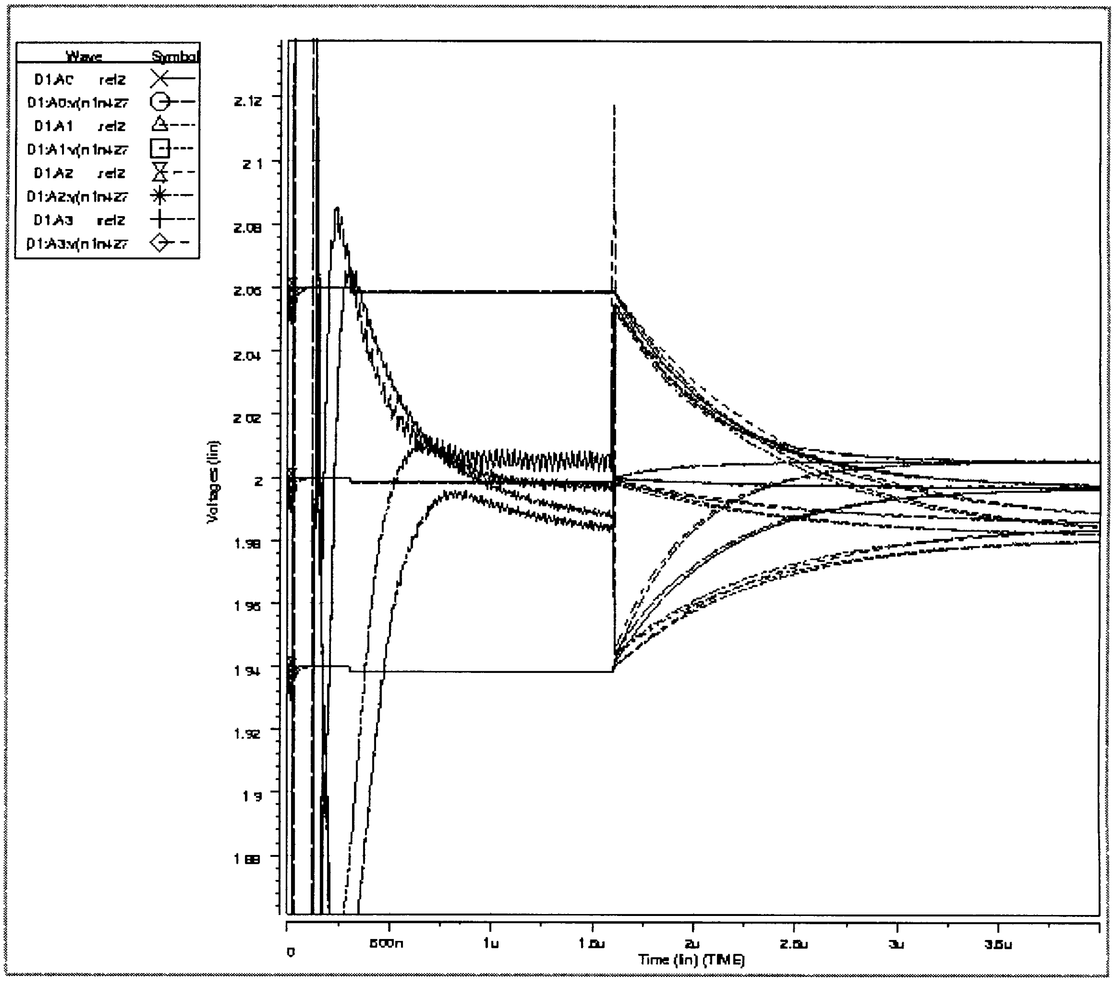


Figure 20: Simulation to find sampling waveforms for PVR

Once the sampling waveforms were determined, simulations were done to verify the timings for the sample and hold. The verification simulations are shown in Figure 21. When SAMPLE is disabled from high to low, a small drop of ~5mv is seen on REF2. This drop is inconsequential since REF2 still remains within 3% of its required value. The drop could be due to noise or leakage on the pass-transistor.

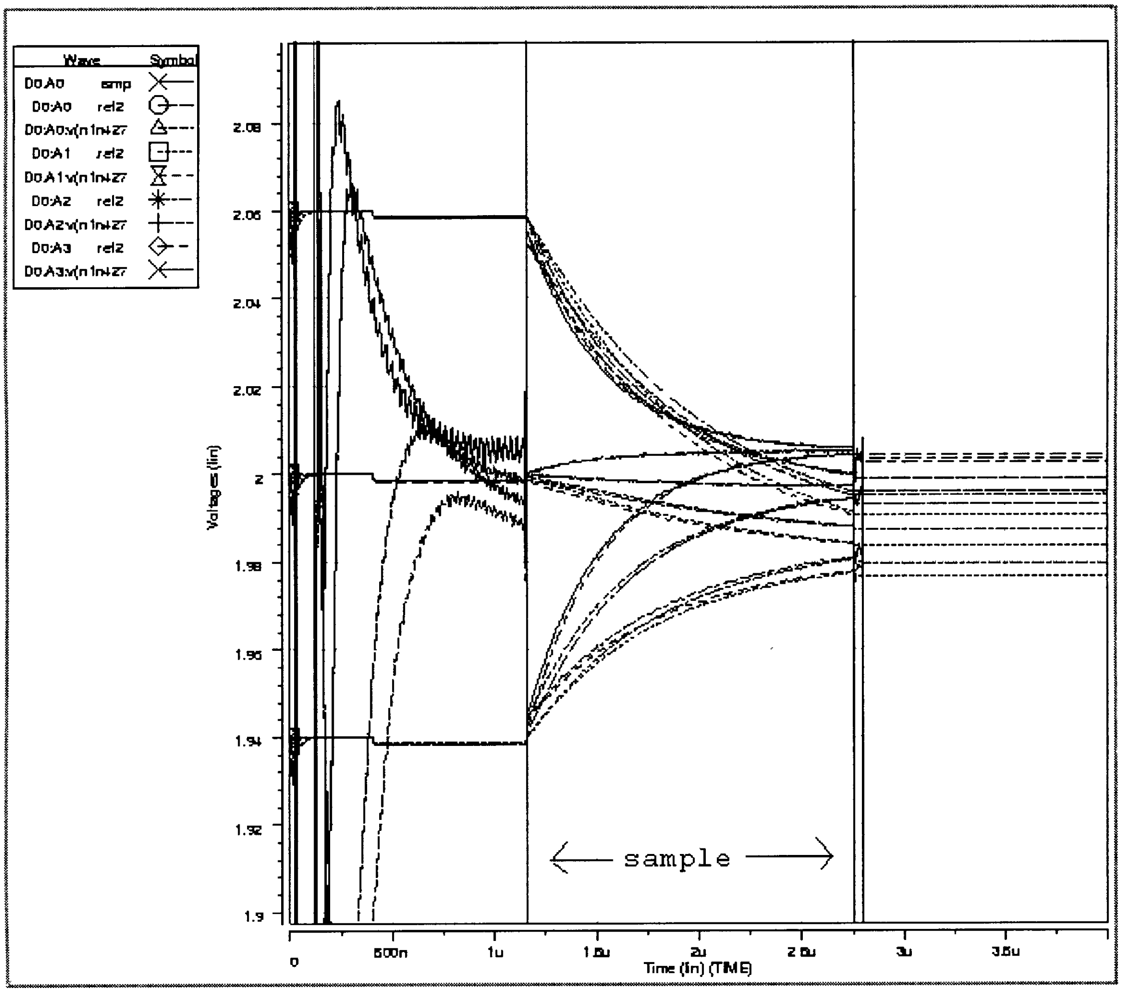


Figure 21: Simulation to verify sample times of PVR

The standby current for the precision voltage reference was also measured. The precision voltage reference was cycled on for the enable time, and the average VCC current burned was measured (Figure 22). The precision voltage reference burns 0mA when off and 3mA when on in the worst case. In the past, the precision voltage reference has not been enabled in Standby mode. Thus, implementing sample and hold will increase the standby current, since the precision voltage reference will burn current every time it is enabled. The current in the worst case (fast n fast p, T=-40, VCC=3.7) is increased by 5.4uA. This is derived from a weighted average:

[(3mA X 9.18us) + 0mA X 5.1ms] / 9.18us + 5.1ms = 5.4uA.

The increase is a small price to pay for the speed increase in wordline select time.

The CAM option implementation of the PVR also had to be tested and verified. Using a CAM option allows the PVR "sample and hold" to be tested by simply setting SELSMPHLD high. This ensures that the original circuits without "sample and hold" could still be used if desired. Unfortunately, the initial design produced undesired results.

Simulations were then run to attempt to minimize the apparent discharge on REF2 after sampling that occurred at slow n slow p, VCC=-2.9, T=140. The additions to the PVR for the CAM option are shown in Figure 17 and Appendix F-Appendix G.

When the SELSMPHLD bit is high, the PVR will act with the "sample and hold" feature. When the SELSMPHLD bit is low, the PVR will act without the new feature. The inverted signal of SELSMPHLD is ORed with the existing sampling signal SAMPLE. SELSMPHLD is ANDed with PVREN so that the oscillator will not falsely enable the PVR when SELSMPHLD is not active. Other additions are an extra voltage divider leg off of REF4 for faster charge of REF2, which is controlled by the SELSMPHLD bit.

An extra pulldown leg to reset node REF2 is also permanently added in. SELSMPHLD determines which pulldown leg off of REF2 is chosen.

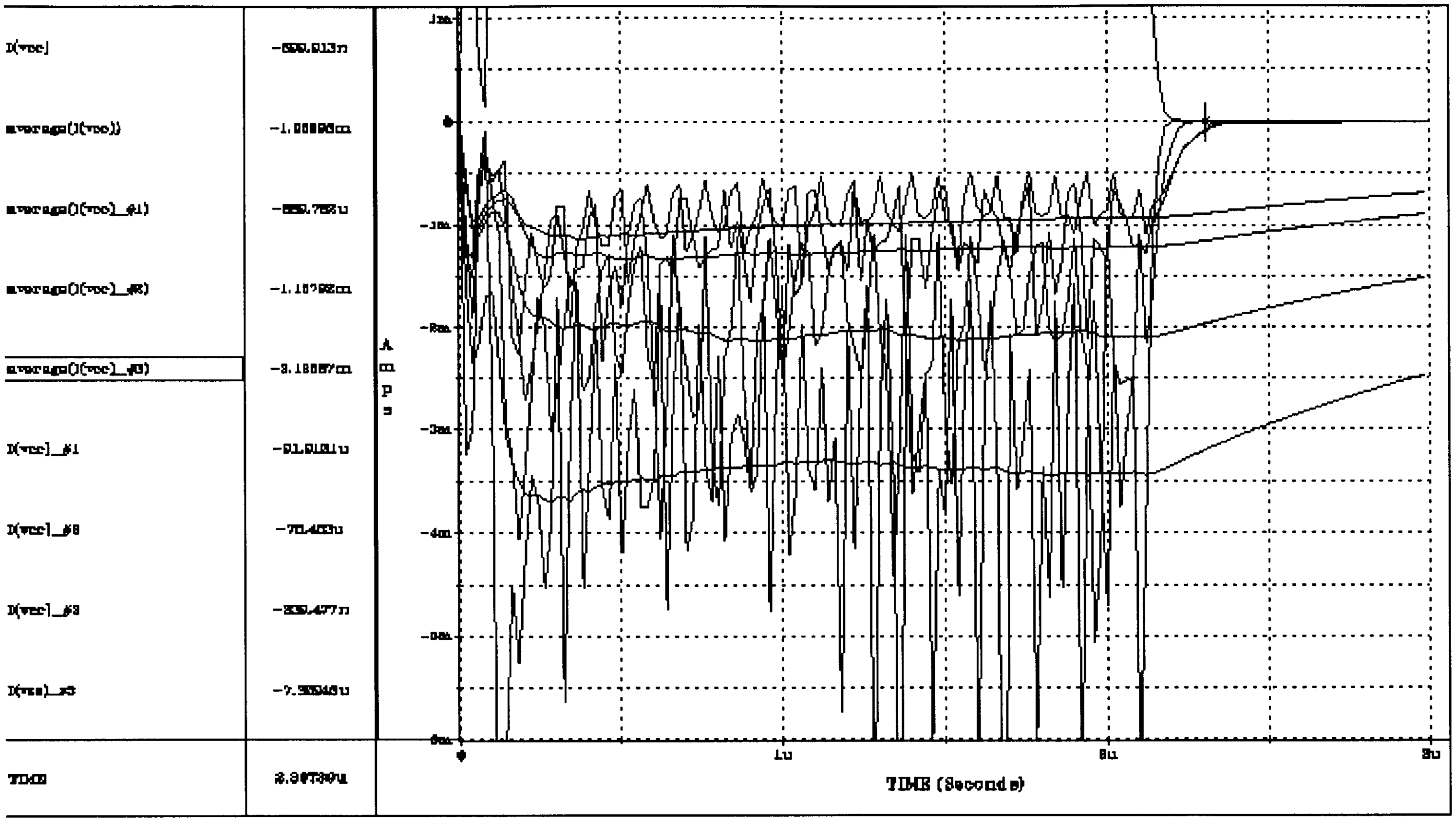
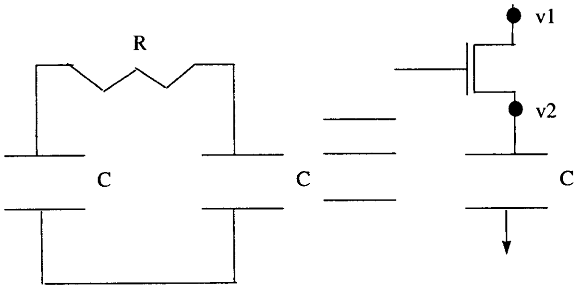


Figure 22: Measurement of current burned by PVR

It is important that the pulldown leg for the "sample and hold" configuration be placed before the sampling pass gate. Otherwise, the capacitor's charge will be reset to zero. An extra NMOS device is also added to control the connection of the sampling capacitor to REF2. A size of 50um for the NMOS device connected to the sampling capacitor appears to minimize the loss of charge. Anything higher does not allow enough current to pass to the capacitor, and a lower value causes too much leakage and capacitance on REF2. A metal option is included to directly connect the sampling capacitor to REF2. This is necessary because simulations at high temperatures shows that the REF2 node is losing or gaining charge even after sampling. This phenomenon is due to the NMOS connected to the capacitor. At high temperatures, the mobility of the electrons decreases. Thus, the effective resistance of the NMOS gate increases. Therefore, the voltage at the drain and source of the NMOS gate is not the same even after the sampling has completed. Thus, the capacitor still "charges up" even after sampling is done. The model of this problem is shown in Figure 23.



At high temperatures, V1 is not always equal to V2; R increases.

Figure 23: Model of NMOS and capacitor in series

The other transistors connected to REF2 were all sized much smaller, and guard rings were placed to minimize leakage. The sampling transistor was also sized much smaller. The effect of a smaller sampling transistor can be seen by comparing the simulations in Figure 24: leakage is less with the smaller sized sampling transistor.

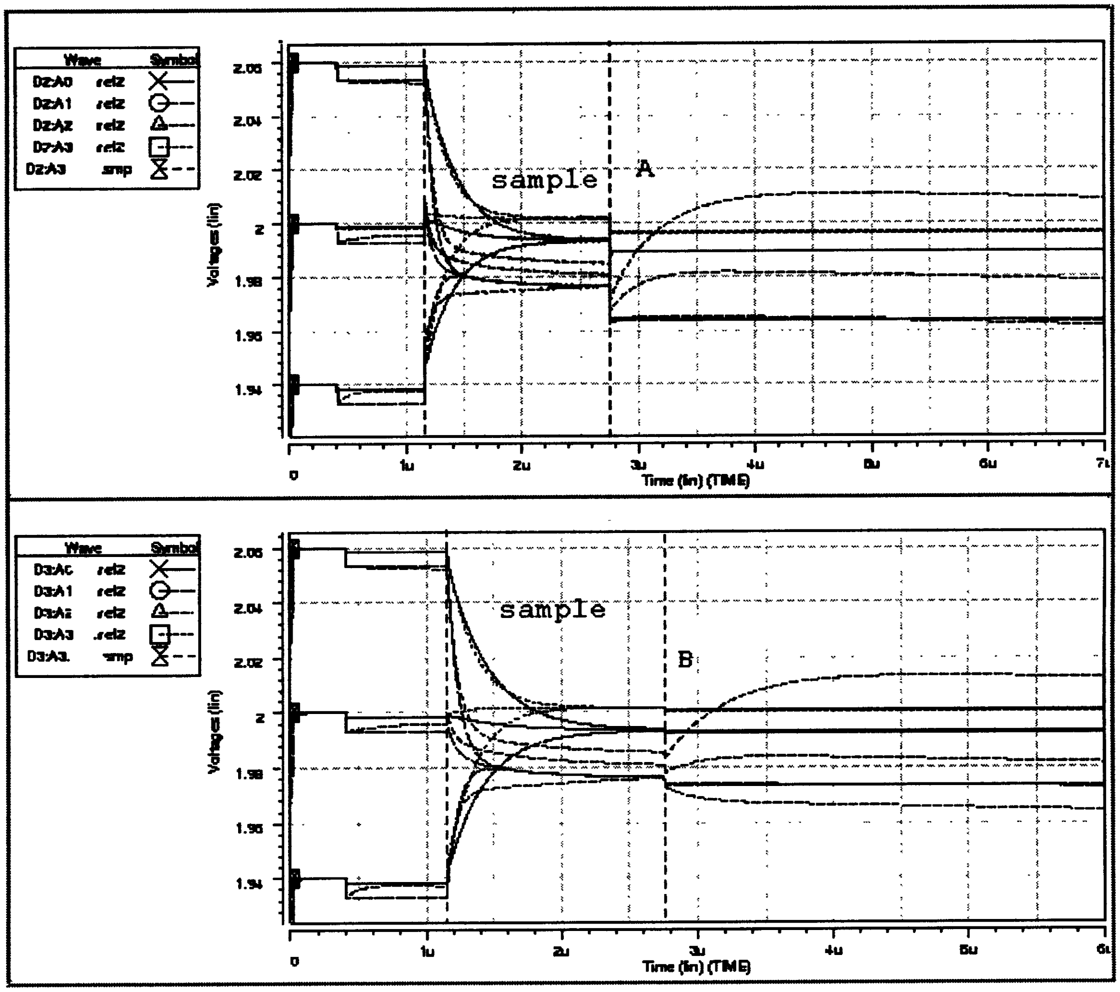


Figure 24: Test CAM option PVR - B has less leakage than A (smaller pass gate)

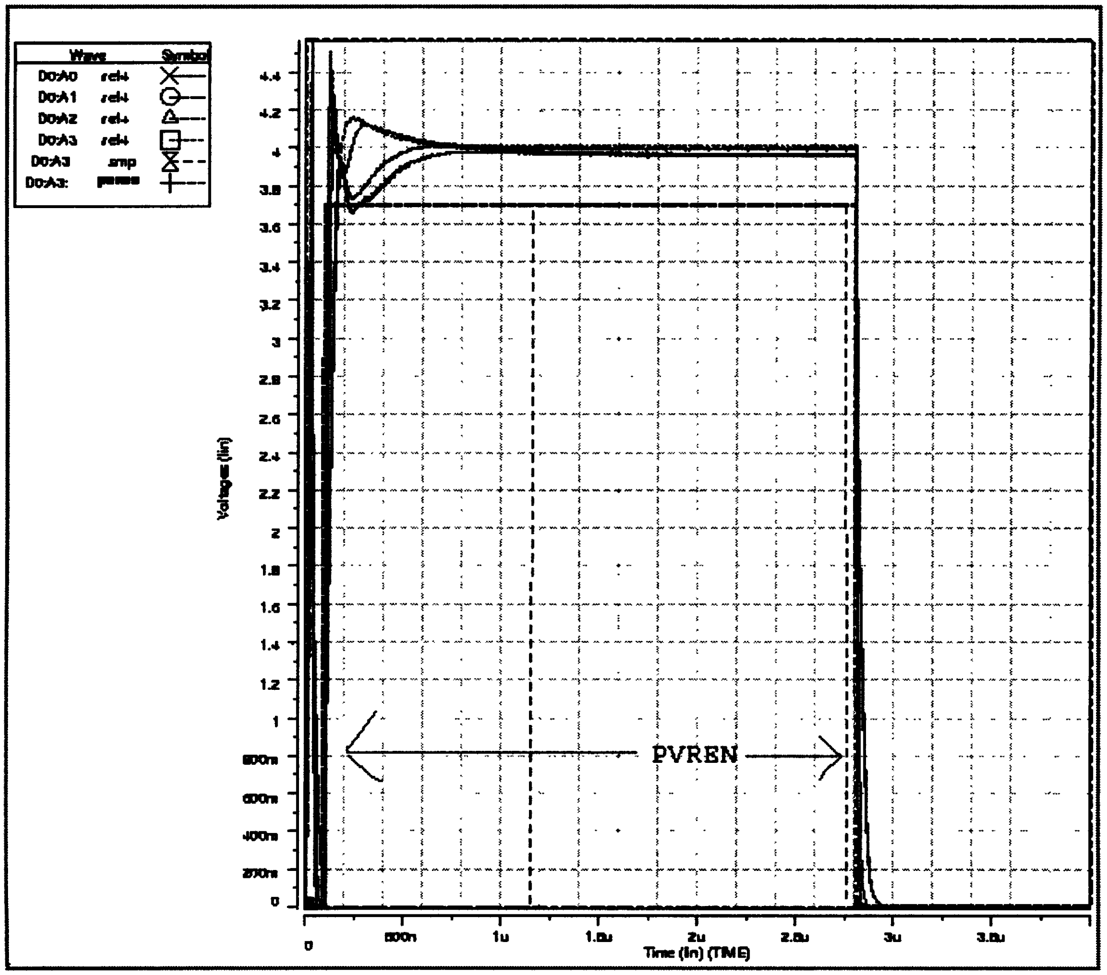
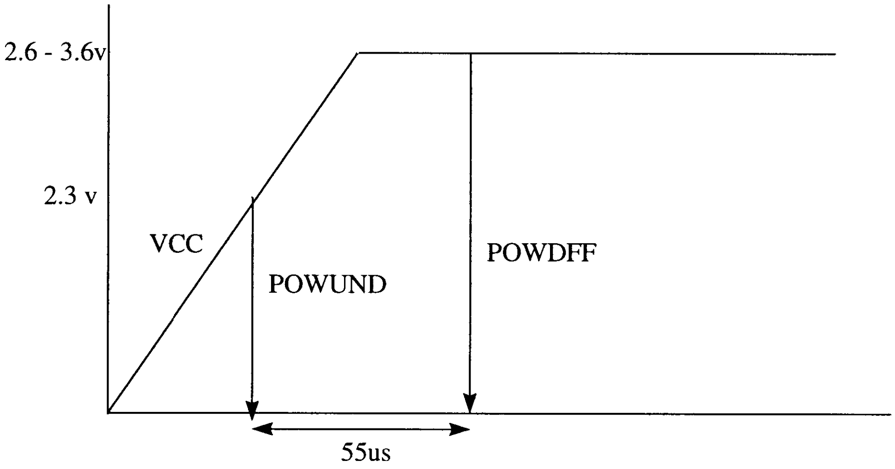


Figure 25: PVR, CAM off

A simulation was also run to verify the old functionality of the PVR when the CAM is off (Figure 25).



