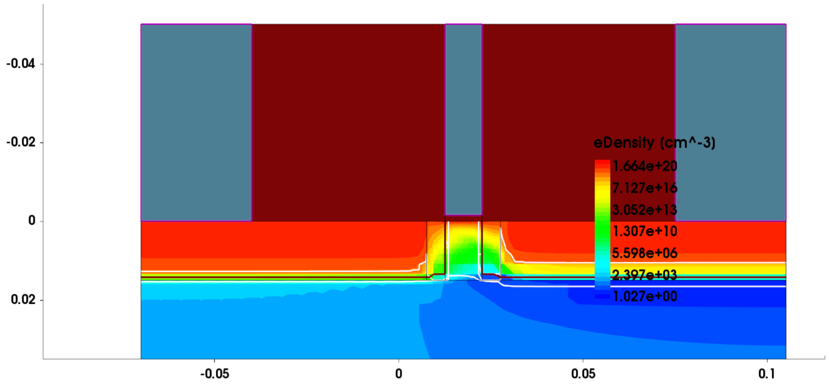
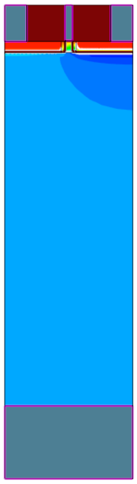
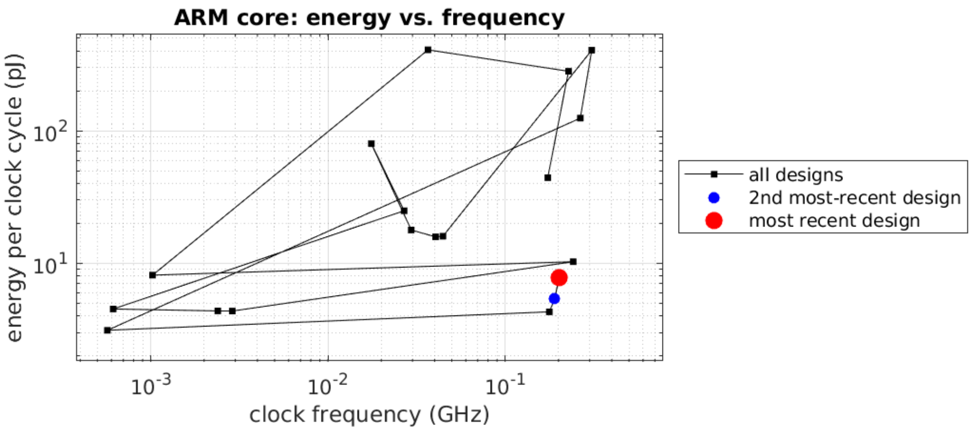
Erick Rodriguez
Hannah Savoldy
Franklin Zhang

SAMTEL: Final Report







MOSFET Changes:

1. Lightly Doped Drain
2. Shallow Junctions
3. Decreased size of channel by decreasing spacers
4. Decreased oxide thickness
5. Increased channel doping
6. Scaled down
7. Decreased Vdd

ARM Microprocessor Performance:

Leakage Energy: 4.757 pJ
Dynamic Energy: 2.972 pJ
Total Energy: 7.73 pJ
Speed: 201 MHz
EDP: 38.29

See appendix for final mosfet geometry file

Equations

1. $V_t = V_{FB} + 2\phi_{Bulk} + \frac{t_{ox}}{\epsilon_{ox}}\sqrt{4 * \phi_{Bulk} * \epsilon_{si} q * N_a}$
2. $I = \frac{W}{L} * \mu * C_{ox}(Vgs - Vt) * Vgs$ for low V_DS,
3. $I_{off} = W * K_{off} exp(-\frac{V_t}{KT/q})$
4. $E_{leak,single} = V_{DD} * I_{off} = V_{DD} W * K_{off} exp(\frac{-V_t}{KT/q})$
5. $E_{dyn,single} = \frac{1}{2}(C_{out} + C_{wire} + F.O.* C_{in}) * \alpha * (V_{DD} - V_{th})^2$
6. $T_{clk} = L.D.* \frac{1}{2} ln(2) * [R_{on} * (C_{out} + C_w + F.O.* C_{in}) + R_w * (C_{w/2} + F.O.* C_{in})]$

