A 2-Mask NMOS Process Design Fabricate and Test Module for Use In Microelectronics Instruction and Process Development

D. W. Parent
Electrical Engineering Department
San Jose State University
One Washington Square Hall, San Jose CA 95192-0084

Abstract— We have developed a simplified 2-mask n-type Metal Oxide Semiconductor (NMOS) transistor process design and verification module for electrical engineering students enrolled in the Microelectronic Manufacturing Methods class/laboratory at San Jose State University. We have run this module for three years and have found that the simplified process allows the students to learn more because they have the time to design the process fabricate and test in one semester. Student learning is also enhanced because it allows students to make and correct mistakes in the processing the devices. We have also found that the simplified process saves time in process development of more complex processes, by reducing the number of photolithography steps required to fabricate a transistor.

I. Introduction

To teach semiconductor process development is resource intensive (instructor time, masks, chemicals, etc.)[1] and it is a risk of these resources to allow novices to design, fabricate and verify a semiconductor device on their own. If the devices do not work, then students are left with a feeling that they did not master the material. (They also believe that instructor did not master the material!) One method to reduce this risk is to have students design a process with Technical Computer aided Design (TCAD) tools, but stop short of fabricating the devices. This allows students to be exposed to the constraints of an advanced deep sub-micron process that might not be available at a university. On the other hand, students will not develop a sense of the statistical variations, and modeled vs. fabricated device differences that can only be encountered in actual device fabrication. Another method to reduce the risk of non-functional devices is to have the students fabricate a process that has already been designed and verified previously[5]. Although the device yields can be over 90%, the students do not learn how to truely design a process. We feel that best way to reduce this risk is to reduce the time and money to design and fabricate devices is to reduce the complexity of the process to an absolute minimum. We have developed a 2-mask NMOS process that by using two masks instead of four, reduces the amount of time and raw materials that is the main cost to fabricating devices. (It also reduces the time to simulate (TCAD) a device structure.) Since this process is less resource intensive it is acceptable to allow novices to design and fabricate their own transistor process. If a student makes an incorrect decision in the processing or design, there is usually enough time to fix the problem because the number of processing steps is reduced.

This module is part of a senior/graduate level course on microelectronic manufacturing methods. The 3-4 student teams are multidisciplinary so a student has to have CMOS circuit design, advanced device physics or semiconductor process knowledge to be a successful team member.

The two-mask device is a simplified form of a non-self aligned 4-mask NMOS process[3]. The major features of the device are (Fig. 1):

- The diffusion windows for the source and drain also are the contact windows when the source drain spin on glass is removed before metalization.
- The field oxide and the gate oxide are the same layer. Since we are interested in transistors, diodes and MOS capacitors, we do not have to prevent leakage current between adjoining devices.
- The gate oxide between the source and drain acts as self aligned diffusion mask.
- The amount the junction diffuses under the gate is critical. The metal gate (which has to be notched in from the edge of the source and drain) has to overlap at least the depletion widths of the source and drains to have current flow at small drain voltages. This notching is required because the standard deviation of our registration error is 1.0 micron. If the gate metal were to drift right or left it would short the gate to the source or the drain.
- The mask set consists of four different NMOS width/length pairs with the notch distance (gate to source/drain offset) between the as edge of the drawn source/drain regions and the edge of the metal gate varied from 0.0 microns to 8.0 microns (Fig. 1).
- The width of the transistor is the as drawn vertical distance of the diffusion cut. Since there is no field oxide the actual width is larger due to the fact the diffusion areas move about 2 microns into the surrounding substrate.
- The length is the as drawn distance separating the

1-4244-0268-9/06 $20.00 ©2006 IEEE

two edges of the diffusion cuts. The true length is reduced by the diffusion into the substrate as well.

- The 2 Mask set is made out of quartz and chrome and has been provided by the instructor and consisting of sheet resistance structures, junction depth monitors, MOS caps, different sized MOSFETS as well as inverters, current mirrors and ring oscillators. Each MOSFET structure comes with an x and y Venier scale to measure registration error.

II. Module Overview

The model begins with a semiconductor process review , which consists of a lecture presentation of a simple metal gate NMOS process that has an Athena run deck, linked to a process traveler, with measured device characteristics. (This is the same process that some team members have fabricated in a previous class.) The major processes reviewed are oxidation, diffusion, and ion-implantation. Then an NWELL self-aligned process is reviewed using the Java Applet Site hosted by Suny[4]. The final lecture shows all the process steps of the 2-Mask NMOS process in conjunction with a calibrated run deck, and process traveler.