POWUND = 55us after POWDFF
POWUND always trips at 2.3V
VCC ramp can vary from 1v/10us to 1v/1ms

Figure 26: Vcc powerup scheme

VCC ramp-up (powerup) was simulated for the precision voltage reference. Figure 26 shows the input vectors for powerup. It is important that the precision voltage reference is able to rise and function at the correct values within the time of VCC ramp-up. This is because other circuitry need the outputs of the precision voltage reference, so it must be operational at powerup.

The typical rampup time for VCC is 1v/10us. The worst case rampup time is 1v/1ms. The latter is worst case because POWUND always drops at 2.3v, and POWDFF always drops at 55us after POWUND. Therefore, if VCC was rising at 2.6v in 2.6ms, then POWUND would rise at 2.3v in 2.3ms. POWDFF would drop at 2.355ms. This only gives the precision voltage reference 2.355ms to warm up at a point when VCC will only be at 2.3V. There are potential problems with 1v/1ms, which has not been previously tested for "sample and hold." Previously, "sample and hold" was implemented on a 5v chip. Therefore, the voltage level applied to the PVR was much higher than for the current 3v chip. Simulation results first indicated that the PVR was not operational at 1v/1ms. However, when the previous version of the PVR was simulated under the same conditions, it also appeared to fail 1v/1ms. This may have occurred because the skew was not tested in GEAR method (explained in Miscellaneous section). Since both the previous simulation and simulation with new additions appeared the same, the issue was determined to be resolved. More accurate results could probably be achieved with a more correct load model (less aggressive). The main goal in simulating the new additions is to ensure that nothing has changed from the previous circuitry, which has been proven to work in lab and in real-life. The result of a 1v/10us vcc ramp for the precision voltage reference is in Figure 27.

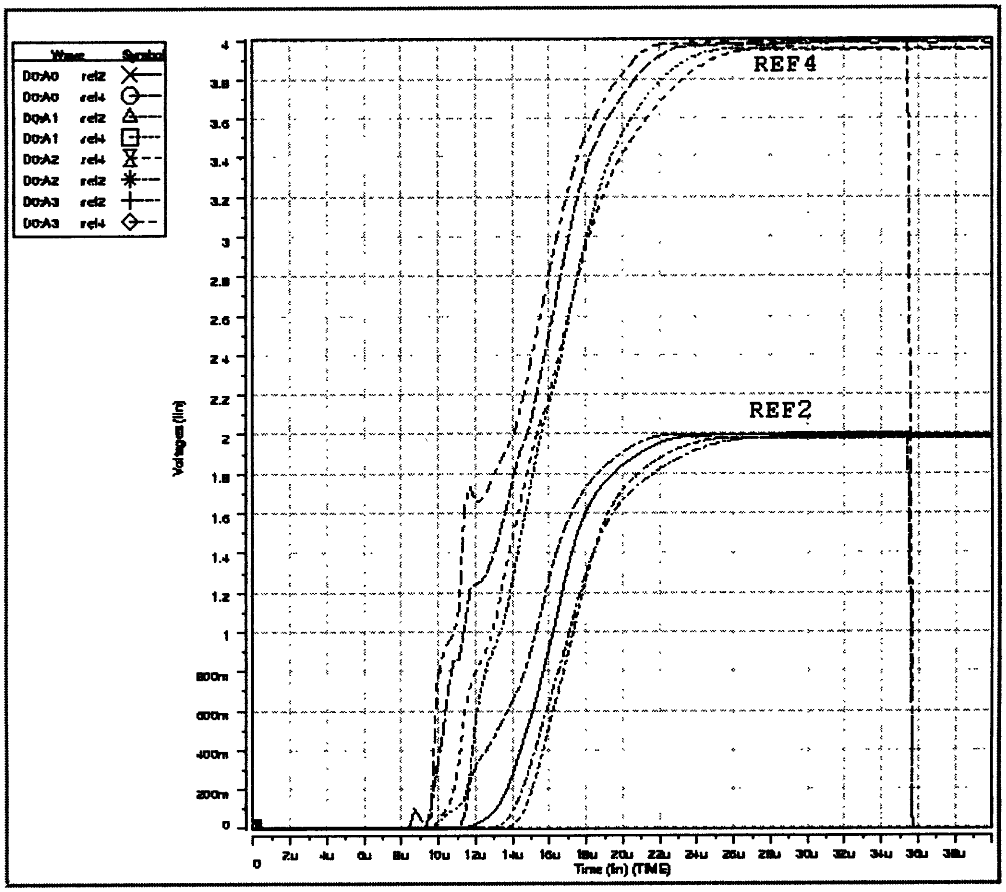


Figure 27: PVR powerup - 1v/10us all skews

2.2 Active Pump

2.2.1 Introduction

2.2.1.1 Purpose and functionality

The active pump is enabled during active mode for pump regulation. Historically, the active pump has used REF2 during powerup as its input reference and then switched to VCC reference for all operations. Figure 7 shows the block diagram of the active pump's interaction with the pumps and wordline. The positive input to the differential amplifier in the active pump is connected to either VCC reference or REF2. The negative input to the differential amplifier of the active pump is connected to output of the regulation dividers, which regulate the positive pump output. In other words, the active pump provides the regulation feedback for the positive high current pump during a read. The positive pump, in turn, outputs HSRC5DRV to the switches (level shifters), which output HSRCDRV to the wordline.

The active pump needed to be altered to always use REF2 as its input instead of using VCC. Originally, it was assumed that a simple change to the active pump logic would suffice. However, the active pump logic turned out to be quite complex, activating many different modes of the active pump. Also, the regulation of the pump itself needed to be altered.

Simulations were done to determine the correct voltage divider ratio for the pump regulation. The initial simulations were done to understand the powerup of the active pump. It was determined that there are many modes of the active pump: DPD powerup, Stby powerup, Active powerup, Active, Stby, and DPD. Therefore, the regulation changes depending upon what mode the chip is in during turn on. For DPD powerup, all regulation comes on initially, using REF2. Then all regulation is turned off. During Stby powerup, all regulation is on initially (using REF2) and then only the standby regulation divider stays on. The VCC regulation is turned on by a pulse from the oscillator. For active powerup, all regulation is on initially (using REF2) and then VCC regulation is used. Active, Stby, and DPD modes are similar to their powerups, except that they do not use REF2. The signals from the active logic which control these regulation dividers are: HBDIVV1 = standby divider, HDIV2B= use ref2 and ref2 divider, HDIV3B= active/VCC divider, CHVCCB= use VCC. Figure 28 shows the different powerup modes of the active pump.

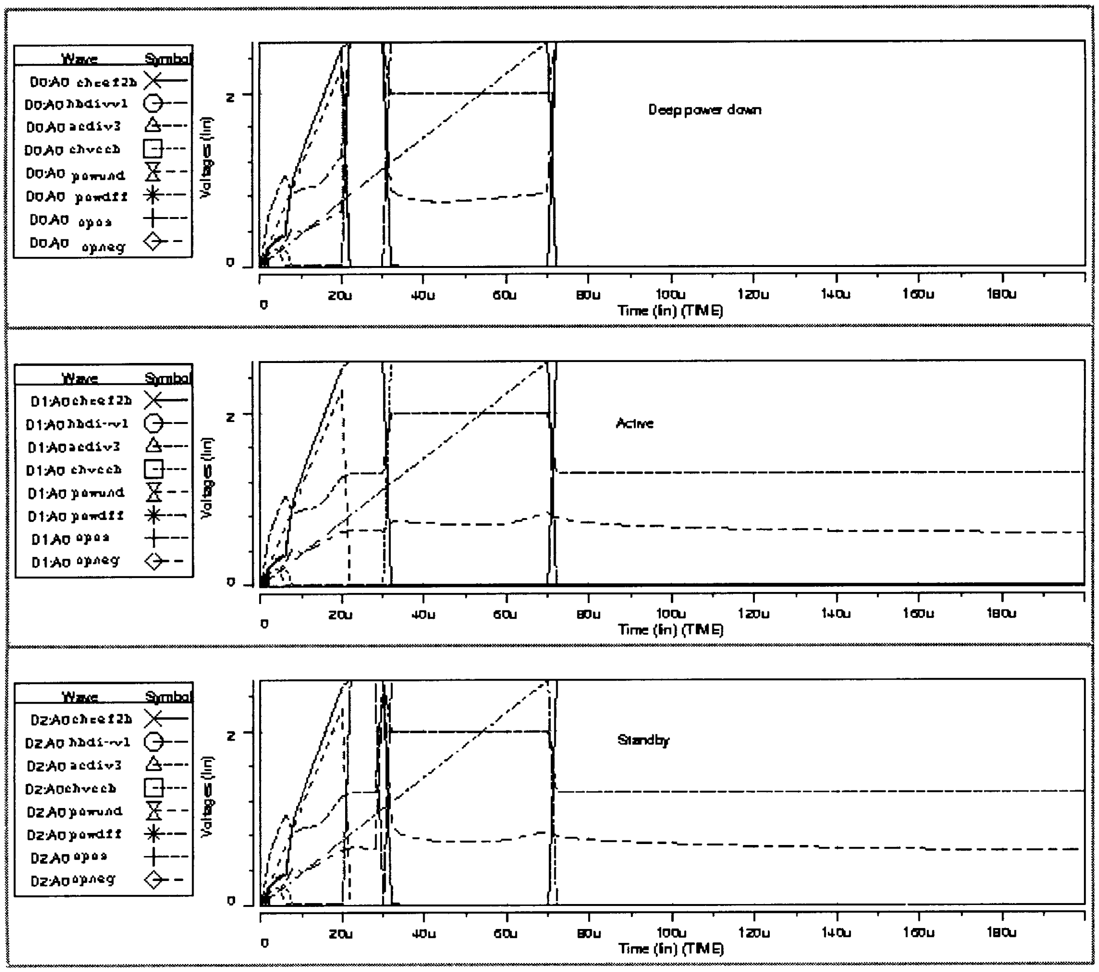


Figure 28: The different powerup modes of the active pump

2.2.1.2 Inputs and Outputs

Table 7 shows the inputs and outputs of the active pump.

Inputs	
ACDIV3	turns on active regulation dividers (in read mode)
HSRC5DRV	regulated by active pump
HBDIVV1	turns on standby dividers (in all modes)
CHVCCB	select VCC regulation (old way)
CHREF2B	select PLRF2 regulation (old way)
SELSMPHLD	test bit to test CAM option
REF2	inputs 2v
Outputs	
OPOUT	regulated output to positive pumps
OPPOS	positive output of diff amp
OPNEG	negative output of diff amp
HSRC5DRV (indirectly)	output to switches → wordline HSRCDRV

Table 7: Active pump inputs and outputs

2.2.2 Method / Results

The original approach was to change the active logic to always use ref2 by setting CHREF2B = 0 and CHVCCB = 1. However, the dividers were not adjusted to the desired value of HSRC5DRV. Also, it was imperative that the original modes and powerup of the chip not change. Therefore, since the VCC and ref 2 dividers are always turned on at powerup, it is easier to simply adjust the VCC divider and the standby divider. In this way, the original logic can still function the same way. The ref2 divider originally added in is detached, since the other dividers are tweaked for ref2 regulation. Also, the input to the diff amp is changed to always take REF2 instead of choosing between REF2 and VCC.

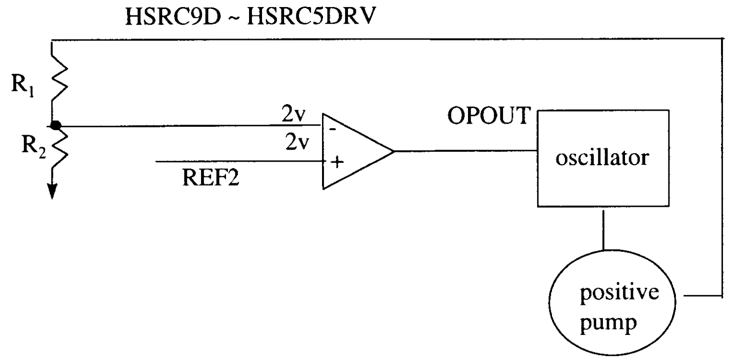


Figure 29: Diagram of regulation circuit

Two different simulations were run to size the new regulation dividers. Both the standby and active regulation dividers were sized the same to avoid contention. The two simulations were run using a model of the regulation circuit, similar to Figure 29, minus the oscillator and positive pump. After the initial sizes were found, simulations including the oscillator and positive pump were then run with the new sizes. In the first simulation (Figure 30, top), the widths of the gates in the divider string were input as a SWEEP variable. HSRC5DRV was set to the desired value of 5.4v. The correct width was found where the output into the differential amplifier crossed 2v. This width=7.23um had the danger of varying greatly over process variation. To limit variation over process, it would be ideal to have all of the gates in the divider be the same width and length. However, this is difficult to achieve for HSRC5DRV = 5.4v.

The second simulation (Figure 30, bottom) attempted to verify the value of W=7.23. HSRC5DRV was varied in a SWEEP and plotted against OPOUT. The trip point occurred at HSRC5DRV=5.4v, confirming the size.

If the dividers were constructed with resistors instead of CMOS gates, it would have been easy to estimate the correct size since one could use voltage divider relationships. However, CMOS gate voltage does not vary linearly as one varies the width; therefore it is difficult to estimate by hand calculations.

$I_{ds} = ß\ [(V_{gs} - V_t)V_{ds} - V_{ds}^2/2]$ linear region

$ß = \mu\varepsilon/t_{ox}(W/L)$

$V_{ds}\ \alpha\ (W/L)^{1/2}$

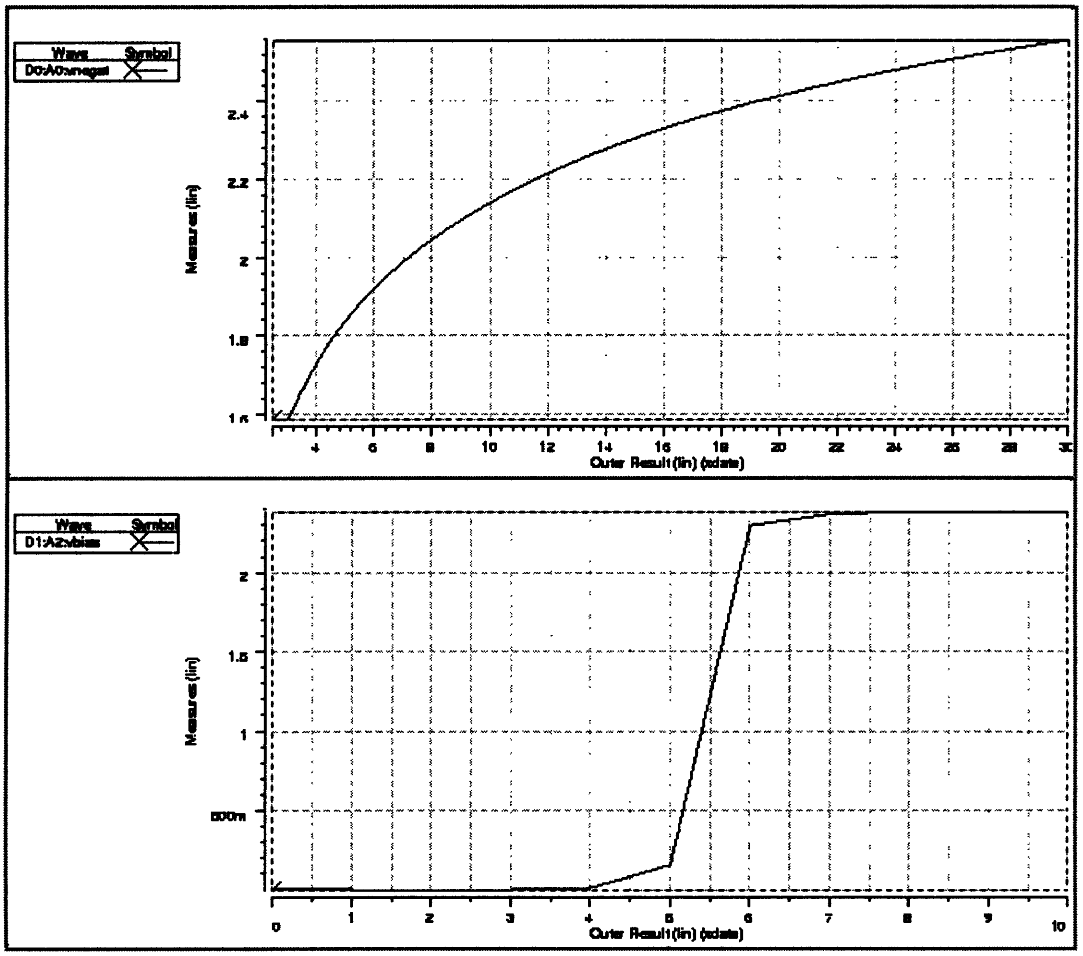


Figure 30: Sizing active pump: top = first simulation, bottom = verify

The current through the new dividers was also measured to ensure that it would not greatly increase during Standby mode. The active pump with metal optioned changes is shown in Figure 31.

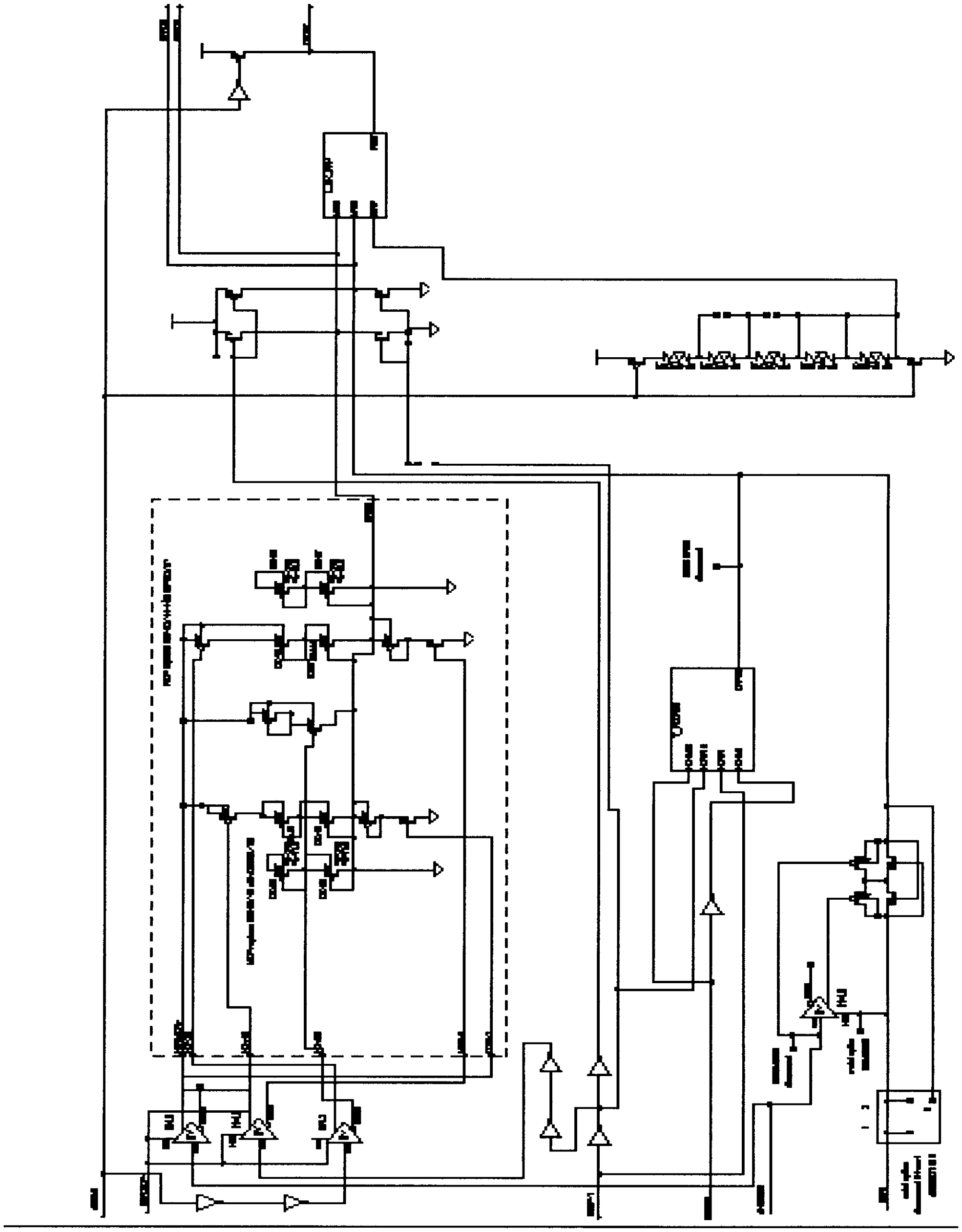


Figure 31: Active Pump in Metal options

Finally, the active pump was simulated in active mode (Figure 32 and Figure 33) with the positive pump and loading attached. The pump regulates to the correct value of

5.4v even with the loading of the other circuits. The loading also assists in smoothing out the ripple of the pump output. The target for HSRC5DRV can be from 5.1-5.4v, as was found in the simulations to determine the effectiveness of applying a direct voltage to the wordline.