Summary

1. Problem 1: Improve Subthreshold Slope

We calculated the subthreshold slope of the basic mosfet from the vgs and current data and realized that this slope was around 200mv/decade, which was much greater than the 60mv/decade theoretical bound. This means that a lot of energy is being wasted due to problems such as short channel effects and poor gate control over certain regions of the channel. In order to combat this, we implemented LDD, shallow junctions, and changed the geometries of the mosfet. The subthreshold slope was decreased from 200mv/decade to 60mv/decade. Along the way, we were able to greatly decrease the leakage energy, but created new issues. Our clock frequency greatly decreased.

2. Problem 2: Increase Clock Speed

After improving our subthreshold slope, our next issue was that the speed of the mosfet was way too slow. Our clock speed decreased from 198 MHz to 27 MHz after making the changes in addressing problem 1. Our 27 MHz speed is way below the 200 MHz clock speed in the design requirements. Before we increase efficiency we must make it faster. The speed is dependent on the RC constant of the circuit as shown in equation 6. Thus we want a low R and a low C. Decreasing resistance while having a constant voltage implies an increase in the on current. Thus we want to maximize current based on equation 2. We increased doping in the channel and modified the capacitor geometries. The change in capacitor geometries impacts both R and C, which greatly reduces the delay. From MOSFET_10 to MOSFET_15, we were successfully able to increase the on-current from 270 uA/m to 840 uA/m. The normalized capacitance reduced from 4.75 to 0.9. The speed increased to 308 MHz. We were successful in increasing speed!

3. Problem 3: Reduce Energy Consumption

The EDP for MOSFET_15 is at a 1319 largely due to the large on current. The threshold voltage is also very small at .1869. The leakage current is high at 11078. We want to reduce leakage current and reduce our on current. Leakage energy accounted for more than 95% of our energy use. Based on equation 1, 4, and 5, we decided to decrease Vdd and increase doping in the channel to lower Vth. We successfully reduced our EDP to 38.29 while keeping our clock speed above 200 MHz.

Introduction

Our first step was to obtain the Ioff, Ion, Cgate, and Armcore data for the basic MOSFET given at the start. The Armcore data refers to the frequency, leakage energy, dynamic energy, subthreshold slope, and EDP. Looking over the course material, we recorded the key equations dictating the MOSFETs behavior, shown above. The initial capacitance Cgate was calculated manually, but Franklin then developed a program that extracts the positional data of the MOSFET geometry and outputs the normalized gate capacitance (see get_capacitance.py). Having obtained the initial data, we then discussed what first steps to take in order to begin altering the MOSFET.

The initial approach was to have each member work on a MOSFET in parallel, and then combine the changes. However, meeting with Professor Shulaker gave us valuable guidance as to how to approach the project. Individual parallel changes to the MOSFET cannot be concatenated due to MOSFET modifications not combining linearly. A smoother approach would be to instead implement changes that increase energy efficiency and cause short channel effects, mitigate the effects, and repeat the process. With this general road plan in mind, we began our MOSFET modification, taking in mind the variables we wished to decrease or increase in the equations of interest.

We began by attempting to change threshold voltage (V_t) by changing the thickness of the oxide. First, we decreased the oxide thickness which should increase V_t (eq 1). For low V_ds, a higher V_t would increase the current and energy consumption (eq 2). We saw an expected increase in off current and a decrease in speed which was probably due to an increase in Cox (eq. 2). Subsequently increasing the oxide thickness from .005 to .01 gave us the anticipated impact. Energy consumption was decreased as expected, but leakage energy comprised 87% of the energy use. The subthreshold slope was also an unfortunate 699 mV per decade.

The next focus was to reduce power consumption by reducing the leakage and dynamic energy. This was done by reducing Vdd from 1.8 to .9 V. This change gave a much lower energy use, as well as about 450 mV per decade. These two factors improved, but the clock frequency was a miniscule 1MHz. This let us know that the Vdd change was too drastic for that MOSFET configuration, which inspired the subsequent change of lowering the Vdd from 1.8 to 1.5 V.