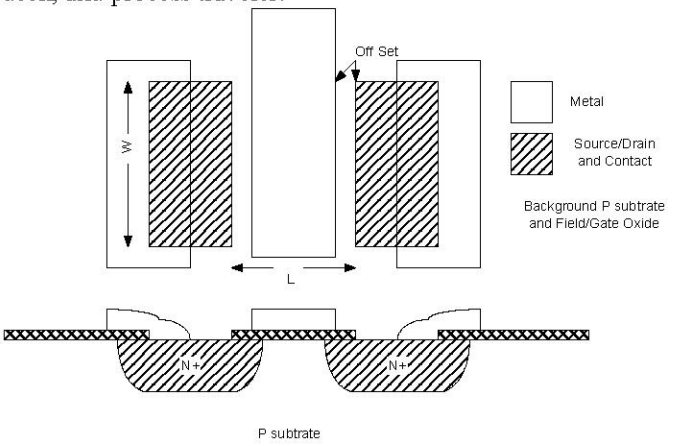

Fig 1. The top view and cross-section of the 2-Mask Process.

In order to design the improve process the students first need to learn the fundamentals of Athena/Atlas by completing a tutorial the guides them though creating their own run deck for a pn junction and learning how and why to refine the mesh[5]. Then the students use the NMOS poly-silicon gate run deck example provided by Silvaco. To prove that they are familiar with TCAD tools the students complete an individual homework assignment to change the channel length and VT of the sample run deck.

The students are then ready to improve the VT of the 2-mask NMOS process. They are given free reign to change any time and temperature of the existing calibrated process run deck (Fig. 2). The students present their improved process to the professor who makes sure that the times and temperatures selected by the students are "reasonable". The students then simulate their design for variations in time and temperatures to anticipate process variances in the laboratory. The final part of the design is to create a process traveler (Fig. 3.) that will be used to guide then through device fabrication in the SJSU micro fabrication facility. The groups give an oral presentation on their process design in the form of a design review.

```
#SCREENING OXIDE GROWTH STEP 2.0 ASSUMED
method fermi compress grid.ox=.003 gridinit.ox=.003
# STEP 2.0 IS ASSUMED
# STEP 2.1 PUSH
diffus time=15   temp=400 t.final=700  nitro
# STEP 2.2 RAMP UP
diffus time=13.3 temp=900 t.final=1000  nitro
# STEP 2.2 STABILIZE
diffus time=10   temp=900   nitro
# STEP 2.2 SOAK
diffus time=30   temp=900   dryo2
#STEP 2.2 PURGE
diffus time=10   temp=900 nitro press=1.00
#STEP 2.2 RAMP DOWN
#diffus time=10   temp=900 t.final=700   nitro
#STEP 2.3 PULL
diffus time=15   temp=700 t.final=400  nitro
```

Fig. 2 Sample code from a calibrated Athena run deck.

2.0		Oxide Growth 1 - Load	• Load cleaned wafers into quartz boat; put boat 6" into furnace. Lowest number, shiny side in first. Note Wafer order: • N2 flow 10 slm Tube furnace heating for few hours
2.1		2 - Push	• Push in slowly to prevent temperature shock using autoloader • 15 min push at 700 Dry N2 • Target temperature = 700°C • Record oven temp in all zones • _____, _____, _____
2.2		3 - Oxidize	• Put on end cap, Leave doors open • 15 min ramp to 900 Dry N2 • 10 minute stabilize at 900, Dry N2 • 100 Min soak Dry O2 (Max Flow) (put on flow restrictor after you switch gas flow or the cap will fly off and break!) • Take off restrictor • 10 minute purge N2 • XX min ramp down to 700 Dry N2
2.3		4 - Pull	• 15 min pull at 700 • Pull out boat slowly, remove, cool under hood

Fig. 3 Sample instructions from the process traveler.

After an equipment process review and safety lecture the students fabricate their devices in an effective open lab[1]. The teams operate independently in lab, asking for help only when something with the process is unexpected.

[1] There are five groups working o different processes at any given time. Therefore even if the instructor is present the students are not being directly supervised at all times.

III. 2 Mask NMOS Process Flow

The starting substrates are <100> P-type silicon wafers with a resistivity of 5 Ω-cm. The substrate doping is so low that given the large observed fixed oxide charge in our lab ~10^{11}q/cm^2, the threshold voltage will be negative for all oxide thicknesses.

To change the substrate doping B diffusion through a screen oxide (~200Å) is carried out[6], after SC1 and SC2 cleaning steps. The thickness of the oxide is critical in determining the final substrate doping (Fig. 4.). If the oxide is too thin the boron concentration at the surface of the substrate will be much higher than the 10^{17}cm^{-3} range, but rather will be in the 10^{20}cm^{-3} range, cause a V_T in excess of 40Volts. Diffusion through a screen oxide is done rather than an implant step due to the fact our facility does not have an implanter. Implant services are available from a local vendor or SNF, but we have found students do not like to outsource their processing steps. The boron diffusion time and temperature are usually the steps the students change to increase the substrate doping. The boron source is spin on glass (SOG) from Emulsitone[7]. The SOG is spun of for 60 seconds at 300rpm and then baked for one minute at 100°C.