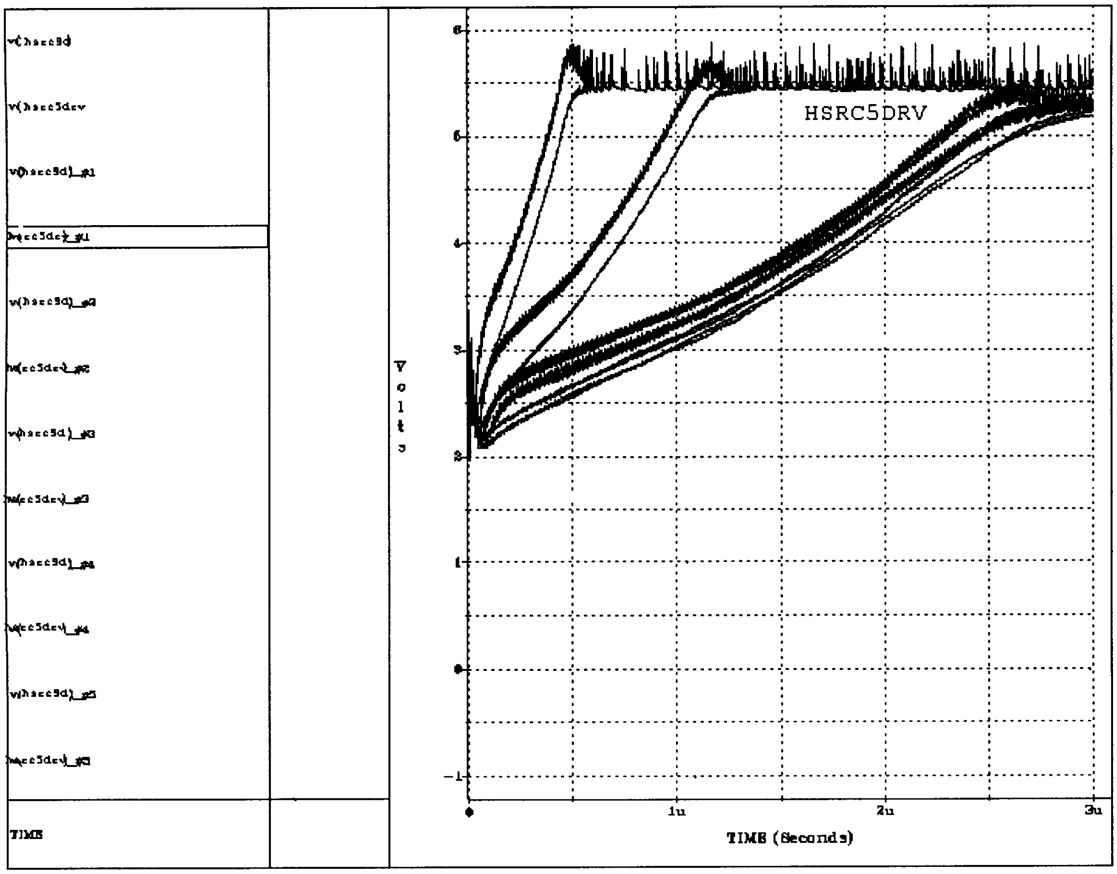


Figure 32: Active mode of pump, all skews (minus slow n slow p)

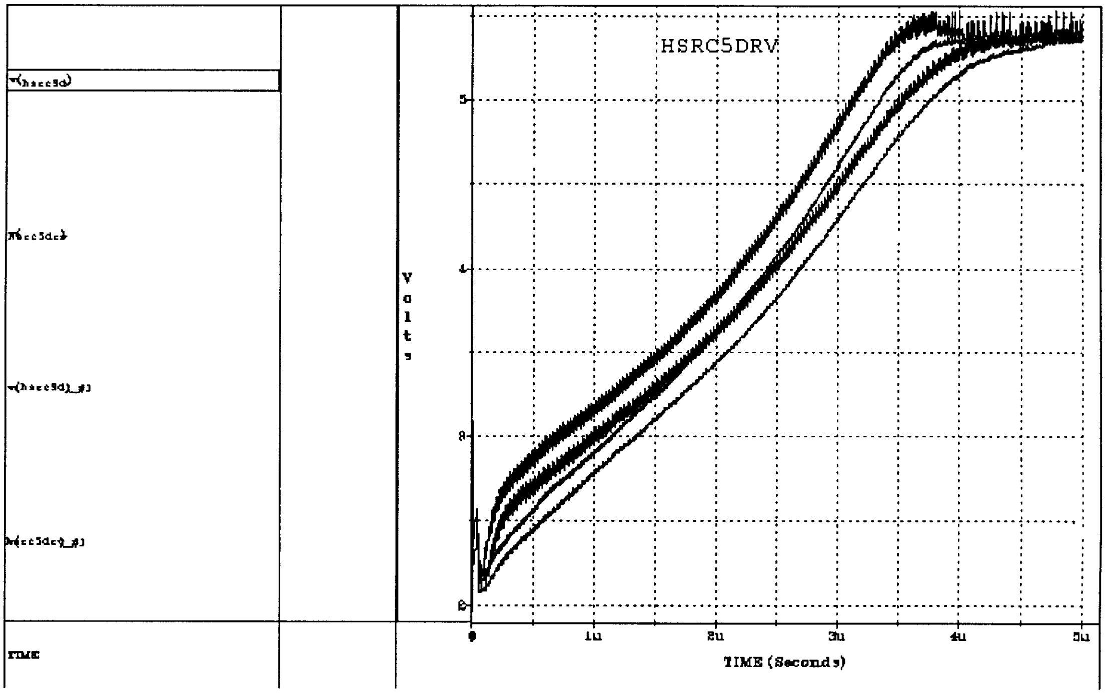


Figure 33: Active mode of pump, slow n slow p

The active pump and positive pump were also simulated in Standby mode to ensure that standby current was not increased. Since we are no longer using VCC regulation, standby current should not increase (less of a draw on VCC). The pump was simulated before and after the changes for comparison. The two simulations were similar. The old simulation had a current of 62uA while on (for 2us). The new simulation had a current of 60.4uA while on (for 2us). Therefore, since the pumps are on every 1ms for 2us, the Icc for both old and new active pumps = 60uA(2us)/2us + 1ms = 0.012uA. Thus, standby current was not altered by the changes to the regulation, as can be seen in Figure 34.

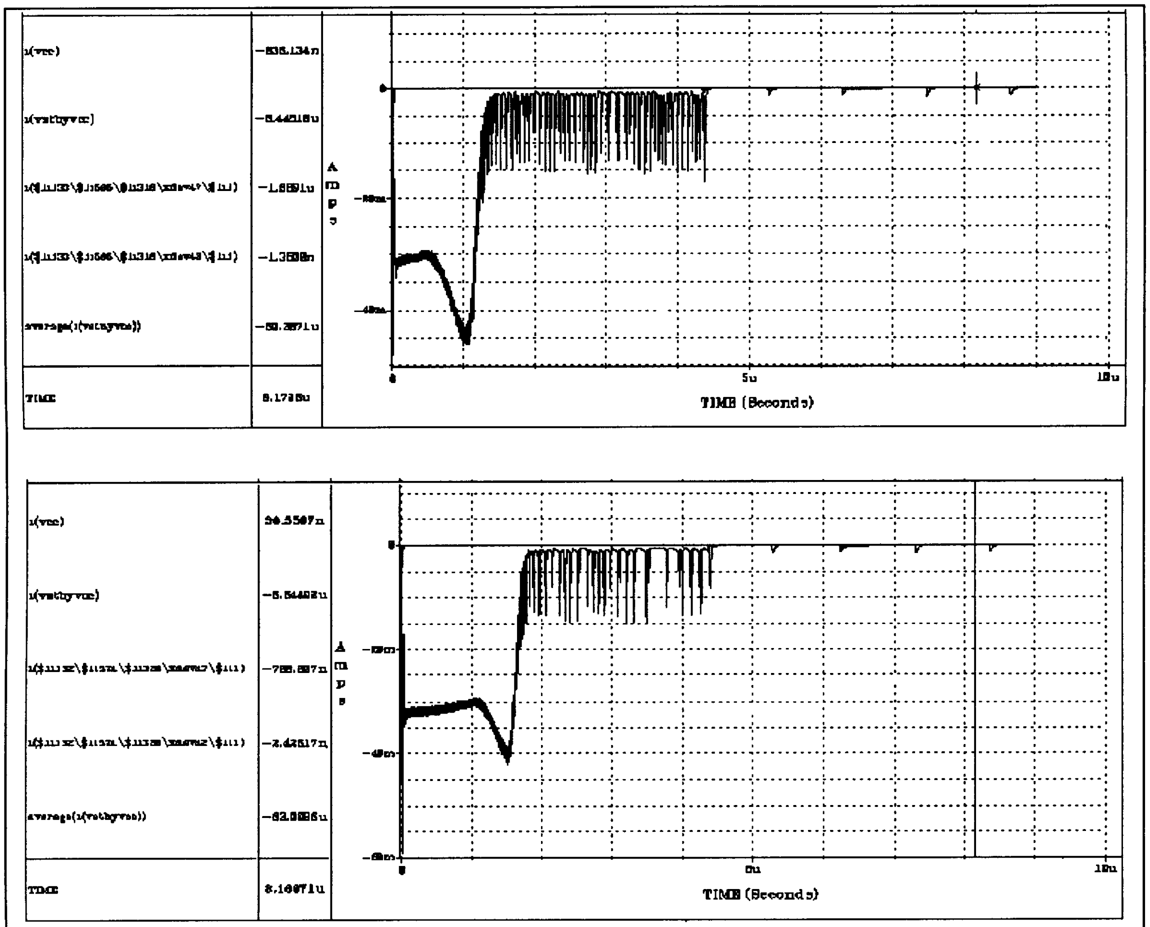


Figure 34: Standby current of pump top - new current bottom - old current

The active pump was also changed to a CAM option for ease of testing. The CAM option implementation of the active pump is pictured in Figure 35. A closeup of the regulation circuit is shown in Appendix H.

Simulations were run to verify the new circuitry and logic added. The SELSMPHLD bit was added as an input. When SELSMPHLD is high, the new regulation is used; when SELSMPHLD is low, the old regulation is used. SELSMPHLD is ORed with

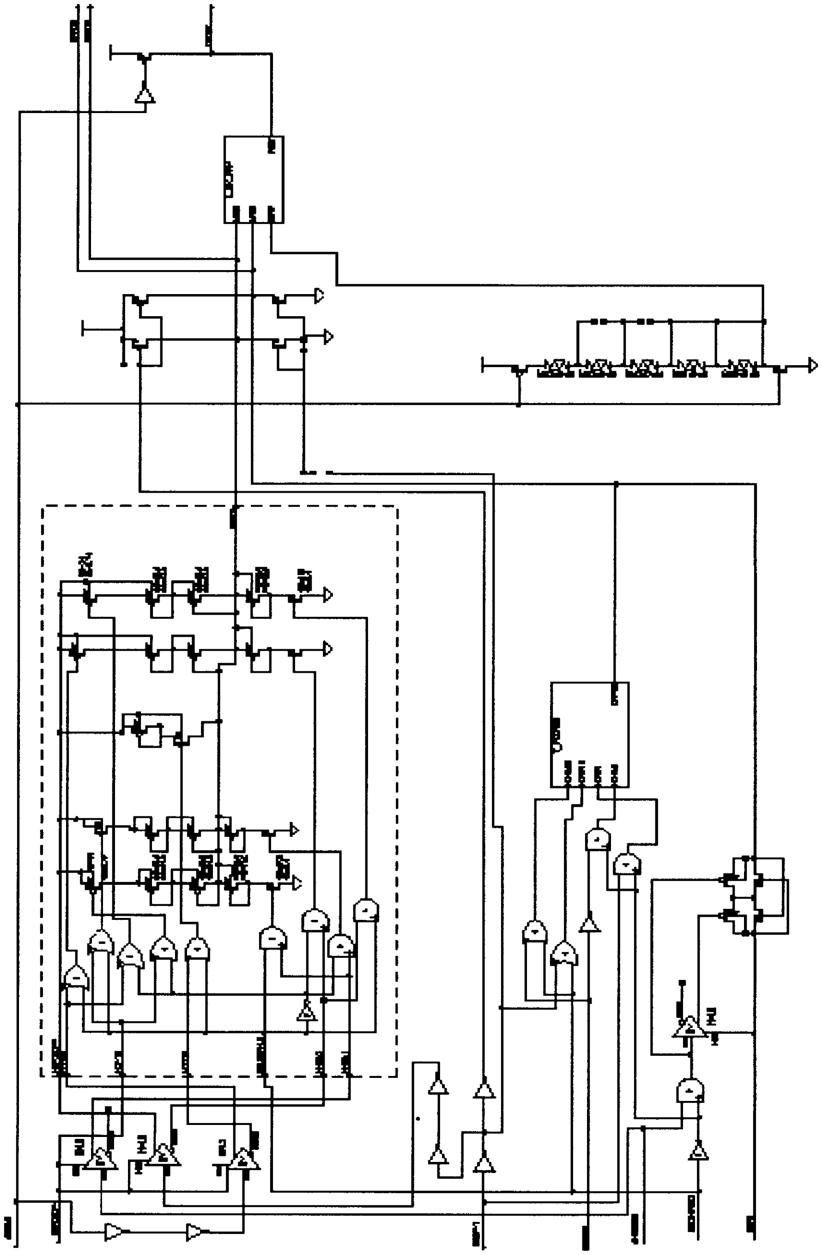


Figure 35: CAM option of active pump

the original inputs to the old regulation PMOS switches. Thus, these PMOS gates are only turned on when SELSMPHLD is off. The inverse of SELSMPHLD is ANDed with the original inputs to the old regulation NMOS switches. Likewise, these NMOS gates are only turned on when SELSMPHLD is off. The inputs to the new regulation PMOS and NMOS switches are similar, except the SELSMPHLD and its inverse are reversed. A similar concept is used as the new inputs to control the VCC vs. REF2 regulation. The simulations tested the active pump with the test bit set to ON and then with the test bit OFF (Figure 36 and Figure 37).

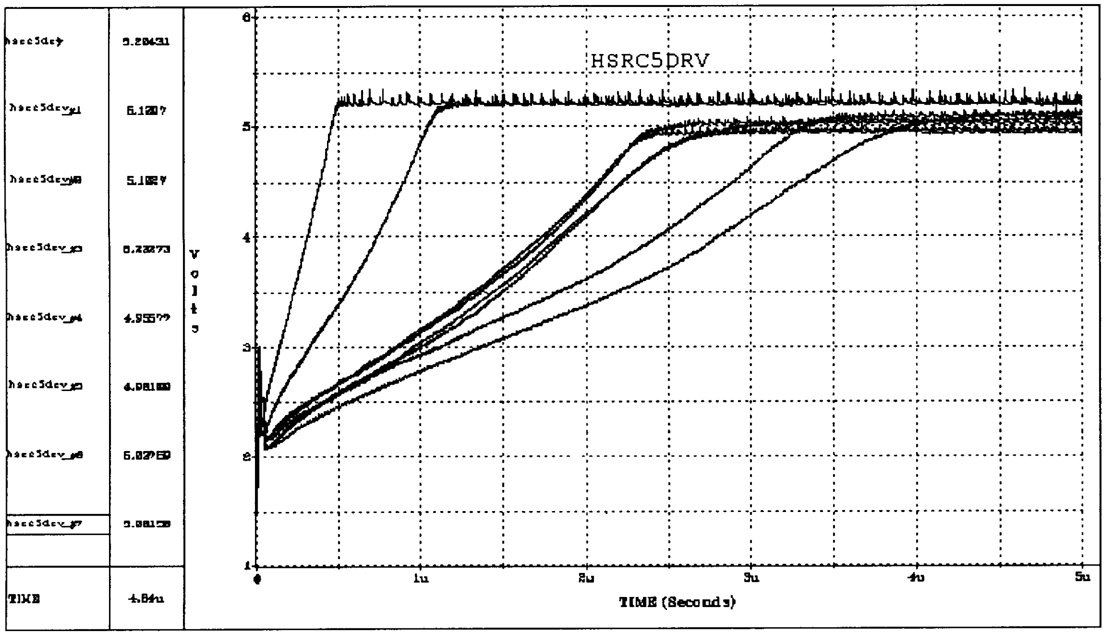


Figure 36: Active pump, CAM selected on

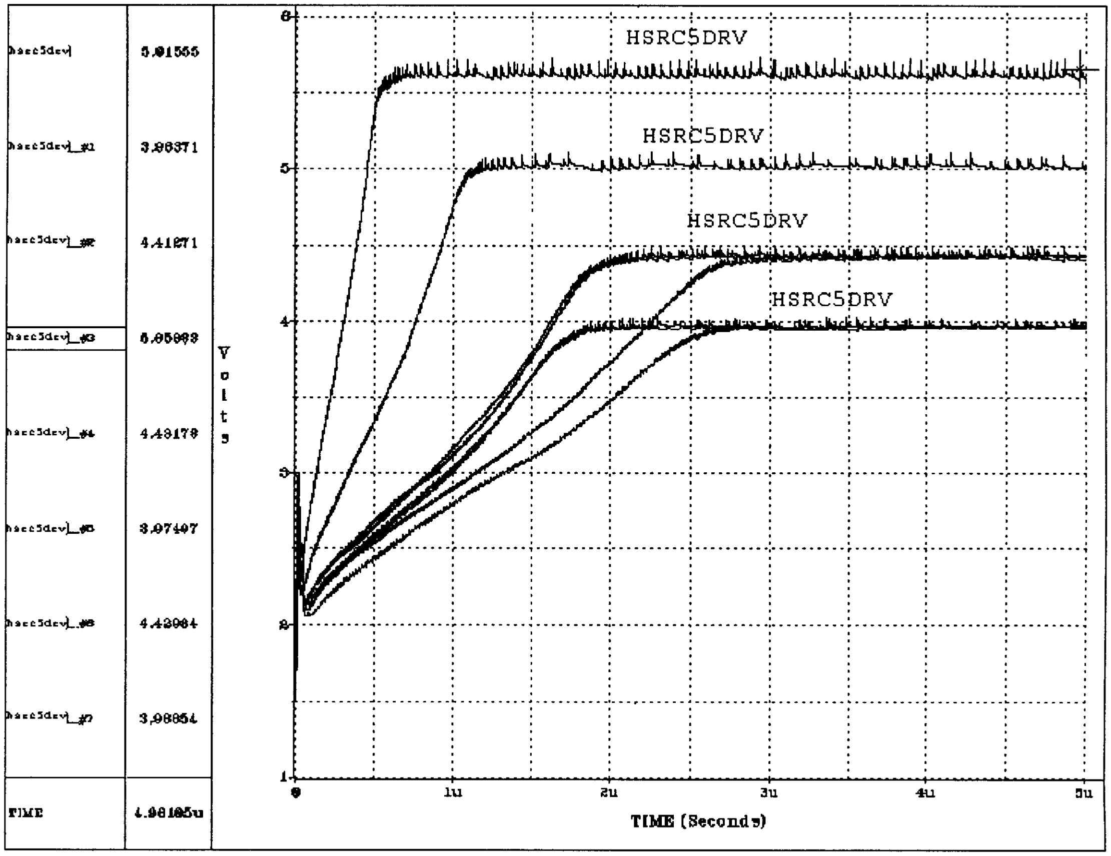


Figure 37: Active pump, CAM selected OFF

The results of the regulated voltage with the new CAM option were slightly lower than without the option. This could be because of added capacitance on the HSRC5DRV node from having all of the possible regulation dividers attached. The new regulated numbers still comply with the necessary voltage to improve read speed (5.0-5.2v) and was determined to not be an issue. Furthermore, once testing is done, it can be possible to take out the CAM option and directly wire the circuits to obtain the more precise results.

Powerup of the active pump is shown in Figure 38:

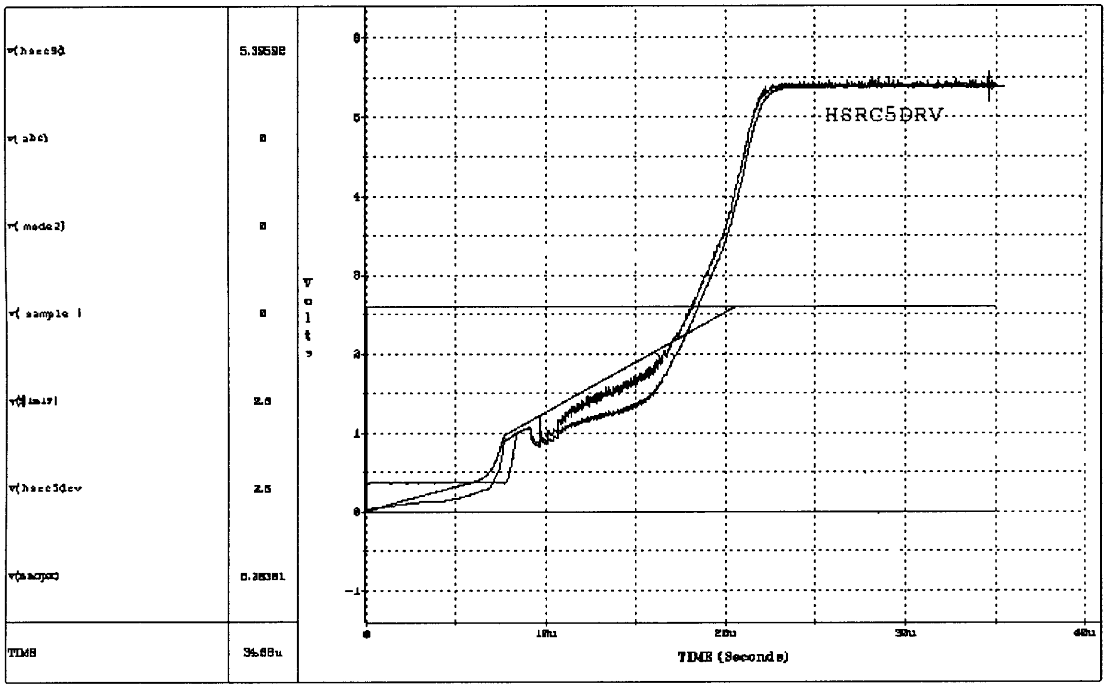


Figure 38: Powerup of Active pump with positive pump at 1v/10us

2.3 Miscellaneous

Several problems were encountered when dealing with HSPICE. For example, when running long simulations, it was important to set .options METHOD=Gear. This is due to the fact that HSPICE cannot handle the oscillations from the oscillators of the pumps. In its iterations, it fails to produce a valid answer and often will output a value below what is expected. The graphical viewer Metawaves also had problems processing large output files. Therefore, the simulations were done using both Metawaves and Viewtrace plotter tools.

Another common problem found using HSPICE was non-convergence. One solution to this problem is to ramp inputs such as VCC and to set as many initial conditions as possible. Otherwise, HSPICE has too many nodes to calculate and initialize in a short amount of time. Other parameters can also be set to help convergence, but it may lessen accuracy and should be used with caution. These settings are:

```
.OPTIONS mbypass=0.1
.OPTIONS rmin=1e-12 absmos=20u absvar=2 relmos=.2 relv=.5
.OPTIONS DCSTEP=.1 VNTOL=.001 reltol = .01
```

HSPICE also cannot model leakage off of nodes very well. The leakage rate instead needs to be determined from test circuits. Since HSPICE is only a tool to assist engineers, there is plenty of room for judgement as to the accuracy of results. Often the results will seem much worse than they are in reality, while in other cases the results will not be worst case.

3. 1v op amp[12]

3.1 Introduction

3.1.1 Purpose and functionality

A 1v operational amplifier is important to the future of Flash Memory. In order to meet the demands of mobile electronic devices such as cellular phones, PDAs, and cameras, it would be ideal to have circuits that can be operational at a low voltage. A 1v supply voltage would be equivalent to battery power; circuits designed at 1v could be used in widespread low voltage applications. The op-amp is important particularly to the Precision Voltage Reference circuit, which depends upon an op-amp, flash cells, and a pump to operate. Therefore, the first step in getting the PVR to operate at low voltage is to investigate possible methods of creating a 1v operational amplifier. This op-amp would also be useful as a differential amplifier in the active pump, which compares a 2v reference to a series of voltage dividers in order to provide regulation for the positive pumps.

An operational amplifier has high forward gain. Ideally, one would like an op-amp to have infinite input resistance, infinite differential voltage gain, and zero output resistance. In other words, $V_{out} = A\ (V_+ - V_-)$

However, the typical op-amp is not ideal. $V_{out} = A_v(V_+ - V_-) + A_c\ [(V_+ + V_-)/ 2]$

[12] One or more patent applications have been filed by Intel Corporation covering portions of the material covered in this thesis.

The first quantity is the differential gain, while the second is the common-mode gain. A typical differential amplifier is shown in Figure 39.

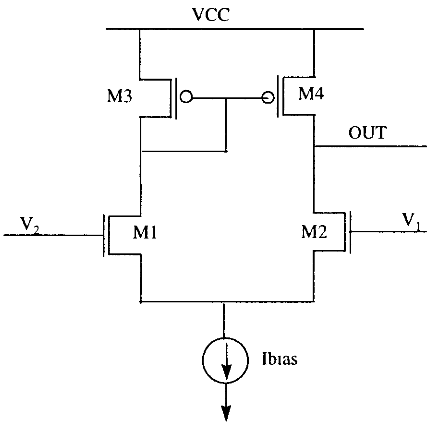


Figure 39: Traditional differential amplifier

M3 and M4 are current loads and M1 and M2 are the differential pair. Ibias is the current source for the amplifier. The current in M1 determined the current in M3. This current is then mirrored in M4. Thus, if $V_{gs1}= V_{gs2}$, then $I_1 = I_2$ and $I_{out} = 0$. However if $V_{gs1} > V_{gs2}$, then $I_1 > I_2$ since Ibias = $I_1 + I_2$. An increase in I_1 implies an increase in I_3 and I_4. Since $I_2 = I_{out} + I_4$, I_{out} = positive. If $V_{gs1} < V_{gs2}$, then Iout = negative. Thus V_2 is the positive input and V_1 is the negative input to the differential amplifier.[13]

There are some limiting factors to consider when designing op-amps. The first is the slew rate, which is the maximum current available to charge or discharge a

[13] Allen, Phillip, and Douglas R Holberg. CMOS Analog Circuit Design. New York: Oxford University Press, 1987, pp.273-276.

capacitance. In other words, this reflects the fastest that the op-amp can go from one voltage level to another when a step input is applied. The slew rate measures the op-amp's maximum output current sourcing and sinking abilities. Another limit is the settling time, or the time for the output to reach a final value when excited by a small signal. Finally, there is a limit on the output voltage range capability.