Addressing Problem 1: Improve Subthreshold Slope

1. Problem

The starting MOSFET has a subthreshold slope of 201mV/decade, which is significantly higher than the ideal 60mV/decade. The device isn't very energy efficient because some of its energy isn't directly used to increase the number of carriers due to either short channel effects or poor gate control over certain regions. This creates problems with altering the device to optimize the EDP. We must first address this problem before we continue because this creates too many nonidealities that are hard to model and predict.

2. Solutions

2.1. Lightly Doped Drain (LDD)

Rationale

By adding a region next to the drain junction that is less heavily doped, the strength of the electric field resulting from the drain voltage is mitigated, decreasing short channel effects. This technique is mirrored on the source side to maintain symmetry for practical uses. LDD is superior to simply lowering the doping of the entire junction because doing so would increase the contact resistance, increasing the EDP.

Details

To implement LDD, we set the doping concentration of Phosphorus in the drain_n_ext and source_n_ext regions to 5*10^(10) (as opposed to the 5*10^(19) concentration in the source and drain junctions). To see how doping values would affect the outcome, we also tried setting the LDD region Phosphorus doping to 1*10^(5).

Results

The results were what we anticipated: LDD decreased the off current by nine orders of magnitude and increased the on current by 2 orders of magnitude. Increasing the doping concentration caused small benefits to both currents but not nearly on the same scale that doping concentration changed. Some negative effects of this was that the clock frequency drastically decreased.

	Before LDD Na = 5*10^(19) (MOSFET_6)	LDD Na = 5*10^(10) (MOSFET_7)	LDD Na = 1*10^(5) (MOSFET_8)
I_on	401 um	4.7 uA	3.9 uA
I_off	117 nm	3.8*10^(-7) nA	3.7*10^(-7) nA
Slope	200mV/decade	125mV/decade	125mV/decade
Clock Freq	242 MHz	2.8 MHz	~

2.2. Shallow Junctions

Rationale

To further attenuate short channel effects, we can simply make the drain region (and source region for symmetry) shallower. This decreases the amount of the channel that is controlled by the drain electric field and increases the control of the gate by removing the depletion region that in closer to the body, thus making the charge sharing factor of the device closer to ideal. This should allow us to create a more efficient MOSFET. In order to extend the gate’s control we can increase our capacitance by decreasing the oxide thickness.

Details

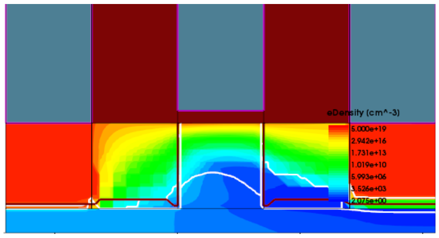
In MOSFET_9, we decreased the vertical size of the source and drain regions from 0.035 nm to 0.015 nm. We also decreased the distance of the oxide capacitor from .001 nm to .005 nm.

Results

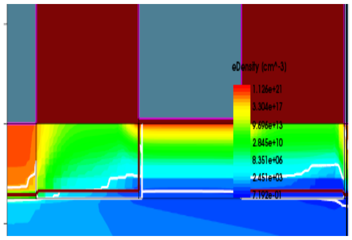
With MOSFET_9, we actually saw a decrease in on current and an increase in off current, which was the opposite of the intended result. Based on the device model showing electron density, we noticed that there appeared to be a pinch off of sorts in the lightly doped regions where electrons were not flowing at a very high density (shown in the lower left image). We do see that the change in oxide thickness allowed the gate to extend its control deeper into the channel (more electrons near the oxide).

	MOSFET_8	MOSFET_9
I_on (Amps)	3.88744e-06	2.87045e-11

I_off (Amps)	3.71152e-16	5.47906e-15



Before Modifications (MOSFET_8)



After Modifications (MOSFET_9)

2.3. Decreasing Gate and Drain-Source Gap

Rationale

Looking at the electron density of MOSFET_9, we see that there is an area in the channel that can't be controlled by the gate. This area is in between the gate controlled region of the channel and the negatively doped regions of the drain and source. The gate's control isn't able to extend to this region simply because it is too far. To combat this, we can make this region smaller.