Another important feature of the op-amp is its Common Mode input Range (CMR). This is the voltage range over which the input signal can vary for the op-amp to continue to operate. It is important to design an op-amp that can have a large CMR, close to the rails of VCC and VSS.

Finally, the op-amp's transconductance must be considered. The transconductance (g_m) is the output current in response to an input voltage. In signal processing applications, it is important to have constant g_m so that the output response exactly mirrors shifts in the AC level. However, for the purposes of the current project, it was determined that constant g_m is not a driving factor for designing a 1v op-amp.

3.1.2 Past Work / Background

3.1.2.1 Lowering the supply voltage

Reducing chip operating voltage has been a widely discussed topic among circuit designers. As the channel length and gate-oxide thickness becomes smaller, the supply voltage must be lowered in order to maintain device reliability. Power dissipation will also increase as the chip density increases, the. Lowering the supply voltage will lower the amount of power dissipated per unit of area. Finally, lower voltage is important for battery-powered equipment.

There are many challenges and problems which must be addressed as the supply voltage is decreased. According to researcher Phillip Allen, the minimum supply voltage necessary for any circuit is equal to $Vdd \geq V_{Tn} + | V_{Tp} |$.[14] If the threshold voltage is maintained around 0.7-0.8 volts, it appears impossible to have a minimum supply voltage of 1v. Hogervorst and Huijsing offer an alternative equation for the minimum supply voltage. For low-voltage circuits, $V_{sup,min} = 2(V_{gs} + V_{dsat})$. This translates to a supply voltage equal to two stacked gate-source voltages and two saturation voltages of a MOS device. For extremely low-voltage circuits, $V_{sup,min} = V_{gs} + V_{dsat}$. Extremely low-voltage circuits operate on one gate-source voltage and one saturation voltage.[15] This equation provides more flexibility than the previous since it does not directly depend upon threshold voltage.

The gate-source voltage (V_{gs}) of a transistor from the minimum supply equation above determines whether a transistor is operating in strong or weak inversion. A transistor is in strong inversion if $V_{gs} > V_T$. Saturation occurs when $V_{ds} > V_{gs} - V_T$. Usually, $V_{ds} = V_{dsat}$ for an op-amp, since all transistors in an op-amp are biased to saturation to obtain the largest voltage gain for a given I_{ds}. A transistor is in weak inversion (subthreshold region) when $V_{gs} < V_T$. Saturation occurs when $V_{ds} > 3$ to 4 V_{th},

[14] Allen, Phillip E., Benjamin J. Blalock, and Gabriel A. Rincon. "Low Voltage Analog Circuits Using Standard CMOS Technology," Proceedings 1995 International Symposium on Low Power Design , 1995, pp. 209-214.

V_{th}= thermal voltage=kT/q = 25 mv at room temperature.[16] Thus the transistor can operate even below the threshold voltage level of the device.

Several researchers have found solutions to the low voltage supply problem. The first solution to the low voltage supply problem involves creating new devices and altering the process. An obvious solution would be to alter the process and technology to create devices with lower threshold voltages. The drawbacks to this method are that it could be very costly, time consuming, and often unreliable. Furthermore, these low V_t devices often only operate below 1v and become non-functional at higher supply voltages. Creating low threshold voltage devices would require much work on improving processes to enable double or triple well technology. A less drastic approach still involves a change in the transistor, but tries to avoid creating a completely new device.

3.1.2.2 Altering threshold voltages / "new" devices

The first solution is a double gate driven MOSFET (DGMOS) by Louis Wong and Graham Rigby. [17] In this configuration, the body of the MOSFET is dynamically connected to the gate by a capacitor. A reverse-biased MOS diode is also inserted between the body and the voltage supply in order to minimize body leakage current. Thus the operating region the device is in determines the potential of the body. When

[15] Hogervorst, Ron and Johan H. Huijsing. Design of Low-Voltage, Low-Power Operational Amplifier Cells. Boston: Kluwer Academic Publishers, 1996, pp.6-7.

[16] Hogervorst, Ron and Johan H. Huijsing. Design of Low-Voltage, Low-Power Operational Amplifier Cells, pp.7-12.

[17] Wong, Louis S.Y. and Graham A. Rigby. "A 1V CMOS Digital Circuits with Double-Gate-Driven MOSFET," ISSCC Digest of Technical Papers (CD-ROM), 1997.

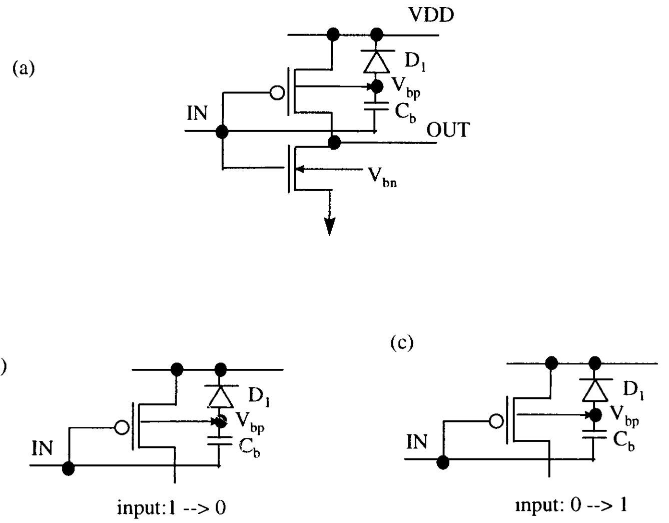


Figure 40: (a) DGMOS inverter (b) operation when input is 1-->0 (c) operation when input is 0--> 1

the transistor is on, it will have a low V_{th}, and when the transistor is off, it will have a high V_{th}. The DGMOS can offer high switching speed and low static power dissipation.

The CMOS inverter in Figure 40(a) has a PMOS transistor with a double gate. If the input changes from a "1" to a "0," (Figure 40(b)) the PMOS turns on, the source-body junction becomes forward biased, and the threshold voltage drops. A larger drain current is produced for faster switching speed, and C_b is charged through capacitive coupling. When C_b is fully charged, the body leakage current is lessened. If the input changes from a "1" to a "0," (Figure 40(c)) the PMOS turns off. C_b discharges through D_l, and the source-body junction becomes reverse biased. The body effect increases the threshold voltage, while the leakage drain current is reduced.

Problems with the DGMOS circuitry are that it requires isolated wells, which could be a significant area cost. Furthermore, the results will probably vary significantly over process and would therefore be unreliable.

Another device which dynamically alters the threshold of a device is the Body biased Controlled SOI (BCSOI) pass-gate.[18] Unlike other low-voltage SOI devices, researcher Tsuneaki Fuse has created the BCSOI pass-gate and the boosted ground scheme which can operate both below and above 1v supply. SOI technology is used because it provides reduced substrate capacitance, reduces leakage and latchup, minimizes body effect, and allows for speed improvements.[19] The BCSOI alters the threshold voltage, also improving speed. The basic concept of the BCSOI pass-gate connects the body of the SOI to the gate. When the pass-gate is on, the threshold voltage is low for a high current drive. When the pass-gate is off, the threshold voltage is high for a stable cut-off. These BCSOI pass-gates can be used to create circuits which can operate at much lower voltages than conventional pass-gate (CPL) technology. The boosted ground scheme, which is beyond the scope of this thesis, can be combined with the BCSOI technology to allow a wider range of operation both above and below 0.8v VCC.

[18] Fuse, Tsuneaki, Oowaki, Yukihito, et al. "A 0.5V 200MHz 1-Stage 32b ALU using a Body Bias Controlled SOI Pass-Gate Logic," ISSCC Digest of Technical Papers (CD-ROM), 1997.
[19] Weste, Neil H.E. and Kamran Eshraghian. Principles of CMOS VLSI Design, pp.125-130.

Although direct control of V_{th} will increase design time, area, and energy, the technique becomes necessary as the supply voltage is lowered furthur. Ricardo Gonzalez explains the necessity of controlling the threshold voltage in "Supply and threshold Voltage Scaling for Low Power CMOS." Controlling the threshold voltage has an advantage over a new device with lower V_t when process and operating point variations are taken into account.[20] Examination of the delay-product associated with circuits also supports the control of the threshold voltage.

3.1.2.3 Techniques

Since altering the process is expensive and long-term, this thesis investigated alternative ways to lower supply voltage without having to create a new Intel process. The first technique involves using low V_t devices in the "peripheral" interfacing circuitry and high V_t devices in the "core" designs.[21] In the past, this technique has been used successfully in level shifters which can operate as low as 1.5v. Both low and high (0.8v) V_t devices are becoming the standard mix of current process technology. Therefore, it is possible to use these different V_t devices together to improve performance. A complex level shifter has been designed using a mix of low and high V_t devices. This level shifter also includes bootstrapping devices which "boost" certain critical voltages. The technique of bootstrapping also becomes important in reducing the supply voltage. It can also improve power performance, as shown in the boosted ground scheme

[20] Gonzalez, Ricardo, Benjamin Gordon, and Mark Horowitz. "Supply and Threshold Voltage Scaling for Low Power CMOS," IEEE Journal of Solid-State Circuits, Vol.32, No.8. August 1997.

aforementioned. A more simplified version of level shifting in the form of an inverter is used in the op-amp design in this thesis. The details of the application of the "peripheral vs. core" technique is explained in the Methods section.

The second technique used to help lower supply voltage relates primarily to op-amps. A cascode circuit is used to allow the input common mode range to extend rail-to-rail. A complex but compact version of a folded cascode along with a complementary input stage allows that input stage to be rail-to-rail. If only a single differential pair with a current mirror load is used, the lower supply-rail can only be reached within one gate-source voltage. Thus the CMR is reduced. A folded cascoded input stage, however, is connected to the lower supply rail by a saturation voltage which is less than the gate-source voltage. Therefore, the CMR can go rail-to-rail with the folded cascode device.[22] A version of cascoding is used to help maximize the CMR in this thesis.

3.1.3 Relation to PVR

The PVR concept is basically a differential amplifier which senses the difference in current through two flash cells. Therefore, research into 1v op-amp can help to understand the path to developing a 1v PVR. Another crucial part of the PVR is the pump, which has been proven by others to be theoretically able to operate as low as 1v. Yet another lingering problem with low voltage PVR is that its threshold voltage cannot

[21] Otsuka, Nobuaki and Mark A. Horowitz. "Circuit Techniques for 1.5V Power Supply Flash Memory," IEEE Journal of Solid-State Circuits, Vol.32, No.8. August 1997.

be scaled in the same way as a traditional transistor. Smaller values can be chosen to trim the Flash cells, but the ramifications of those smaller values must be studied. The operation of a flash cell at 1v must also be verified. It will also be difficult to make a truly efficient PVR operational at 1v on the current process, especially without growing the die size.

3.2 Methods

An initial simple differential amplifier is tested to determine its feasibility at 1v. The schematic of the differential amplifier used is pictured in Figure 41.

[22] Hogervorst, Ron and Johan H. Huijsing. Design of Low-Voltage, Low-Power Operational Amplifier Cells, p. 22.

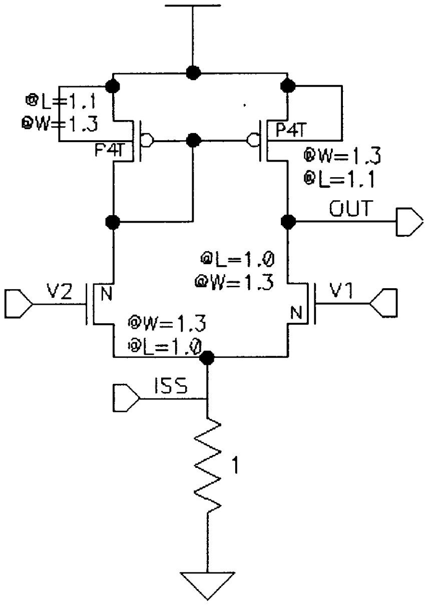


Figure 41: Simple differential amplifier

From simulation results, it is clear that the simple n-pair differential amplifier is not operational in the 1v supply range. This may be due to the high threshold voltages (0.8v) of the transistors and the inability to send enough current through the devices to put them in saturation.

Another drawback with the simple n-pair differential amplifier is that it will only operate well for the high end rail of input. That is, the NMOS pair works well for the

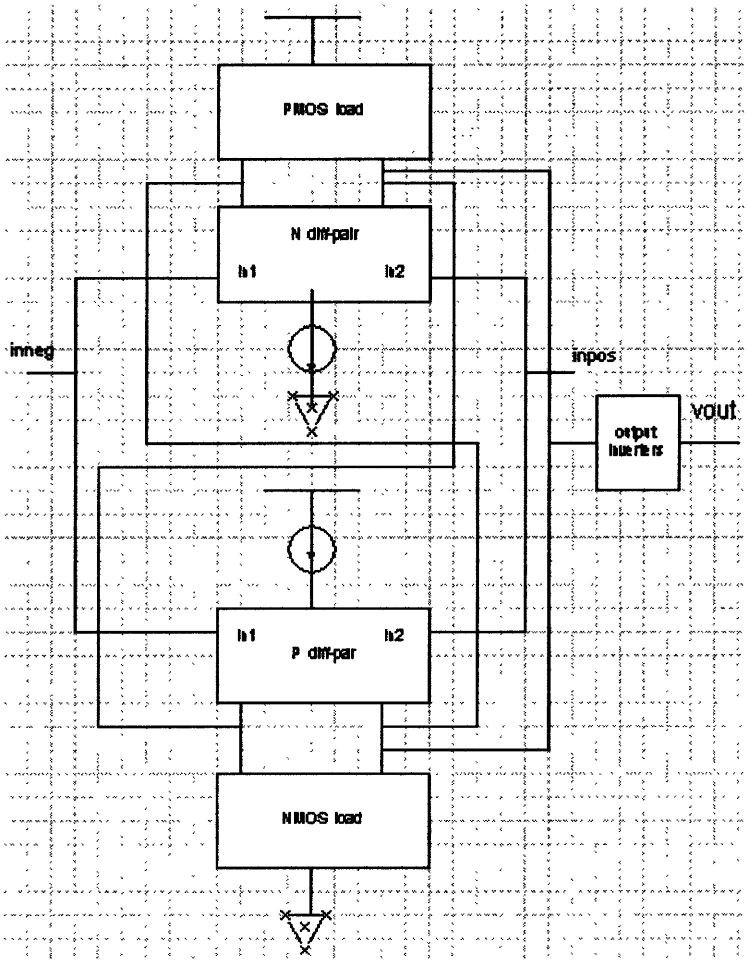


Figure 42: Initial block diagram of differential amplifier

input common-mode range of $V_{in,cm} > V_{ss} + V_{gs1,n} + V_{dsat3,n}$. A p-pair differential amplifier works well for the input range of $V_{in,cm} < V_{dd} - V_{gs1,p} - V_{dsat3,p}$.[23] Therefore, putting the n and p-pair differential amplifiers together in a complementary arrangement would enable a better response for both the high and low end input CMR. A past Intel group has designed a complementary op-amp which can work as low as 1.5v. This op-amp served as a basis for determining the feasibility of a 1v op-amp (Figure 42).

The first step in designing the op-amp at 1v is to determine if the existing 1.5v op-amp can be resized to be operational at 1v. Initially, the op-amp was resized to operate at 1v with a low bias current of 15uA. The results of the simulations are shown in Figure 43-Figure 45.

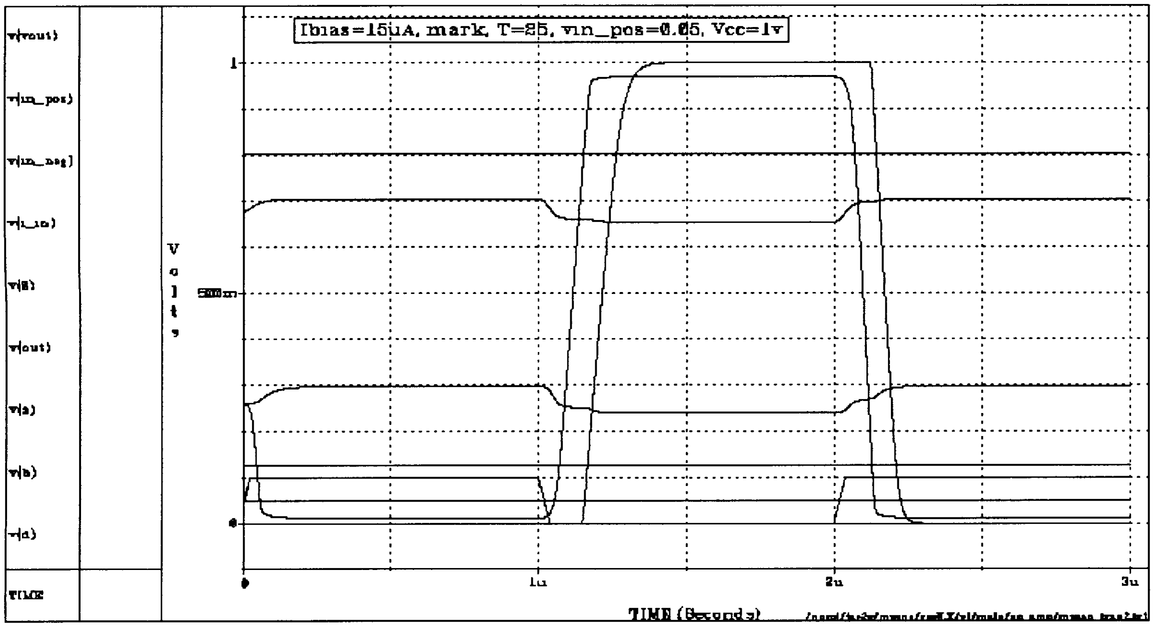


Figure 43: Initial op-amp, resized with Ibias=15uA, vin_pos=0.05

[23] Ferri, Giuseppe and Willy Sansen. "A Rail-to-Rail Constant-gm Low-Voltage CMOS Operational Transconductance Amplifier," IEEE Journal of Solid-State Circuits, Vol. 32, No.10, October 1997.

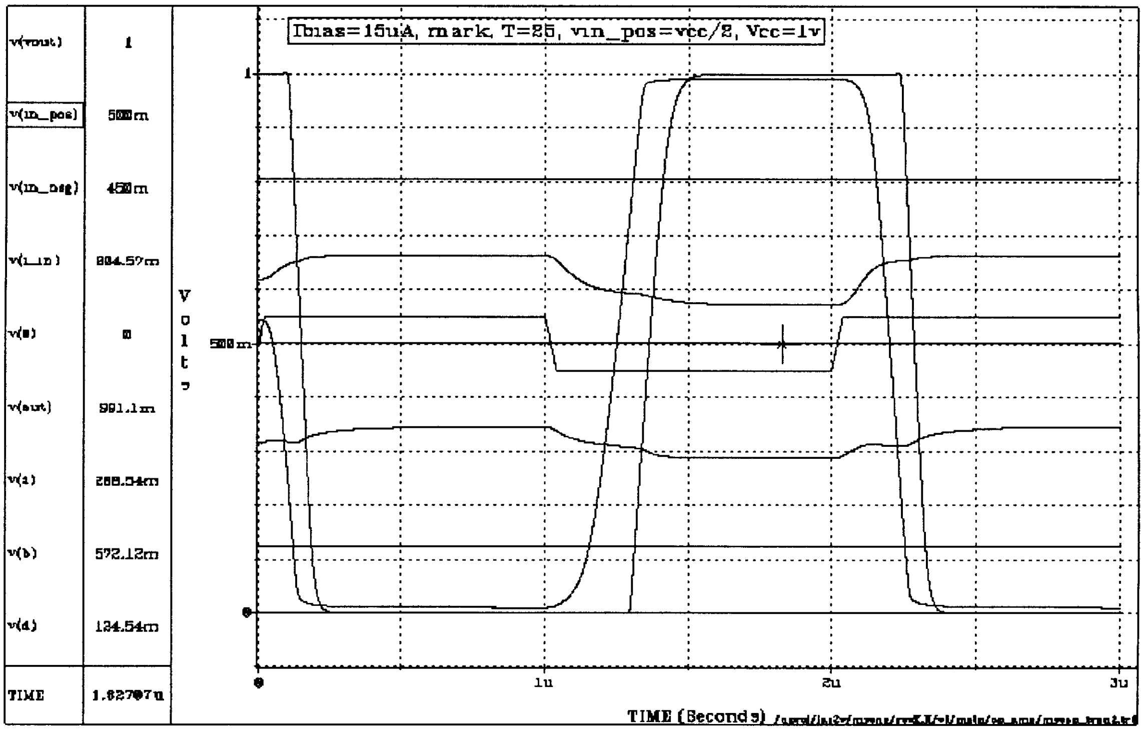


Figure 44:Initial op-amp, resized with Ibias=15uA, vin_pos=vcc/2

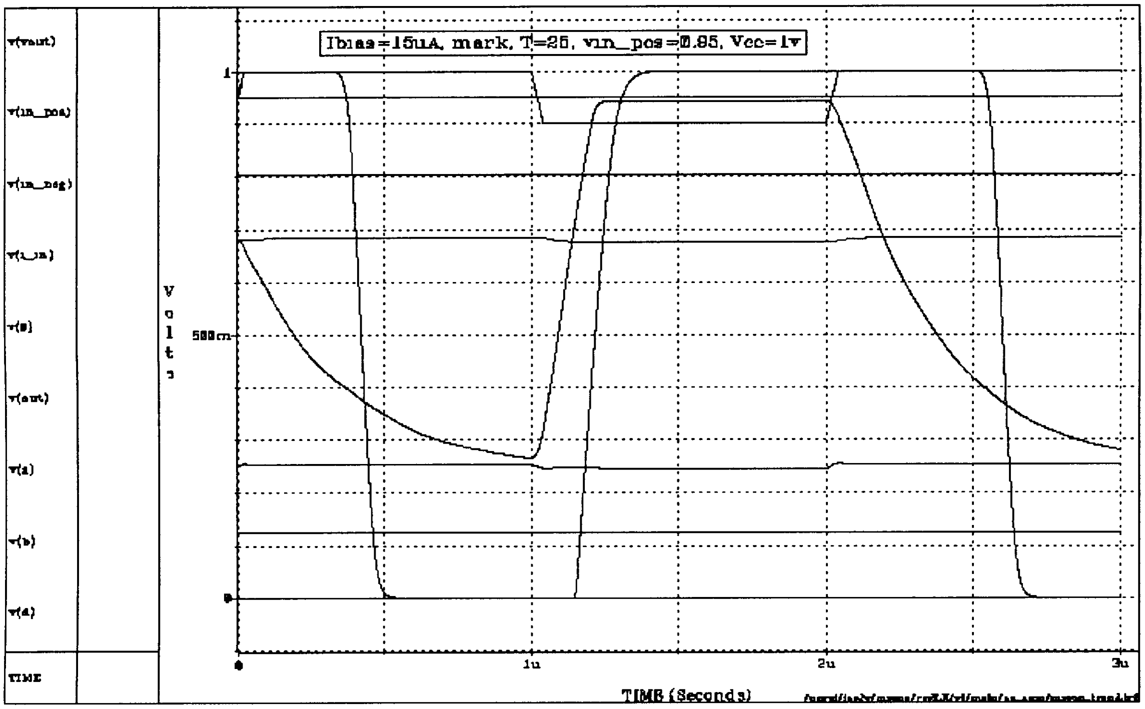


Figure 45: Initial op-amp, resized with Ibias=15uA, vin_pos=0.95

However, this low bias current of 15uA implies very large propagation delays, as can be seen in Figure 45. Generation of a small bias current would also require very large transistors or resistors. For example, R= V/I = 1v/15uA = 6.7X10^4 ohms = 67 Kohms. Therefore, the circuit was again resized for a bias current of 30uA. The ideal bias current would be 50-100uA. However, the circuit is non-operational at these high values. Therefore, a lower current was used with the penalty of a larger delay. A larger bias current requires a larger V_{dsat}, as can be seen in a typical I_{ds} vs. V_{ds} curve of a transistor driver by a voltage source with a load attached (Figure 46).

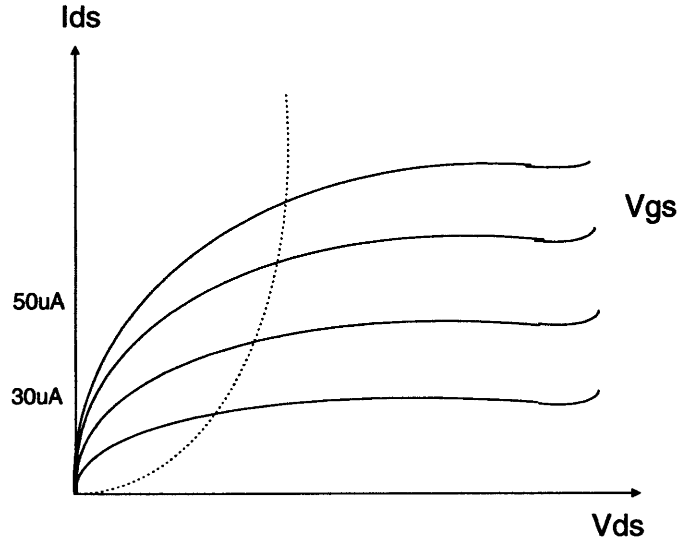


Figure 46: Ids vs. Vds curve - lower bias implies lower Vdsat

The optimal sizes for operation at 1v were determined by initial guesses and simulation. Then the optimize tool in HSPICE was used to get the optimal sizing which would give the least delay. The simulation results for the mark skew are shown in Figure 47-Figure 49.

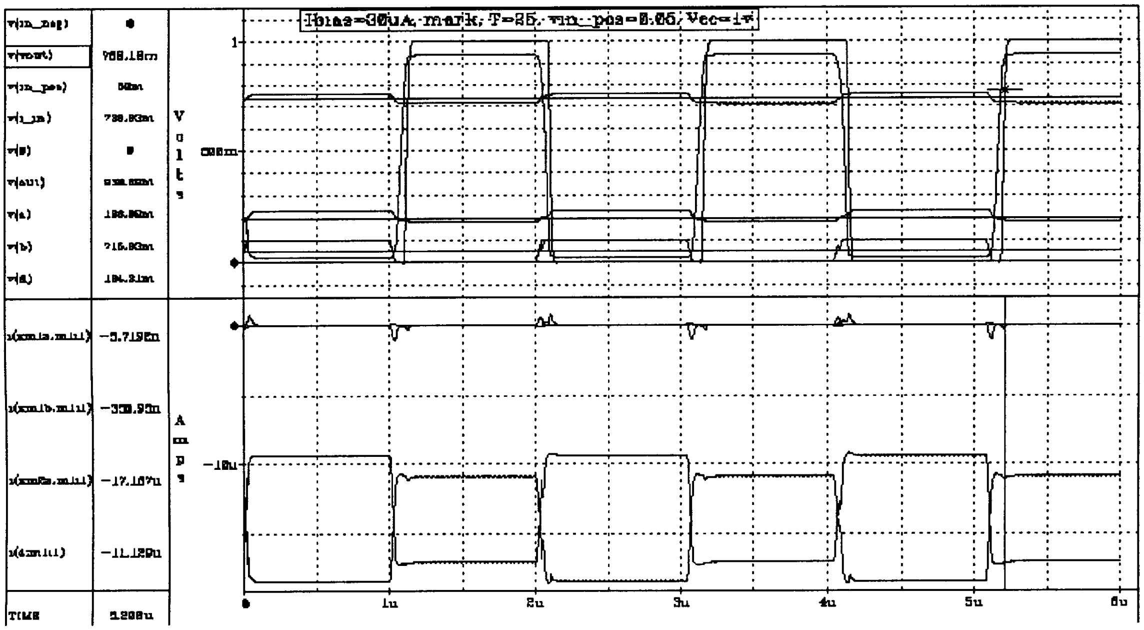


Figure 47: Initial op-amp, resized with Ibias=30uA, vin_pos=0.05, mark

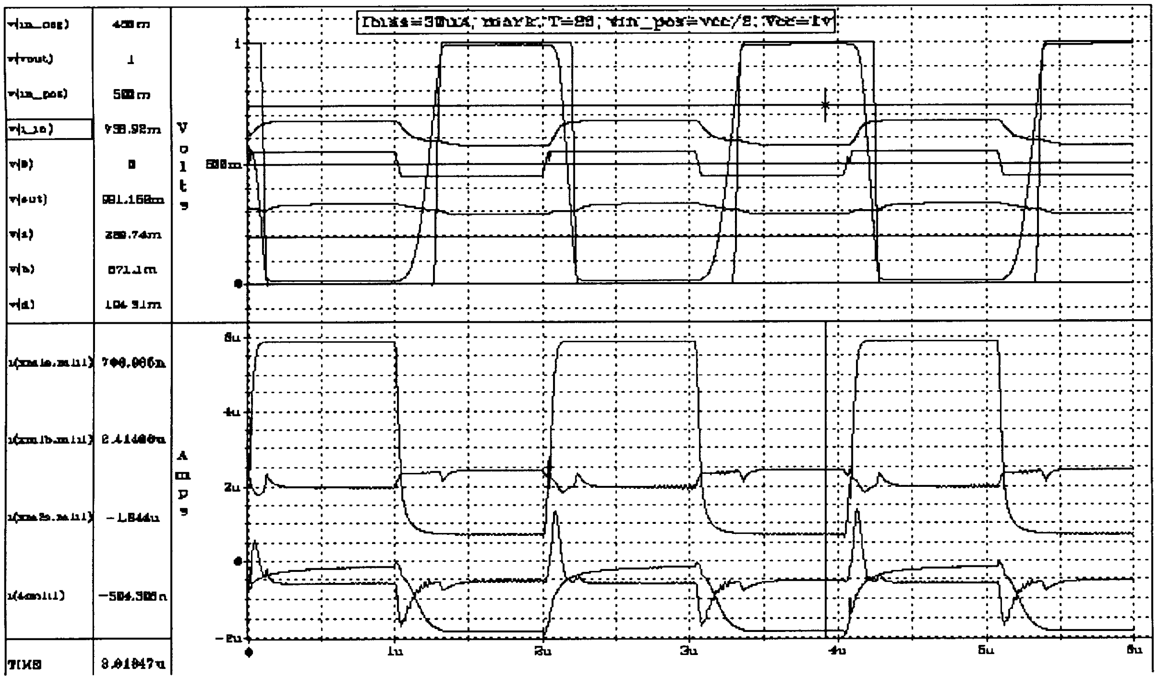


Figure 48: Initial op-amp, resized with Ibias=30uA, vin_pos=vcc/2, mark

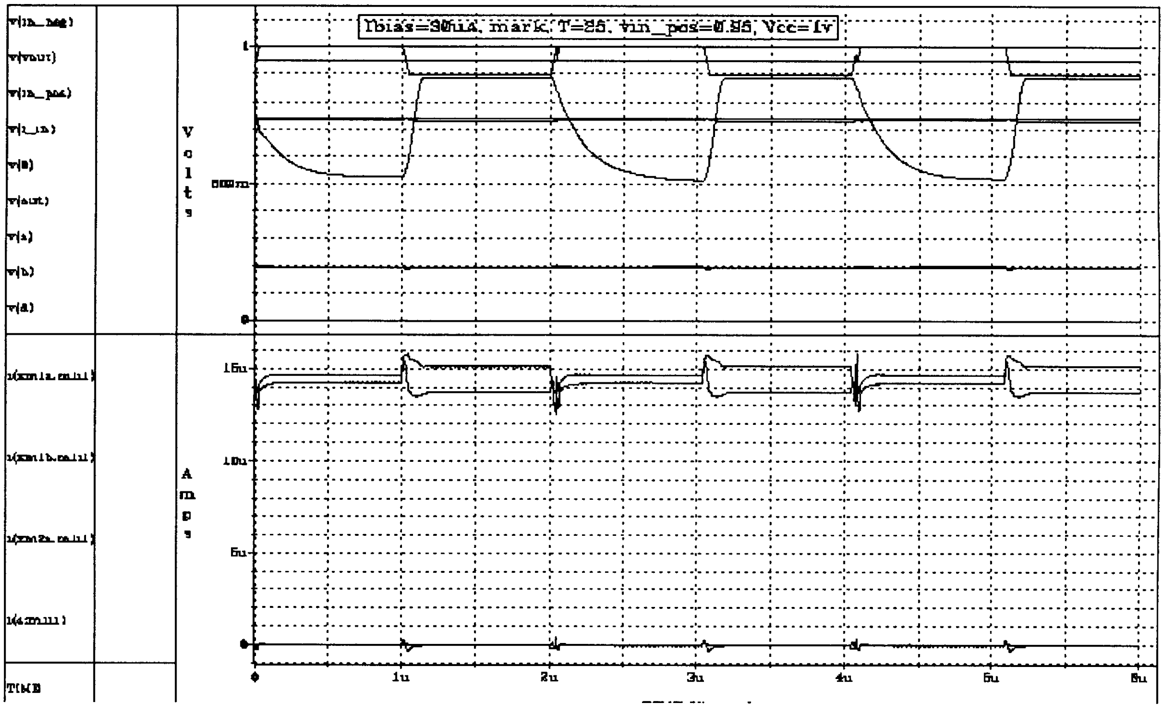


Figure 49: Initial op-amp, resized with Ibias=30uA, vin_pos=0.95, mark

All other skews are shown in Appendix I. As can be seen from Figure 49, the op-amp has trouble operating at the high input rail (vin_pos=0.95). Although the initial output OUT fluctuates, the output VOUT does not swing or change. Therefore, the inverters which are placed between OUT and VOUT must be altered to allow the output of the op-amp to swing rail-to-rail. This can be accomplished by using prime (low V_t) devices in the output inverters.

Prime devices for the output inverters are also used to try to reduce the propagation delay of the output of the op-amp. These prime devices have lower threshold voltages (0.3-0.4v). The output inverters are essentially simple level shifters. In other words, a signal fluctuation on the input of the inverters becomes magnified to VCC or VSS so that the output appears rail-to-rail. Therefore, lower V_t devices can be used in these "level shifter" inverters to provide for faster transitions. This concept is similar to the "core vs. peripheral" technique mentioned in the previous section. The simulations and speed improvements are shown in Figure 50-Figure 52. However, the prime devices will have to be guard ringed to prevent leakage, since these devices have a higher leakage current than traditional transistors.

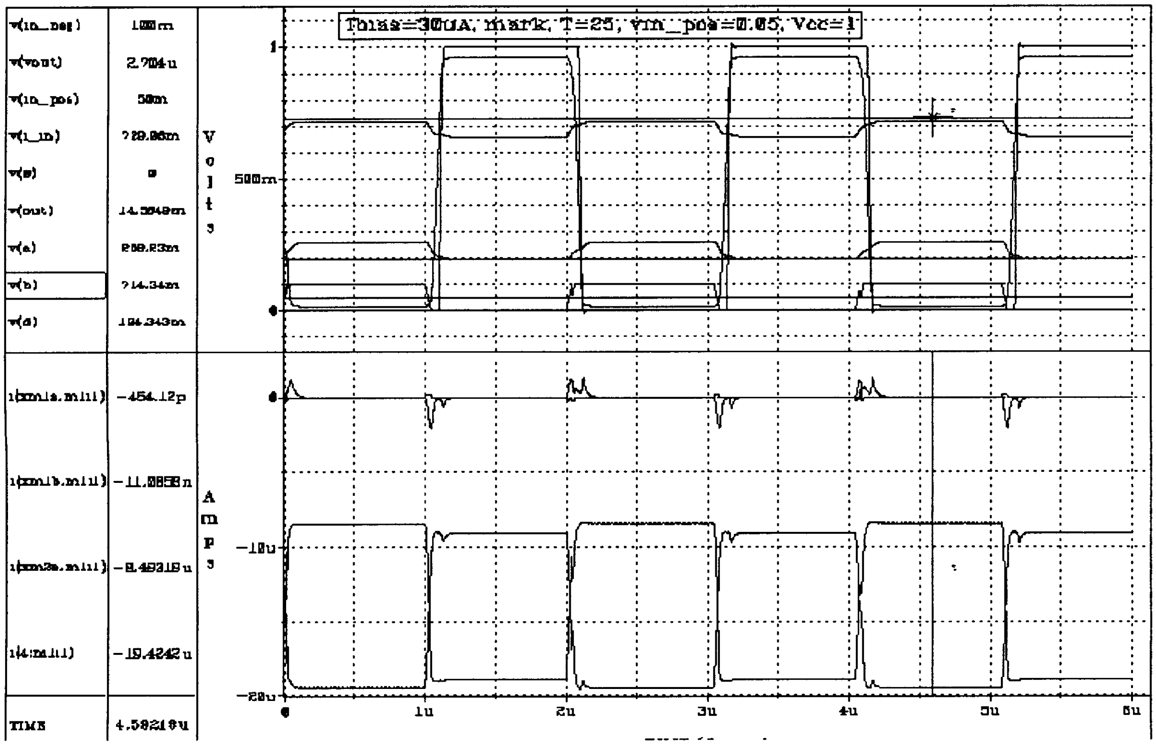


Figure 50: Op-amp with prime devices Ibias=30uA, vin_pos=0.05, mark

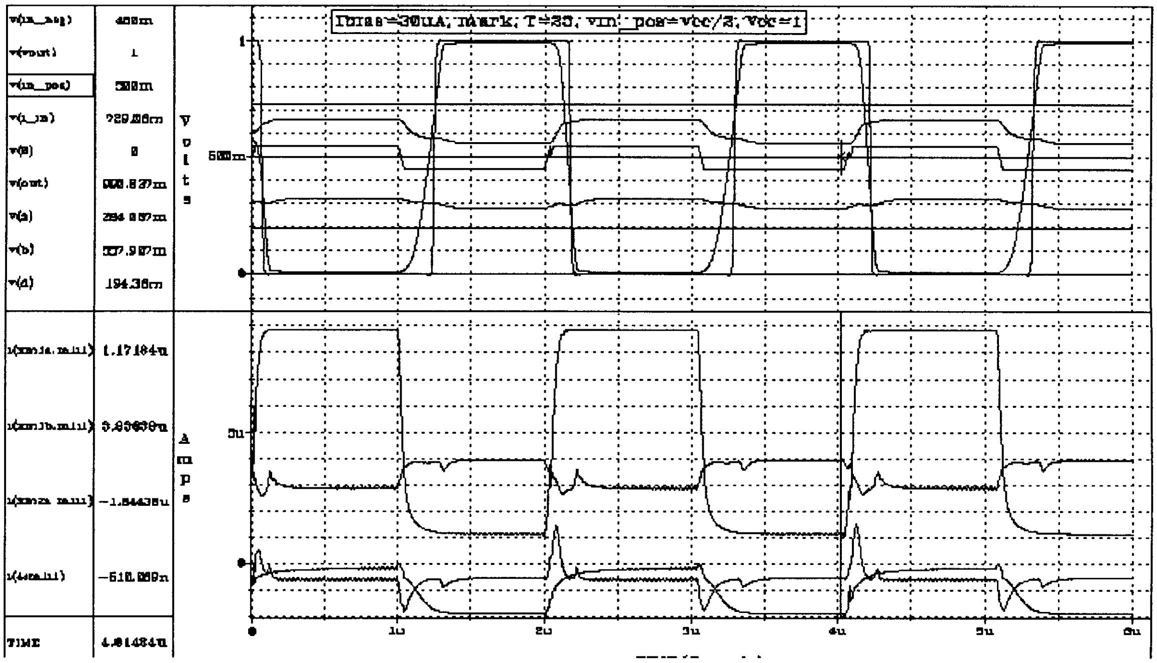


Figure 51: Op-amp with prime devices Ibias=30uA, vin_pos=vcc/2, mark

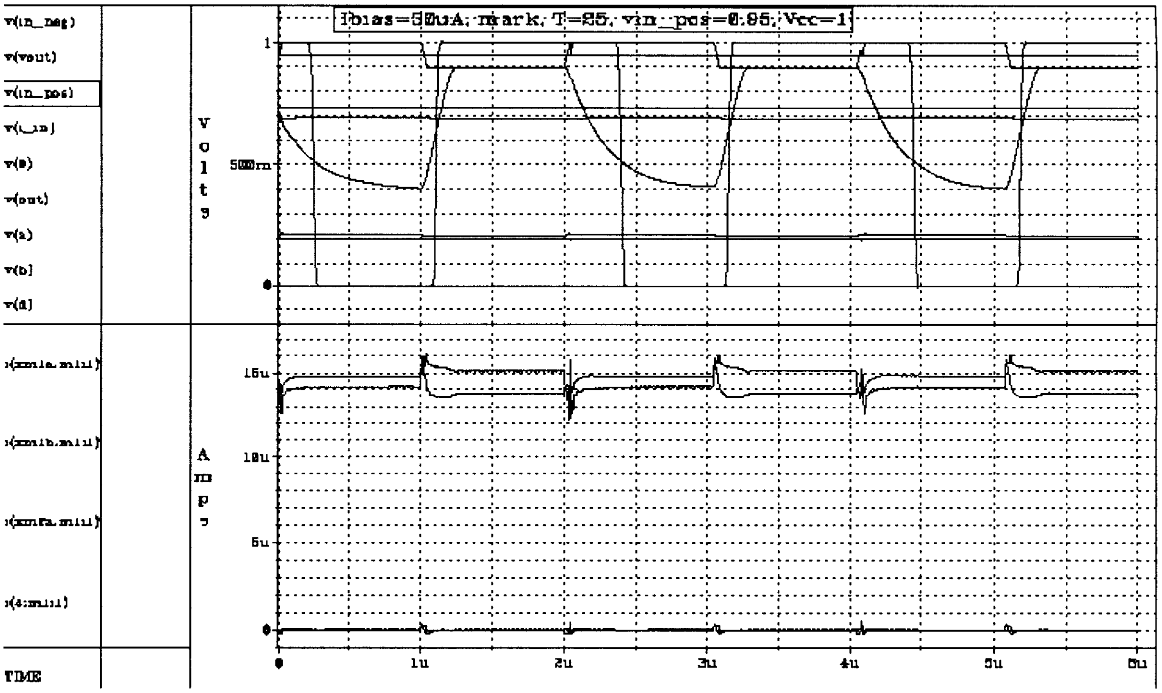


Figure 52: Op-amp with prime devices Ibias=30uA, vin_pos=0.95, mark

As can be seen from Figure 52, the prime devices allow VOUT to swing from rail-to-rail.

3.3 Results

The final op-amp also includes a cascode stage load. This cascode guarantees that the input CMR can go rail-to-rail. A more compact approach would have been to use a folded cascode as a load to the complementary pair as pictured in Figure 53.

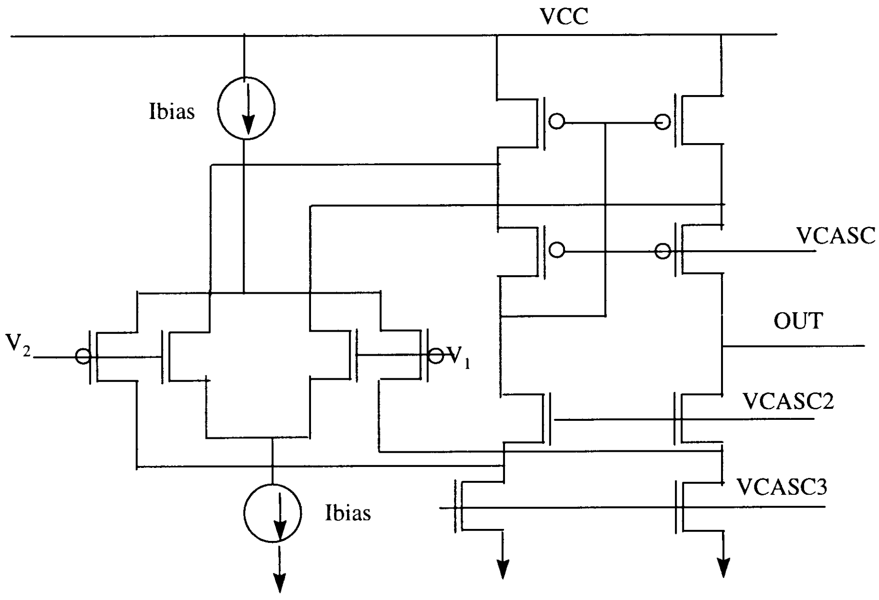


Figure 53: Folded cascoded input[24]

[24] Hogervorst, Ron and Johan H. Huijsing, Design of Low-Voltage, Low-Power Operational Amplifier Cells, p. 29.

However, this folded cascode was not operational with the Intel design because traditional current mirrors were not used in the Intel design as loads. The folded cascode is designed built on the assumption that traditional current mirrors are used as loads. The Intel design is unique and uses simple transistors as loads and then connects its outputs to other current mirrors and amplifies its signals. Therefore a modified cascode circuit was added rather than the completed folded cascode, which would have required putting in current mirror loads. The block diagram of the op-amp with cascode and prime devices is shown in Figure 54.

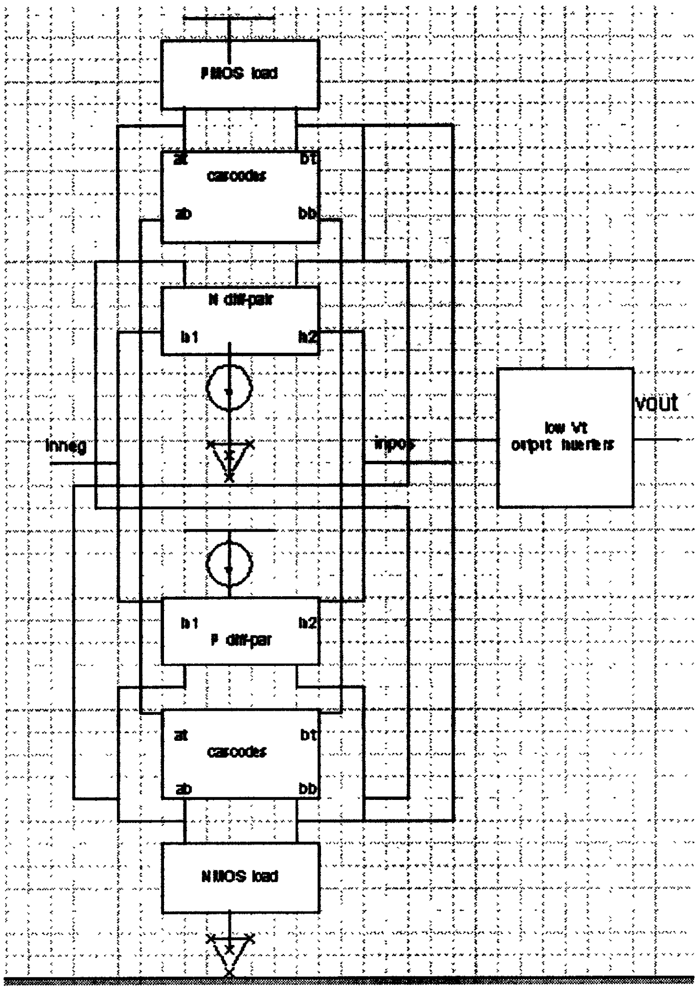


Figure 54: Block diagram of op-amp with cascode and prime devices

The final simulation with prime output inverters and cascode stage is shown in Figure 55-Figure 57 for mark skew.