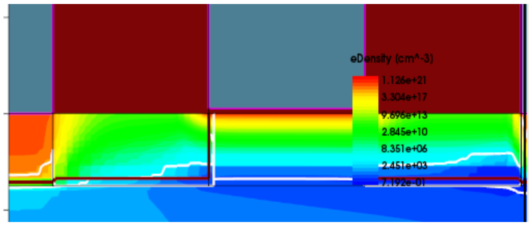
Details

MOSFET_9 was taken and the width of the drain and source was increased, while keeping the gate width the same to create MOSFET_10.

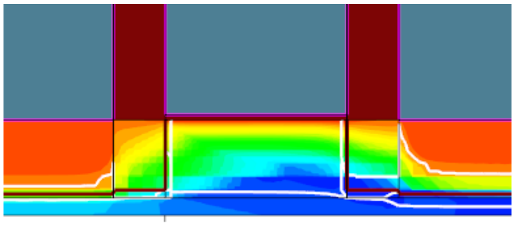
Results

We achieved a significantly higher current value for both on and off states (shown in the lower right image). MOSFET_10 achieved a subthreshold slope of 80mV/decade, meaning that we now that our slope is much closer to the theoretical threshold, we can now work on improving speed.

	MOSFET_9	MOSFET_10
Slope (mV/decade)	200	80
I_on	2.87045e-5 uA	270 uA
I_off	5.47906e-6 nA	1.95 nA



MOSFET_9 (before)



MOSFET_10 (after increasing gap)

Addressing Problem 2: Increase Clock Speed

1. Problem

Even though our current MOSFET has a close to ideal subthreshold slope, the clock frequency is far too low. The specifications for our final device require it to be 10 times faster. To create an effective MOSFET, it must be efficient and fast enough for the ARM microprocessor.

2. Solutions

2.1. Decrease Threshold Voltage

Rationale

We had the idea that decreasing V_th would increase the on current and subsequently speed up the device. Based on equation 2, I would increase if V_th is smaller. Increasing the on-current would hopefully mean a decrease in R, which decrease the RC constant.

Details

To achieve a lower V_t, we decided to decrease the width of the oxide from 0.001 to 0.0005 nm.

Results

This attempt had the opposite of the desired effect. Instead of speeding up the device, decreasing the oxide width actually decreased clock frequency of MOSFET_11 by an order of magnitude. On further investigation, we found that equation 6 explained this result. Decreasing the oxide increased the capacitance of the oxide, which comprises most of C_on, drastically decreasing the speed.

	MOSFET_10 t_ox = .001	MOSFET_11 t_ox = .0005
Clock Freq	26 MHz	17 MHz
Normalized Capacitance	4.7	8.9

2.2. Decrease Oxide Capacitance

Rationale

Because decreasing the width of the oxide had such a large impact on C_gate, we decided that increasing it may decrease the capacitance enough to speed up the device. We can see this by looking at equation 6.

Details

We increased the width of the oxide to 0.0015 in MOSFET_12 and increased the width of the spacers by 0.003 in MOSFET_13.

Results

This had the desired result, but it was not very drastic. The clock frequency of MOSFET_12 increased only by 3MHz, and MOSFET_13 increased only by 14MHz, from the device at the beginning of this step, so it is still far too slow to meet specifications.

	MOSFET_10 t_ox = .001	MOSFET_11 t_ox = .0005	MOSFET_12 T_ox = .003
Clock Freq	26 MHz	17 MHz	29 MHz
Normalized Capacitance	4.7	8.9	3.36

2.3. Scale Down

Rationale

As shown by Dennard scaling, decreasing the dimensions of the device, while maintaining high enough V_t and V_DD values, greatly increases the speed of the MOSFET because we can reduce the capacitance to reduce the RC constant. Also we can reduce R by increasing the current (Ohms Law). Current is proportional to the electric field (E) and inversely proportional to the channel length (E = Vdd/L).

Details

In MOSFET_13, MOSFET_14 and MOSFET_15, we decreased the length of the channel from 0.035 to 0.01 nm and decreased the capacitance by increasing the thickness of the spacers.