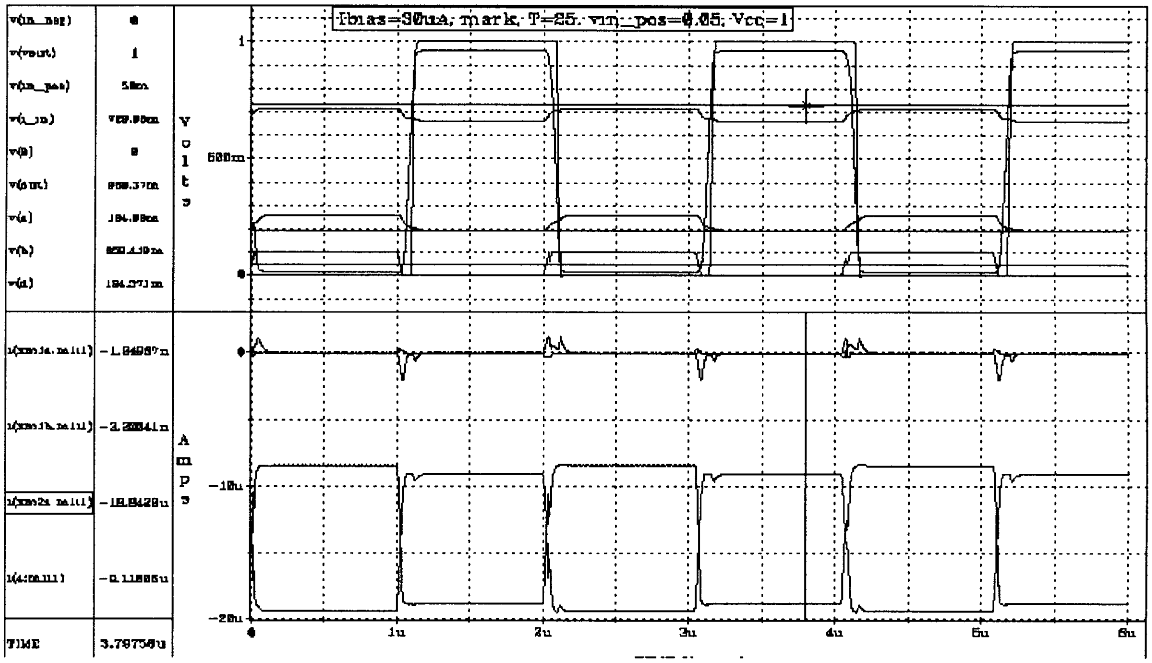


Figure 55: Op-amp with cascode and prime devices, vin_pos=0.05, mark

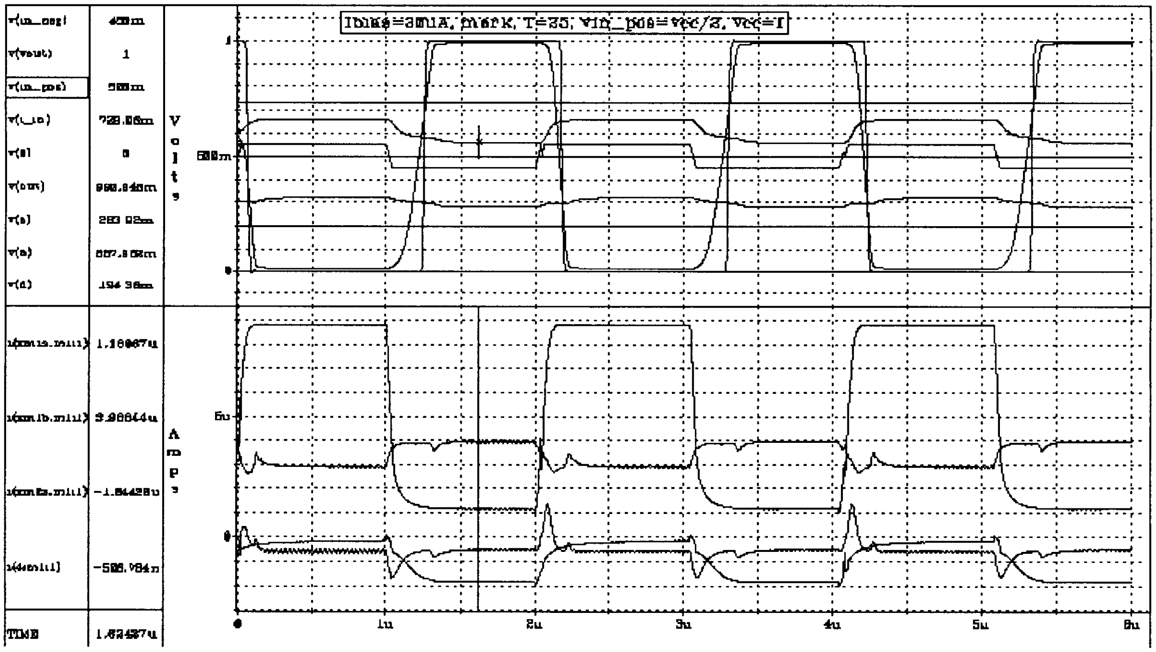


Figure 56: Op-amp with cascode and prime devices, vin_pos=vcc/2, mark

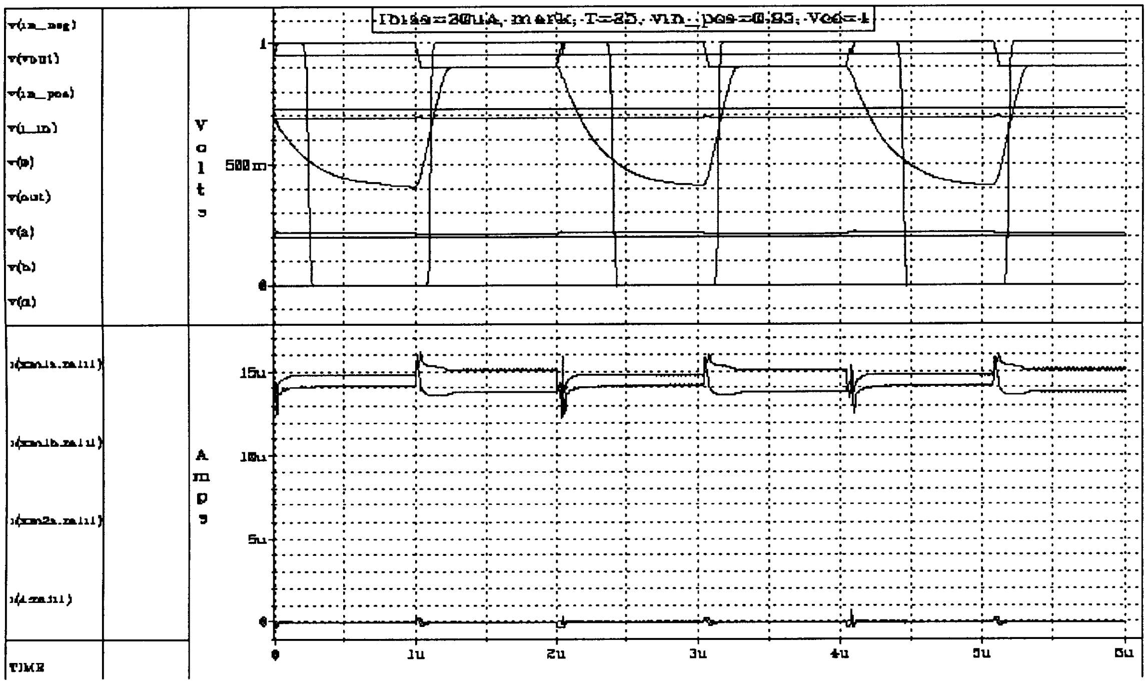


Figure 57: Op-amp with cascode and prime devices, vin_pos=0.95, mark

All other skews are in Appendix J. The positive input to the op-amp is set to a reference voltage VREF, while the negative input is varied in a waveform ±0.05v around VREF. VREF is tested for the values 0.05, VCC/2, and 0.95. This would give a CMR of 0.05-0.95v. Some problems occur on the high input range of 0.95v. The low V_t prime devices still allow a rail-to-rail output on VOUT. However, the true output OUT of the op-amp does not swing as far as desired (Figure 57).

The op-amp is also tested as a unity gain buffer. The output VOUT is connected to the negative input. The positive input vin_pos is then given a linear series of points from 0-1.1. VOUT should follow the action of the input vin_pos, which is shown in Figure 58.

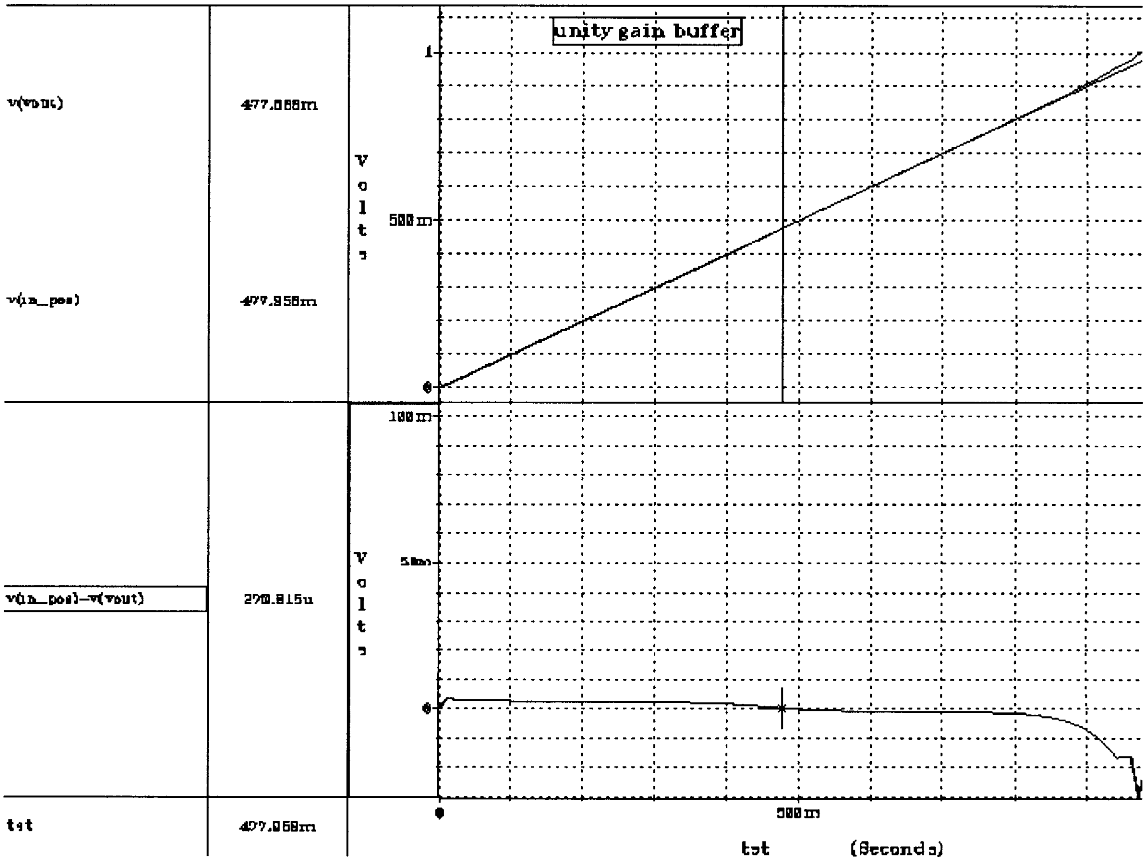


Figure 58: Op-amp as a unity gain buffer

The transfer curve of the op-amp is shown in Figure 59.

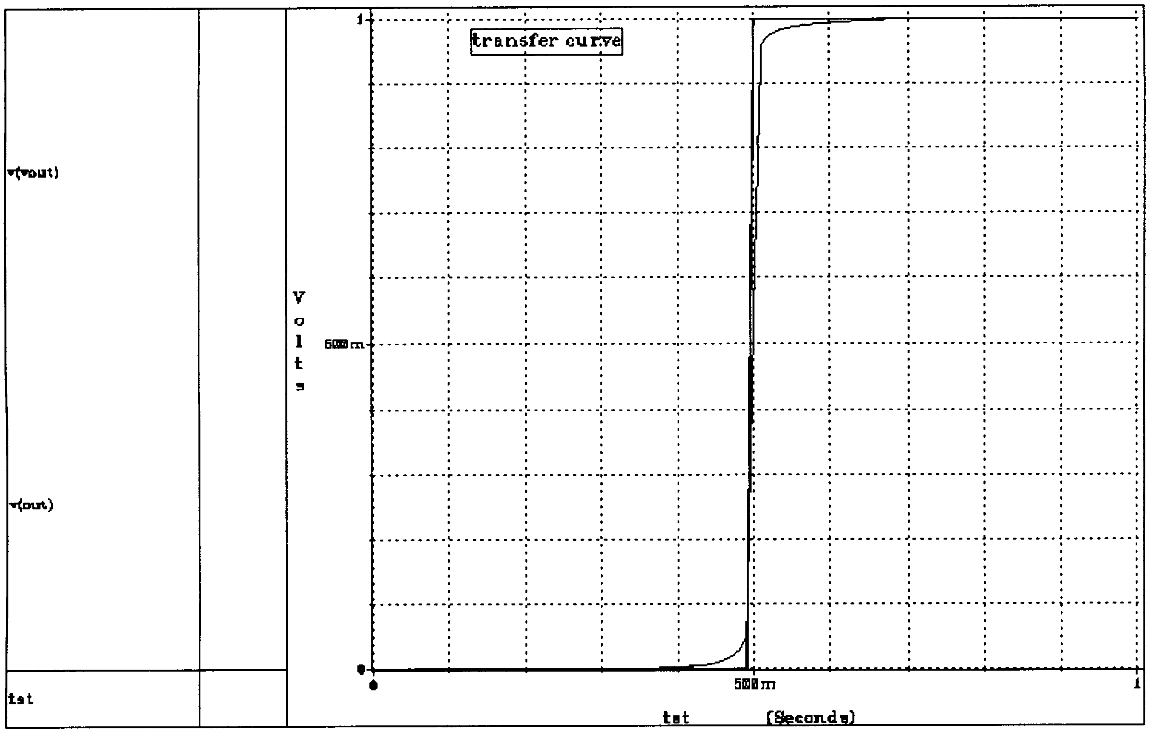


Figure 59: Transfer curve of op-amp

Although the op-amp proved operational at 1v, it is still not practical. The op-amp still suffers from much reduced performance in the form of very long delays. The delays for the op-amp are shown in Table 8. The op-amp still has an undesired delay for the high input range (VREF=0.95) for mark and slow n slow p skews.

Skew	Temp	VCC	VREF	delay of rise of VOUT (ns)	delay of fall of VOUT (ns)
mark	25	1v	0.05	75.188	97.721
slow n slow p	100	1v	0.05	119.75	124.37
fast n fast p	-40	1.5v	0.05	57.260	73.872
mark	25	1v	vcc/2	234.24	158.70
slow n slow p	100	1v	vcc/2	183.30	162.70
fast n fast p	-40	1.5v	vcc/2	43.764	48.474
mark	25	1v	0.95	102.78	356.41
slow n slow p	100	1v	0.95	131.33	387.42
fast n fast p	-40	1.5v	0.95	64.482	67.975

Table 8: Propagation delays from input to output of op-amp, with a 2pF load

Therefore, the initial solution can still be optimized for speed by sizing more wisely. However, one must keep in mind that creating a 1v op-amp without changing the process requires much added circuitry which will grow the die. Also, much efficiency and performance will probably be lost. The solution also could not be tested below 1v since the threshold voltages of typical NMOS and PMOS were too high. However, realistically, the circuit needs to be operable for 0.8v-1.5v VCC. Therefore, more research needs to be done to overcome these boundaries. The best approach for the 0.8v limit will probably be to invest time and money in researching the modified devices mentioned in the previous section.

4. Conclusions

4.1 3v Sample and Hold

Through HSPICE simulations, the implementation of 3v sample and hold was proven to be operational. This methodology of 3v sample and hold will improve read speeds up to 10%. Reducing the read speed of a flash memory chip will greatly enhance the performance and usefulness of the chip. If flash memory read speeds could be further reduced, it may become a viable alternative to other memories that may be higher cost. Furthermore, faster read speeds will aid future developments in personal handheld computers and other digital portable equipment.

Altering the 5v sample and hold theory to be operational at 3v proved challenging. The original goals of the 5v project were to use sample and hold to provide tighter sensing levels for a multi-level cell flash memory. Sample and hold strictly regulated the values during a read, therefore restricting the sensing levels. The design objectives for 3v sample and hold, however, were different. The methodology was used to try to increase read speed and also reliability. Furthermore, implementation of 3v sample and hold was entirely different from the 5v project since the two projects had completely different circuitry and design. The most challenging part of implementing 3v sample and hold was determining a clean and concise way to alter the Active Pump to always use REF2 as a reference voltage. The use of a 30pF sampling capacitor and new sampling waveforms also needed to be determined and verified.

The changes to the PVR and Active Pump to implement sample and hold were done using CAM options. Using the bit SELSMPHLD enabled the chip to use the old methods (SELSMPHLD = 0), or the new sample and hold (SELSMPHLD=1). Implementing this CAM option created some problems. The most important concern was a glitch in the circuitry that occurred at high temperatures. The PVR seemed to operate incorrectly; however, the source of the problem proved to be an NMOS switch that had to be added for the CAM that was creating an effective resistance between the output node REF2 and the sampling capacitor.

Some positive aspects of the 3v sample and hold are that it maintains warmup time of the PVR. It also meets the specifications for ramp-up time. Furthermore, the standby current in the Active pump is not affected. The 3v sample and hold will also provide a more reliable source of 2v and 4v references.

A negative aspect of the 3v sample and hold circuitry is that it will increase standby current in the PVR by 5.4uA. This is a small price to pay, however, for improved read speed performance.

Future improvements on the 3v sample and hold may involve more in depth and in lab testing of actual circuits. Also, the circuitry can be redesigned for greater efficiency by reducing the standby current. A possible way to reduce this current is to redesign the PVR to burn less current when it is active.

4.2 1v op-amp

The 1v op-amp discussed in this thesis was created as an experiment in new low voltage techniques. The compact design involves a few key features. A complementary input-stage, which has been shown to work before at voltages as low as 1.5v, was the starting point of the research. To make the comparator operational at 1v, the circuit was first resized with a lower Ibias = 30uA. Then, the input stage was altered into a complementary cascoded input stage. The cascoded input allows the input common mode range to swing rail-to-rail at low voltage. Finally, the output inverters, which were used as level shifters in the original design, were redesigned using low threshold voltage devices. These low V_t devices allow the output voltage of the op-amp to swing rail-to-rail on a small signal change of the op-amp.

The positive aspects of the 1v op-amp are that the comparator proved operational without a change in the process. The output swing is decent in all skews; however, the performance of the op-amp is very slow and could use improvement. Furthermore, it was not possible to go below 1v, since the V_t of a typical transistor is about 0.8v. However, in practical applications, the VCC supply would range from 0.8-1.5 for a 1v supply. Therefore, more research is needed in this area. Ultimately, operation at such a low voltage will probably require a process change which would enable both CMOS and bi-CMOS technologies to be used effectively and efficiently in combination. Alternate devices which require minimal process change but entail addition of capacitors and diodes could also be used to improve performance.

Future improvements on the existing design could include speed and reliability improvements. These could be approached initially by a re-optimization of the transistor sizes. Devices which do not require process change could also be tested in the existing design for speed improvements.

5. Appendix

5.1 Appendix A: Original simulation of wordline

(Note: all results have been multiplied by an undisclosed constant).

device	**run 0** **(size um)**	**1**	**2**	**3**	**4**
md	153	95.2	85	85	112
mb	170	102	136	136	170
mc	142.8	85	85	85	170
ma	16.32	12.9	12.9	12.9	16.32
mf	27.2	27.2	27.2	27.2	34
mg	108.8	75	75	75	108
mh	108.8	75	75	75	108
mi	7.14	7.14	7.14	17	7.14
WL select time (ns)	139.6	157.8	144.5	126.8	138

Table 9: Original simulation - doubling mi decreases WL select time even while decreasing sizes of other devices in decoder.