Results

We were successfully able to increase our speed dramatically by reducing our channel size and capacitance. The on-current increased by an order of magnitude from MOSFET_12 to MOSFET_15.

	MOSFET_12	MOSFET_15
Clock Freq	29 MHz	308 MHz

Addressing Problem 3: Reduce Energy Consumption

1. Problem

The third performance issue was that of energy consumption, and it was also the primary goal of the project. The difficulty was in having to maintain a minimum speed of 200 MHz while also minimizing energy use. However, reducing energy cost comes with the trade-off of reducing speed, so solutions had to be devised in order to improve performance enough to lower the voltages needed to operate at sufficient speeds. Starting off from MOSFET_15, the off current is 11078 nA and the on current is 840 uA which are both very high, giving us an EDP of 1319.

2. Solutions

2.1. Decrease Vdd

Rationale

The previous solution's large tradeoff of energy and speed was due in part to the shift in proportions between Vdd and Vt. After reducing the gate size and reducing Vt, the subsequent step was to reduce Vdd and bring the two voltages closer to one another but at a reduced magnitude compared to its original values. As per equations 4 and 5, this would have the desired effect of bringing down both leakage and dynamic energy.

Details

Vdd's reduction was a trivial matter, in that we had free control over its applied value so long as it remained between 400 mV and 4 V. Changing its value inside the simulation control text file brought it down from 1.5 V to 1.0 V.

Results

As anticipated, reducing Vdd brought down the on and off currents to less dramatic values. I_on was reduced from 800 uA to around 450, which is preferable due to not wanting to bring I_on down too much. The off current went down to around 4,000 nA, which is greatly improved from before but still extremely high. The EDP was brought back down to the 400's range, but with a clock speed remaining above 200 MHz. This was a great improvement over the earlier 400 EDP iteration with a pitiful clock speed of 40 MHz.   With speed in check and voltages reduced, the final step was to reduce I_off as much as possible while remaining above 200 MHz.

2.2. Increase Channel Doping

Rationale

With the geometry and voltages set, our attention turned to the doping concentrations. We decided to increase the doping of the channel in order to reduce the current through the channel. Bringing down the current would then lower the dynamic and leakage energy, but would also slow down the clock speed. This trade-off meant our limit to increasing the doping concentration would be until the clock speed was brought down as close to 200 MHz as possible. This would thereby minimize the energy cost while having us remain at an acceptable speed.

Details

The doping increase would have to be tuned due to the precise nature of this solution. We would have to get as close to 200 MHz as possible without going under it. The first doping concentration change was an increase to 9e19, a value near the upper limits of our allowed concentration. This value put us greatly below our minimum clock speed but gave us an upper bound to start with. Bringing the doping concentration down to 1e19 put us closer at 177 MHz. Acknowledging our proximity to the desired value, the subsequent tweaks were much smaller, going down to 9e18 before finally settling on our final concentration of 8e18.

Results

The results of each doping concentration iteration informed us of the direction and approximate magnitude of the subsequent change. The first increase to 9e19 brought our currents down, but to extremely tiny values too far in the other extreme to be of value. The decrease to 1e19 brought the values to more reasonable levels. The I_off current was now 28.6 nA, and I_on was 317.9. These values were with a clock speed of 177.6 MHz, which was still below our minimum but was much closer to the end value. The 9e18 doping concentration brought the clock speed up to 190.8 with an EDP of 28.41, indicating only a slight shift in concentration was needed. The change to 8e18 brought the clock speed to its final value of 201 MHz. Unfortunately, this came with the unexpected side effect of a large relative increase in EDP from 28 to 38 from only a slight tweak in concentration. However, this would be the best result we would be able to obtain using this solution method, as any further change in concentration would only move the clock speed further from 200 MHz.

Appendix

- Work Folder: This contains all the work we did for this project, including code https://drive.google.com/drive/folders/1OqAX4sR_2Spp_3uf4PmsXcBHt1I9HegX?usp=sharing
- Get_capactiance.py: a python script used to get the capacitance by looking at the geometry files of the mosfet.
- Project: A detailed google doc that summarizes every change we made.
- Final mosfet design: See Drive:~/mosfet_20/mosfet_20.txt