5.2 Appendix B: Simulation results for direct wordline regulation

HSRCDRV (V)	WL select time (ns)
1.5VCC	127.36
3.9	134.5
4.1	124.17
4.5	111.49
4.9	103.80
5.1	100.88
5.3	98.57
5.5	98.57

Table 10: WL select time VCC=2.7v, T=100, slow n slow p

HSRCDRV (V)	WL select time (ns)
1.5VCC	67.72
3.9	89.25
4.1	83.57
4.5	76.4
4.9	71.91
5.1	68.4
5.3	68
5.5	68

Table 11: WL select time VCC=3.65v, T=-40, fast n fast p

5.3 Appendix C: Updated capacitance values for PVR

outputs		interconnect(um)			
	node	M1	M2	M1	(total M1)
REF4 -->	negpmp	102	4430	850	952
	nepmp	1200	170	0	1200
	hladout	2140	0	0	2140
	powerde	2080	0	0	2080
REF2 -->	actpmp	1360	0	0	1360
	hsrcdrvreg	2930	0	0	2930
	hpld	2930			2930
	nepmp	1200	4600	850	2050
	negpmp	102			102
	powerde	2080	0	0	2080
	pmppos	1360	0	0	1360
	lcpmp	1300	0	0	1300
inputs		interconnect(um)			
	node	M1	M2	M1	(total M1)
IS5V <--		170	4430	850	1020
POWDFF <--		170	4430	850	1020
PGMEN <--		1360	0	0	1360
MF1DRAIN <--		850	4430	0	850
MF2DRAIN <--		850	4430	0	850
MFGATE <--		1320	0	0	1320
PVREN <--		3230	4600	850	4080
SAMPLE <--		3230	4600	850	4080
SELSMPHLD <--					0

Table 12: FTRC values for flash-pair with new layout

outputs		interconnect(um)			
	node	M1	M2	M1	(total M1)
HSRC5DRV -->	hm	340	0	0	340
	swth	4430	0	0	4430
	pvr	1360	0	0	1360
	pmppos	n/a			n/a
	sd	510	4430	850	1360
OPPOS -->	pmppos	n/a			n/a
OPNEG -->	pmppos	n/a			n/a
OPOUT -->	pmppos	n/a			n/a

Table 13: FTRC values for active pump with new layout

5.4 Appendix D: Zoomed left schematic of PVR, MOP

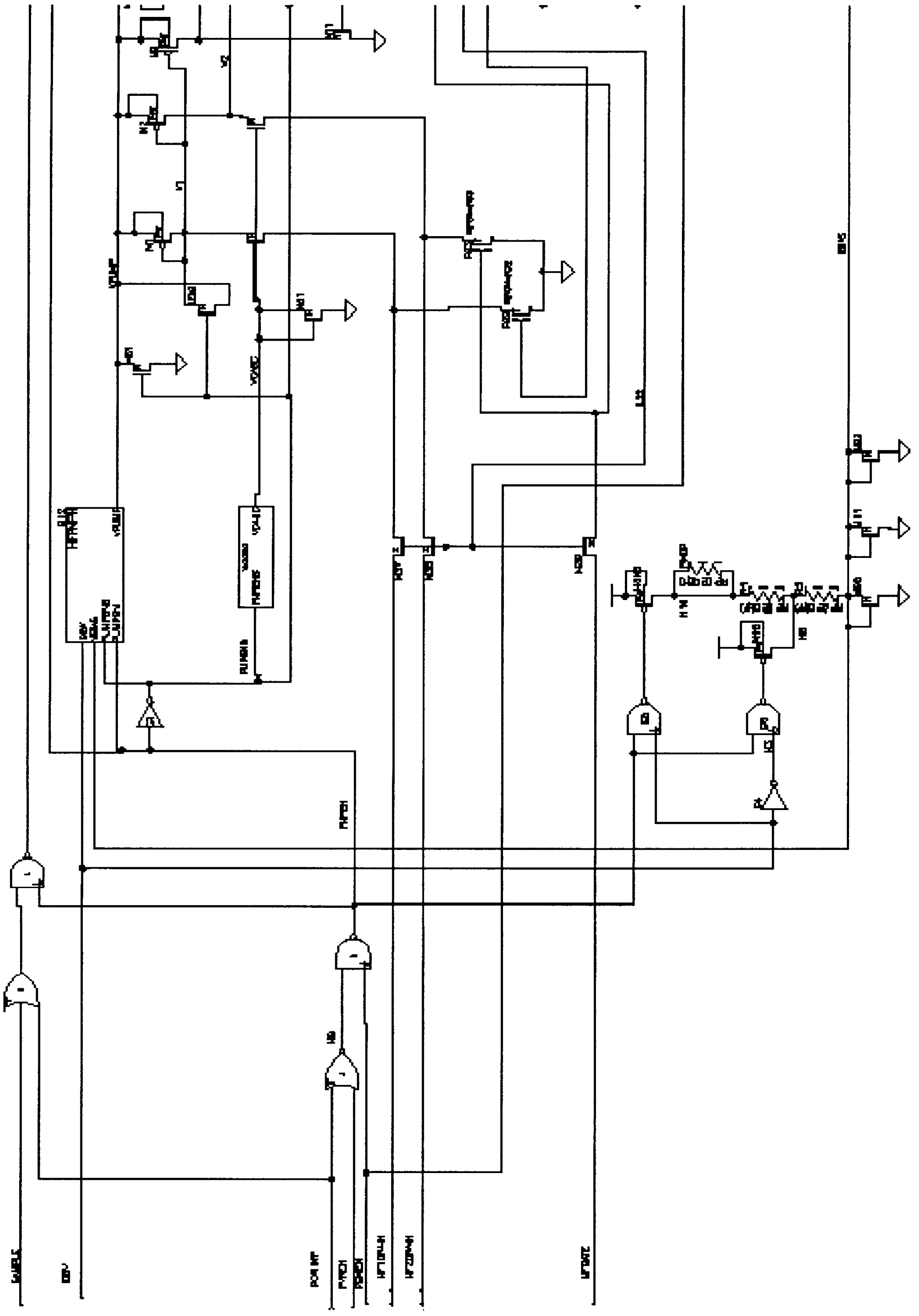


Figure 60: Zoomed left side of PVR in metal options

5.5 Appendix E: Zoomed right side of PVR, MOP

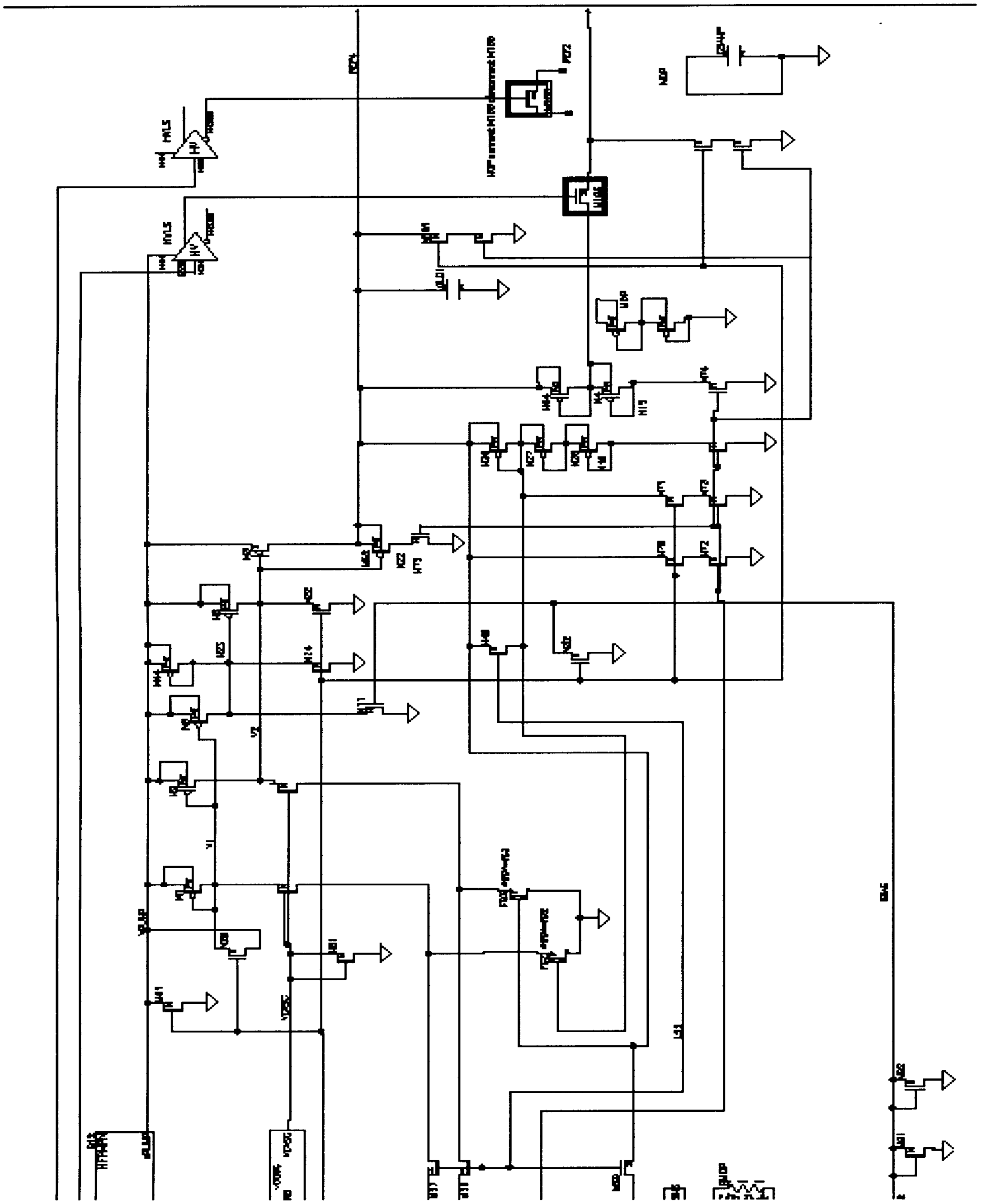


Figure 61: Zoomed right side of PVR in metal options

5.6 Appendix F: Zoomed left side of PVR, CAM

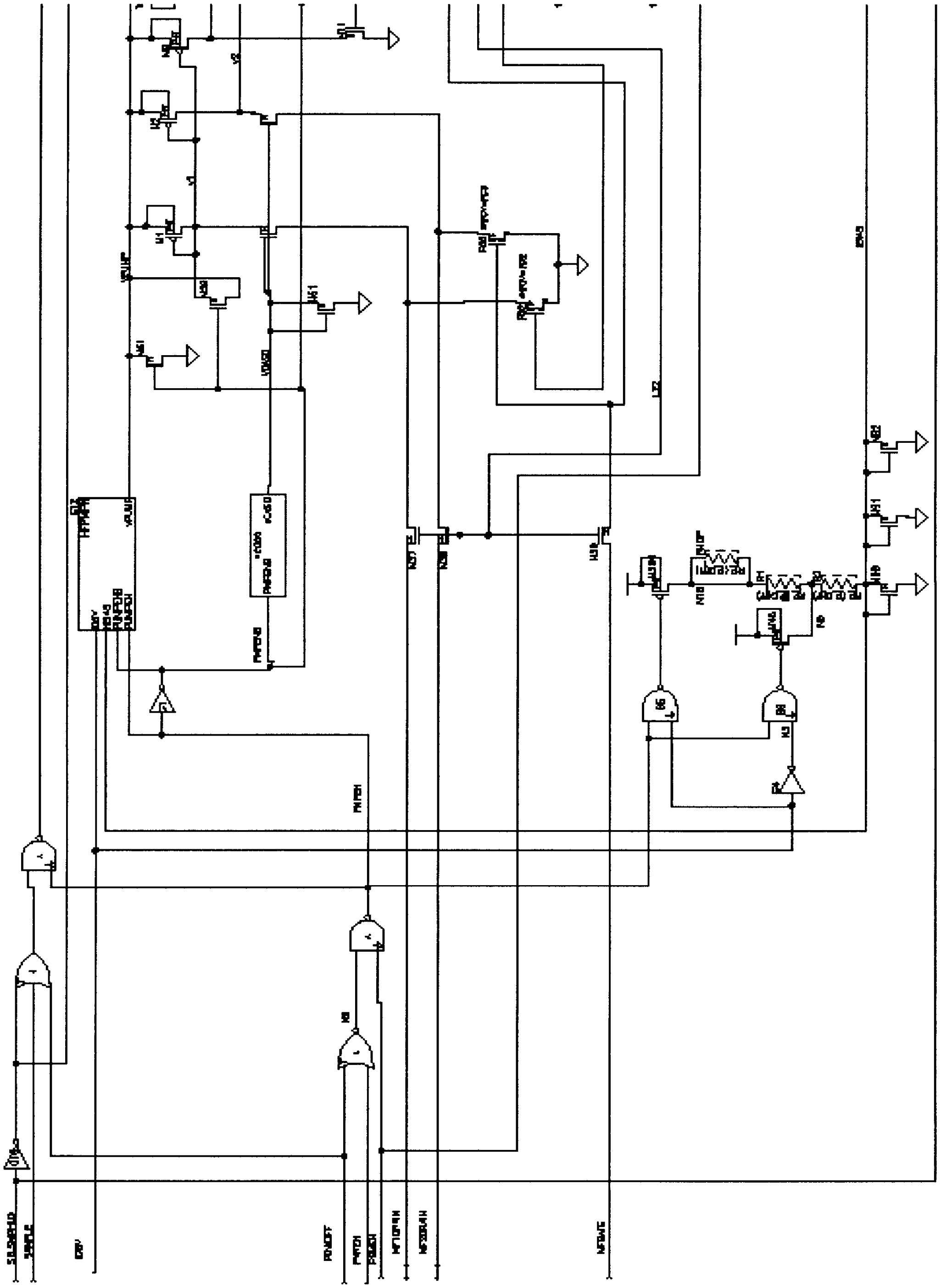


Figure 62: Zoomed left side of PVR with CAM option

5.7 Appendix G: Zoomed right side of PVR, CAM

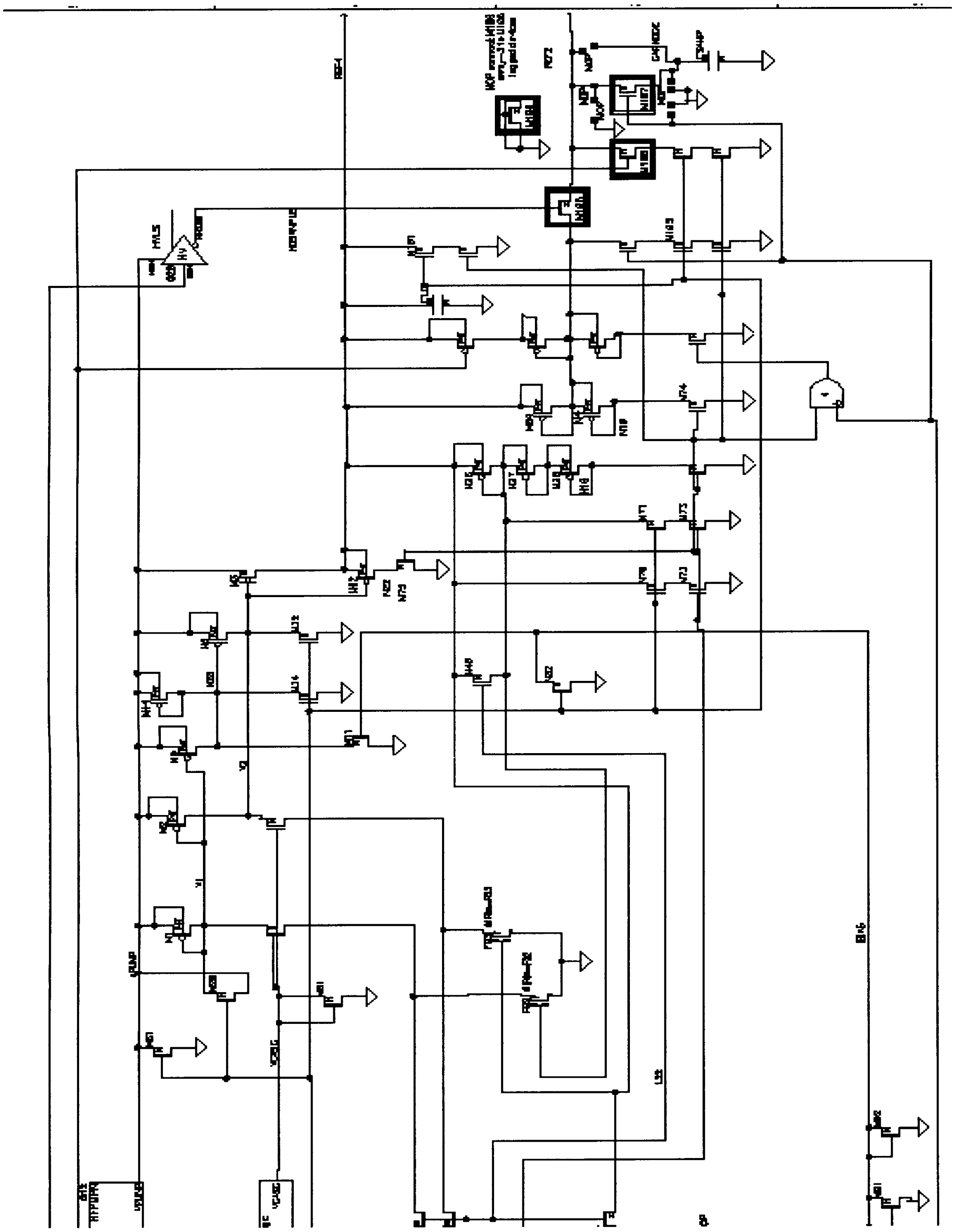


Figure 63: Zoomed right side of PVR with CAM option

5.8 Appendix H: Schematic of active pump regulation

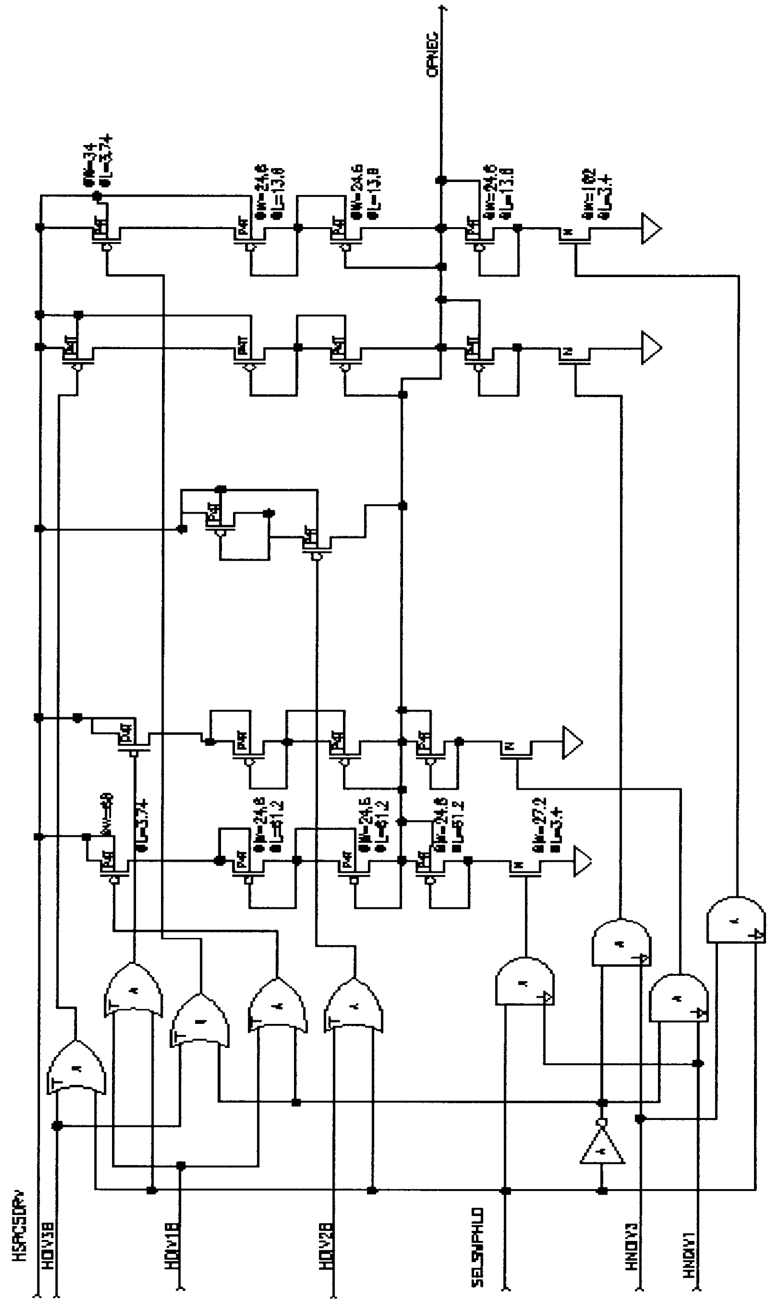


Figure 64: regulation circuit of active pump

5.9 Appendix I: Initial op-amp simulation results

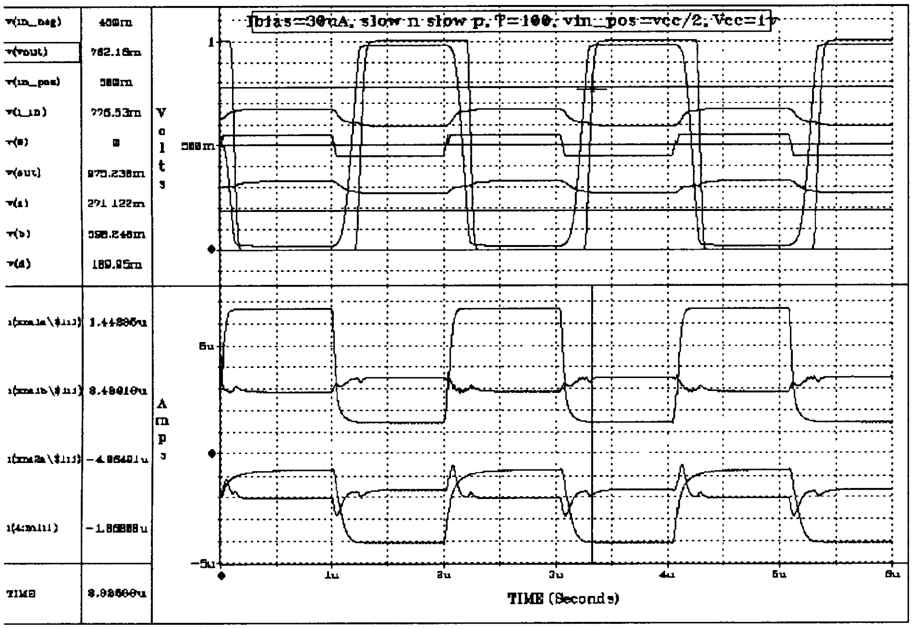


Figure 65: Initial op-amp, resized with Ibias=30uA, vin_pos=vcc/2, slow n slow p

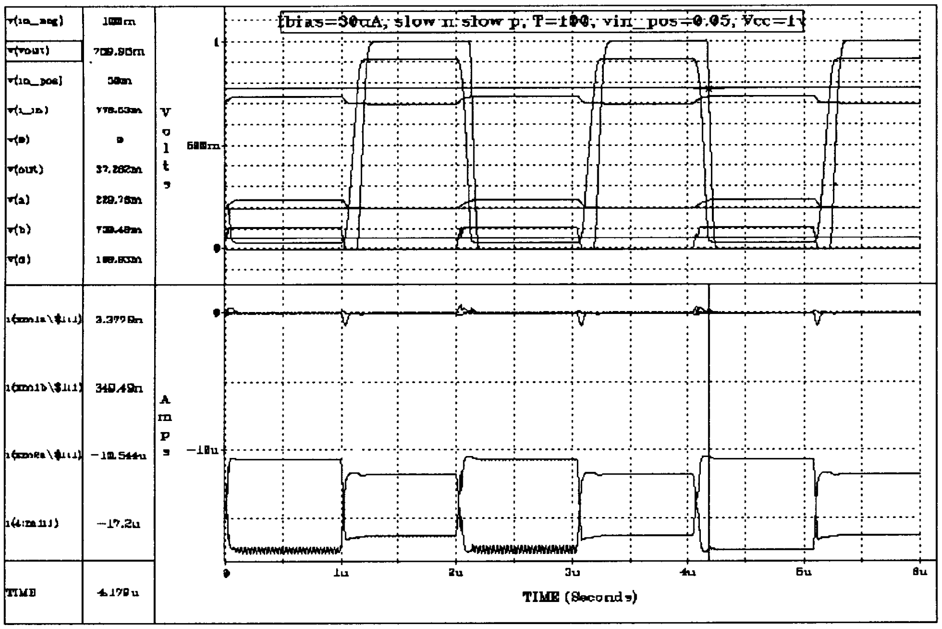


Figure 66: Initial op-amp, resized with Ibias=30uA, vin_pos=0.05, slow n slow p

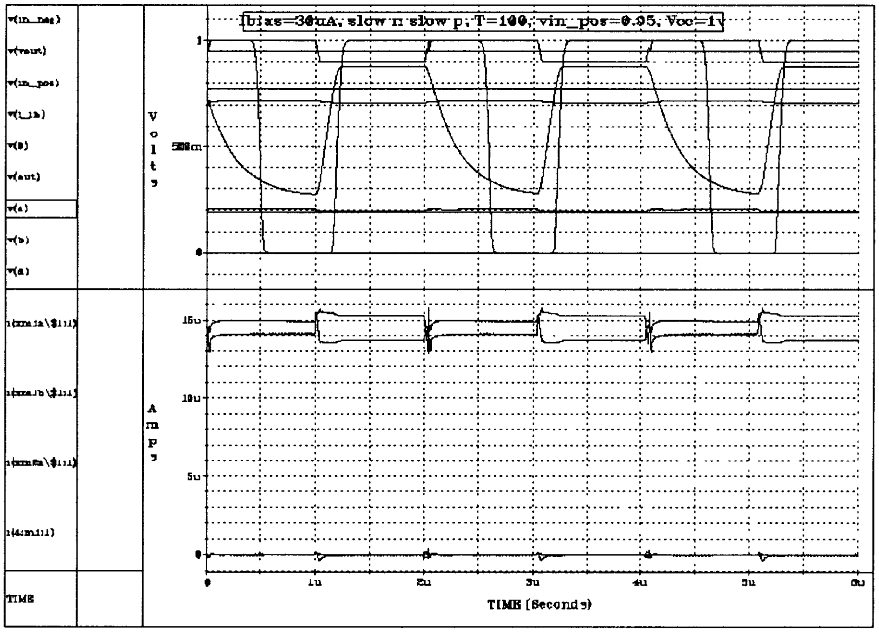


Figure 67: Initial op-amp, resized with Ibias=30uA, vin_pos=0.95, slow n slow p

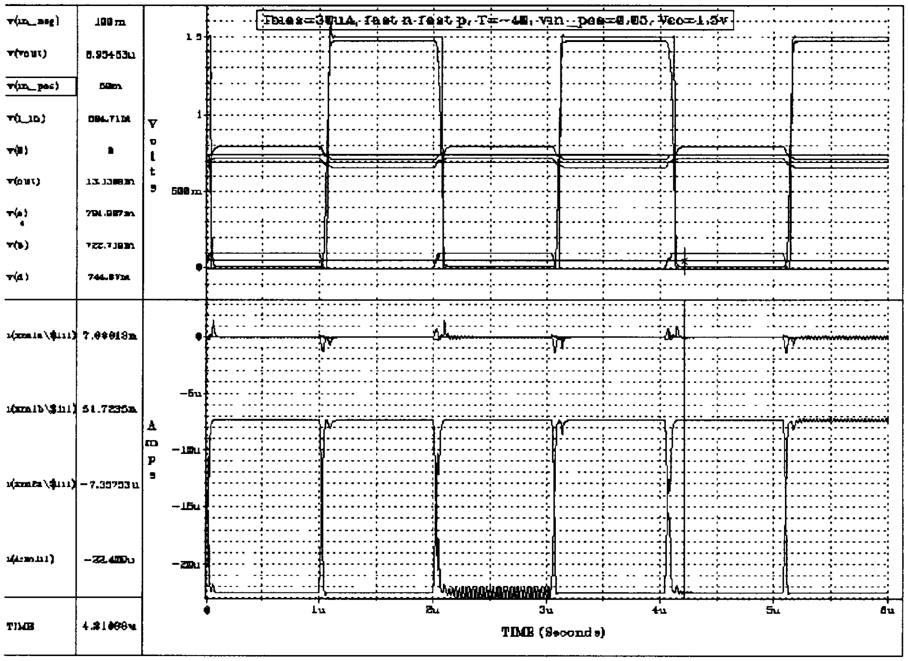


Figure 68: Initial op-amp, resized with Ibias=30uA, vin_pos=0.05, fast n fast p

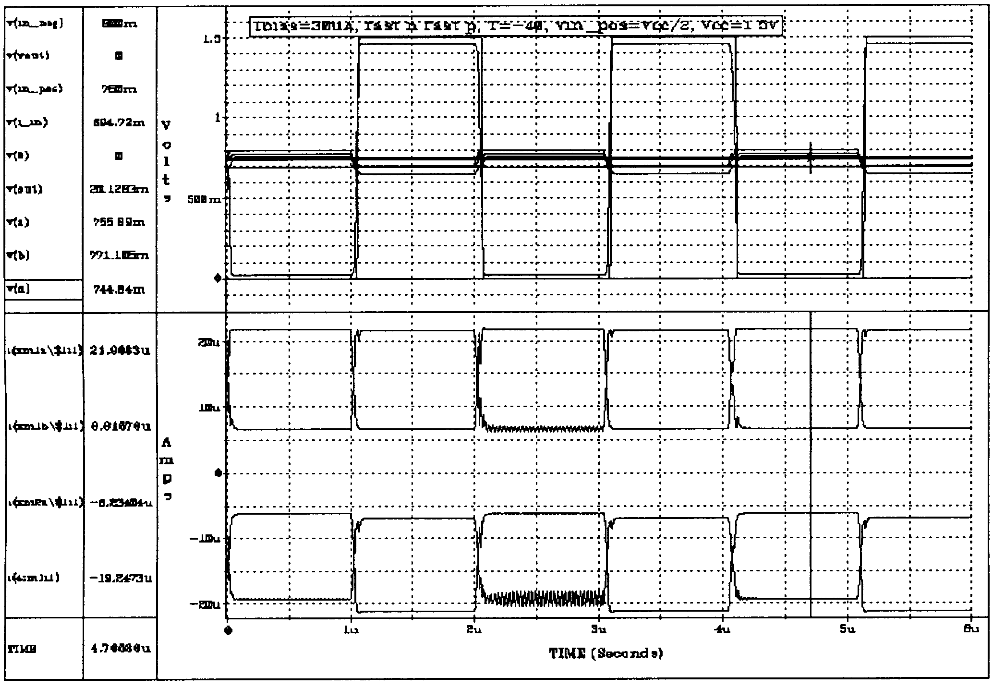


Figure 69: Initial op-amp, resized with Ibias=30uA, vin_pos=vcc/2, fast n fast p

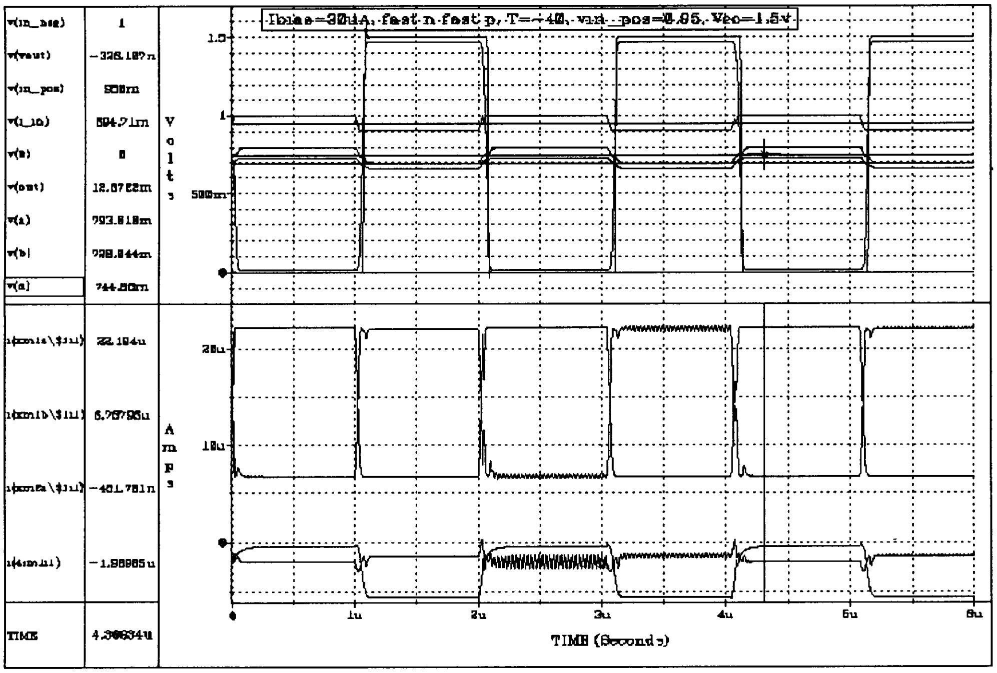


Figure 70: Initial op-amp, resized with Ibias=30uA, vin_pos=0.95, fast n fast p

5.10 Appendix J: New op-amp simulation results

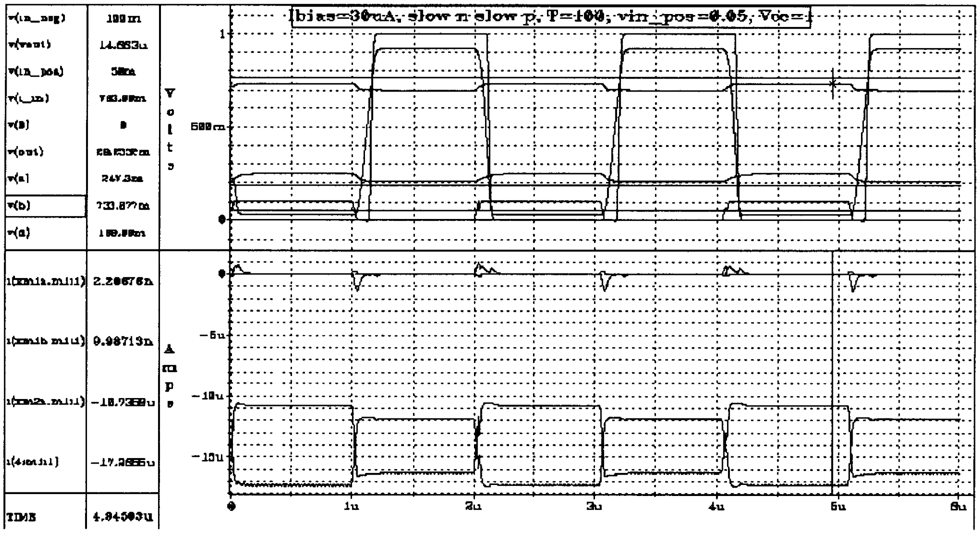


Figure 71: Op-amp with cascode and prime devices, vin_pos=0.05, slow n slow p

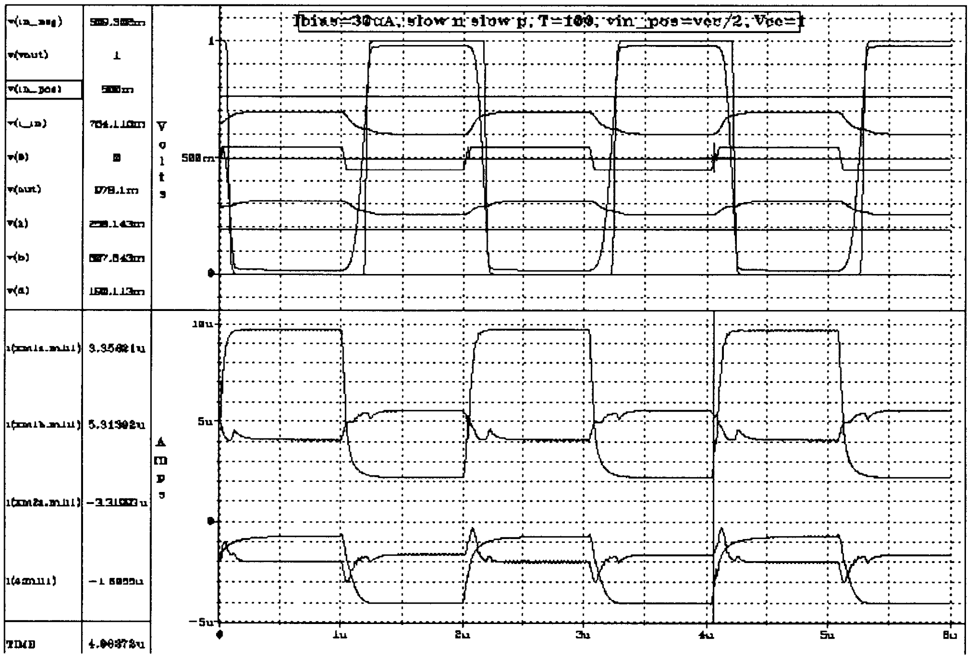


Figure 72: Op-amp with cascode and prime devices, vin_pos=vcc/2, slow n slow p

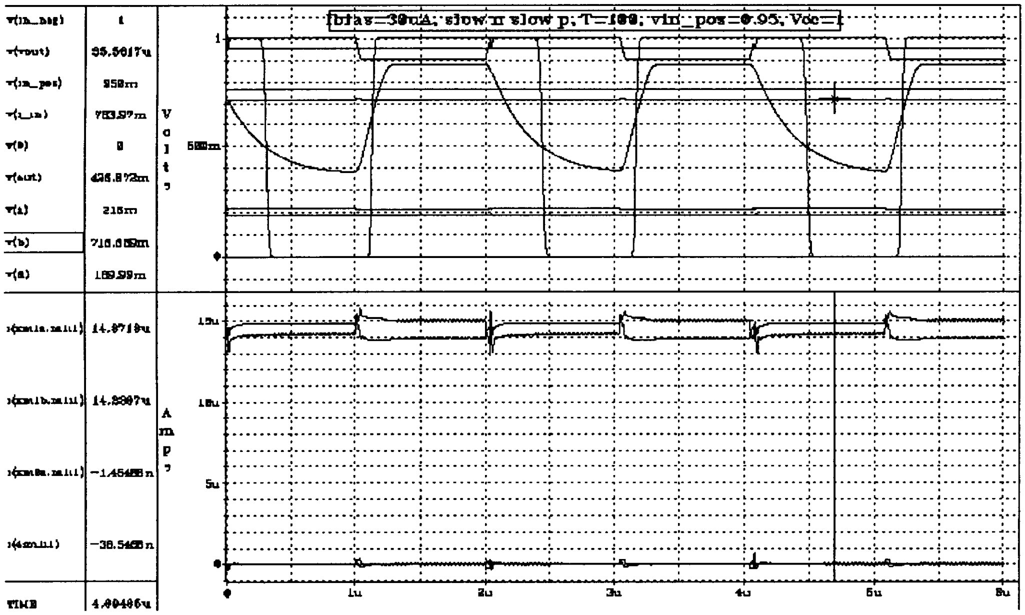


Figure 73: Op-amp with cascode and prime devices, vin_pos=0.95, slow n slow p

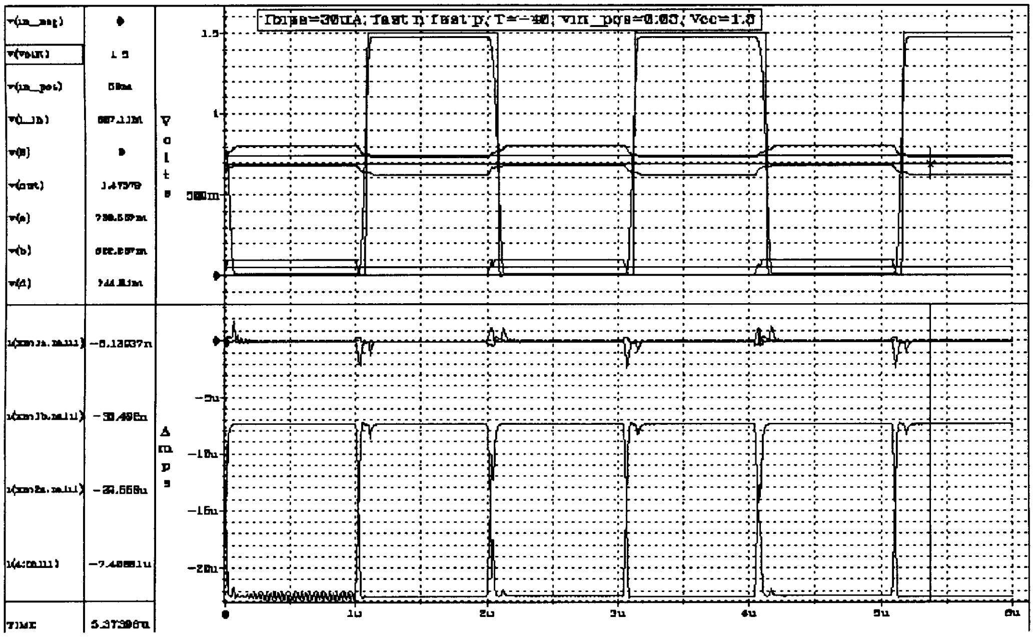


Figure 74: Op-amp with cascode and prime devices, vin_pos=0.05, fast n fast p

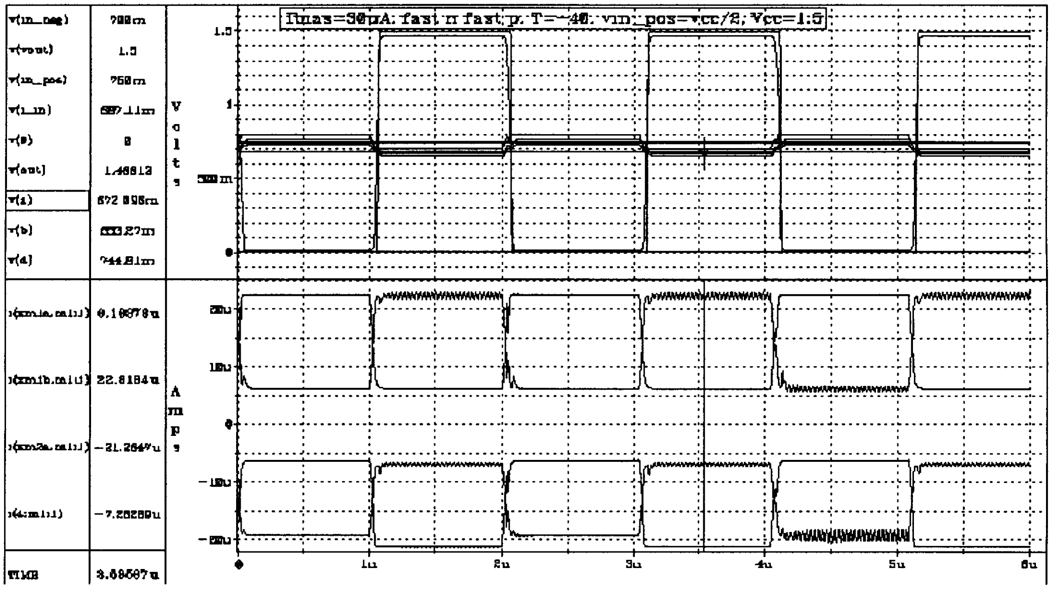


Figure 75: Op-amp with cascode and prime devices, vin_pos=vcc/2, fast n fast p

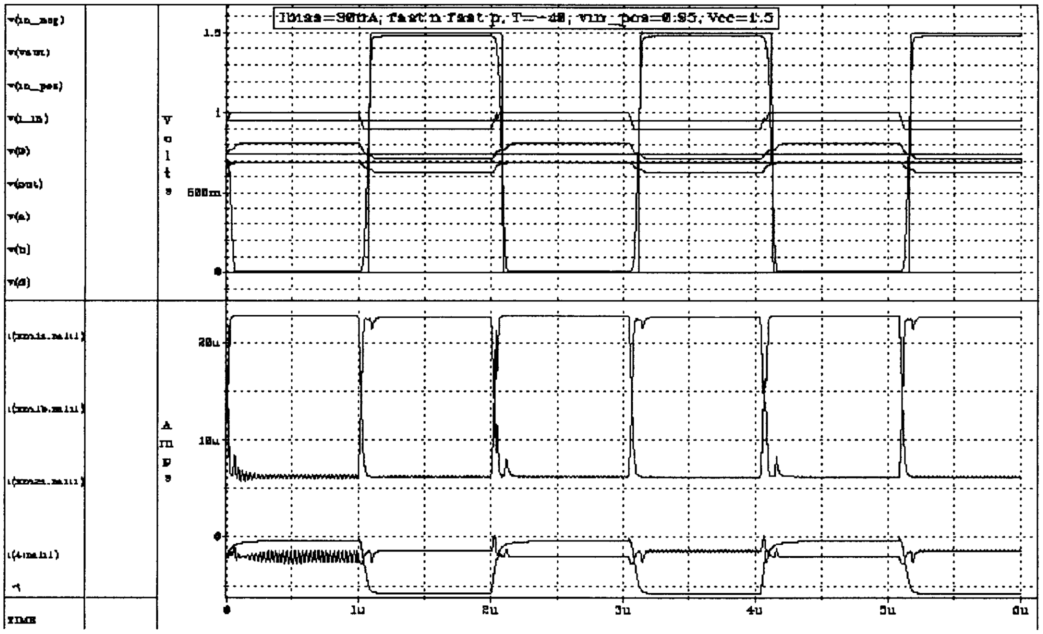


Figure 76: Op-amp with cascode and prime devices, vin_pos=0.95, fast n fast

6. References

Avanti. Star-Hspice User's Manual. 3 vols. Campbell, CA: Meta-Software, Inc., 1996.

Allen, Phillip, and Douglas R. Holberg. CMOS Analog Circuit Design. New York: Oxford University Press, 1987.

Allen, Phillip E., Benjamin J. Blalock, and Gabriel A. Rincon. "Low Voltage Analog Circuits Using Standard CMOS Technology," Proceedings 1995 International Symposium on Low Power Design, 1995.

Chandrakasan, Anantha P. and Robert Brodersen. Low Power Digital CMOS Design. Boston, MA: Kluwer Academic Publishers, 1995.

Dipert, Brian. "Data Storage in a Flash," EDN Access for Design, by Design. (HTTP://www.ednmag.com/current/070397/14df_01.phtm), July 1997.

Dipert, Brian, and Markus Levy. Designing with Flash Memory. San Diego, CA: Annabooks, 1993.

Ferri, Giuseppe and Willy Sansen. "A Rail-to-Rail Constant-gm Low-Voltage CMOS Operational Transconductance Amplifier," IEEE Journal of Solid-State Circuits, Vol. 32, No.10, October 1997.

Flash Memory. 2 vols. Santa Clara, CA: Intel Corporation, 1997.

Fuse, Tsuneaki, Yukihito Oowaki, et al. "A 0.5V 200MHz 1-Stage 32b ALU using a Body Bias Controlled SOI Pass-Gate Logic," ISSCC Digest of Technical Papers (CD-ROM), 1997.

Gonzalez, Ricardo, Benjamin Gordon, and Mark Horowitz. "Supply and Threshold Voltage Scaling for Low Power CMOS," IEEE Journal of Solid-State Circuits, Vol.32, No.8. August 1997.

Hogervorst, Ron and Johan H. Huijsing. Design of Low-Voltage, Low-Power Operational Amplifier Cells. Boston: Kluwer Academic Publishers, 1996.

Kawahara, Takayuki, et. al. "Bit-line Clamped Sensing Multiplex and Accurate High-Voltage Generator for 0.25um Flash Memories," IEEE Solid-State Circuits Conference, Vol. 39, February, 1996.

Otsuka, Nobuaki and Mark A. Horowitz. "Circuit Techniques for 1.5V Power Supply Flash Memory," IEEE Journal of Solid-State Circuits, Vol.32, No.8. August 1997.

Rabaey, Jan M. Digital Integrated Circuits. New Jersey: Prentice Hall Electronics and VLSI Series, 1996.

Sakurai, Satoshi and Mohammed Ismail. Low-Voltage CMOS Operational Amplifiers: Theory, Design and Implementation. Boston, MA: Kluwer Academic Publishers, 1995.

Tedrow, Kerry, et al. "High precision Voltage Regulation Circuit for Programming Multiple Bit Flash Memory," U.S. Patent No: 5546042. Issued: Aug. 13, 1996.

Tedrow, Kerry, et. al. "Precision Voltage Reference, " U.S. Patent No: 5339272. Issued: Aug. 16, 1994.

Tedrow, Kerry, Jahanshir J. Javanifard, et. al. "System having multiple phase boosted charge pump with a plurality of stages," U.S. Patent No. 5524266. Issued: June 4, 1996.

Weste, Neil and Kamran Eshraghian. Principles of CMOS VLSI Design. 2nd edition. Reading, MA: Addison-Wesley Publishing Co., 1993.

Wong, Louis S.Y. and Graham A. Rigby. "A 1V CMOS Digital Circuits with Double-Gate-Driven MOSFET," ISSCC Digest of Technical Papers (CD-ROM), 1997.

Wu, Jieh-Tsomg, et. al. "1.2V CMOS Switched-Capacitor Circuits," IEEE Solid-State Circuits Conference, Vol. 39, February, 1996.