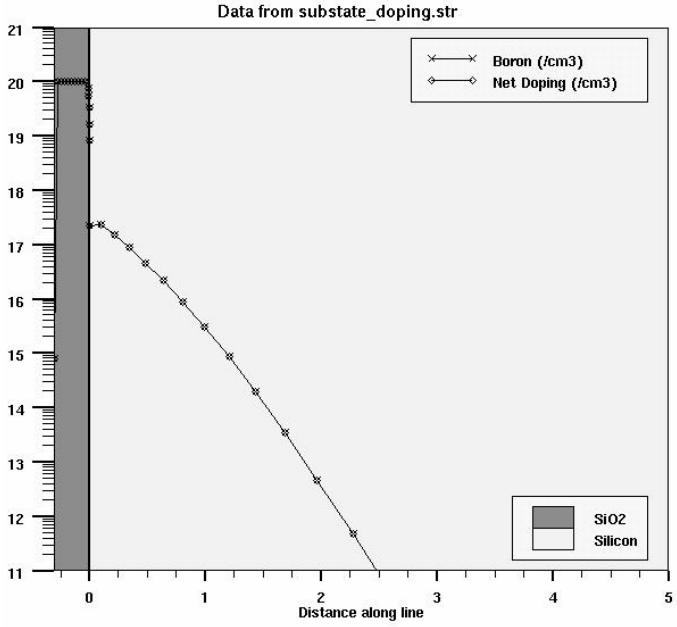

Fig 4. Substrate doping profile after B diffusion step.

Once the substrate-doping step has been carried out the SOG and screen oxide are etched off in buffered hydrofluoric acid (HF). The gate oxide is grown next (Fig. 5.), by means of dry thermal oxidation. The oxidation temperature is 1100°C and a post oxide anneal is done at the oxidation temperature for at least 50minutes. The gate oxide thickness is usually in the range of 3500 Å, as it is over etched in a later processing step. The students adjust the oxidation time but not the post oxide anneal time or the oxidation temperature to increase the threshold voltage.

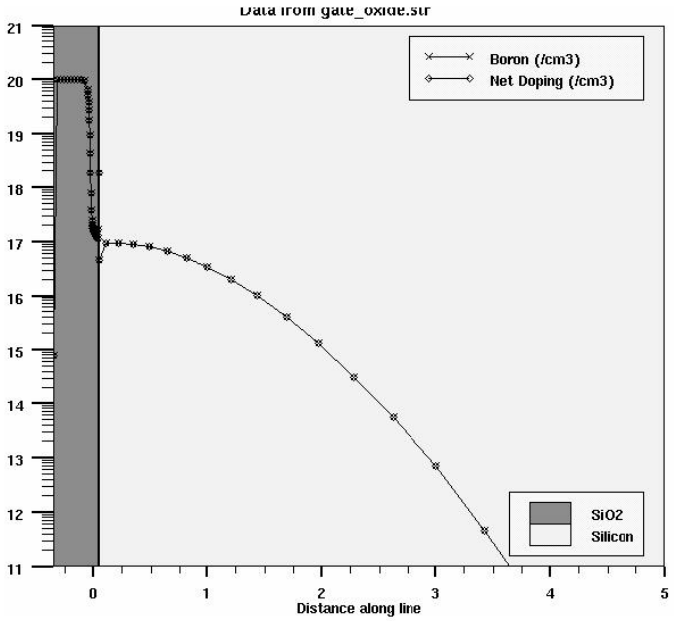

Fig. 5. Substrate doping profile after gate oxidation step.

The first photolithography step (Fig. 6.), source drain diffusion/contact (mask 1) is completed after the gate oxidation step. The source and drains are etched down to below 20Å.

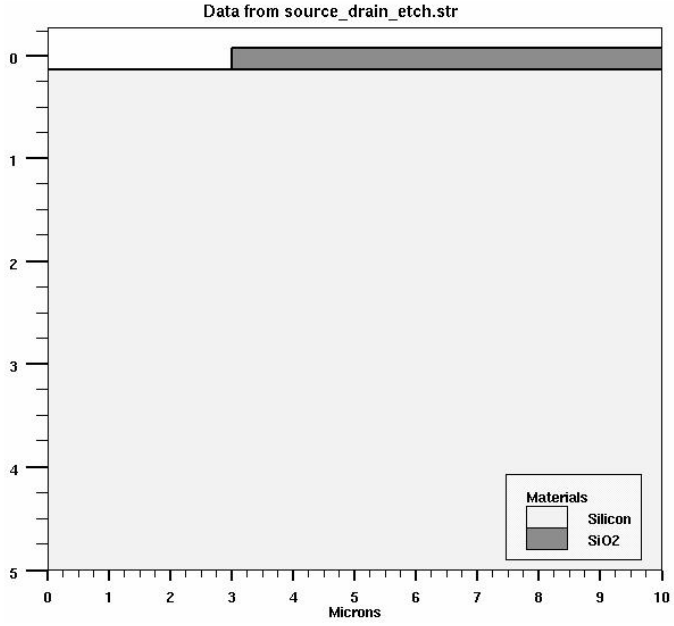

Fig. 6. Source/Drain etch step.

To dope the source and drains, P-SOG is applied to the wafer in a similar manner as the B-SOG. The source and drain diffusion is carried out at 1100°C. The time of the diffusion is critical because the source and drain need to diffuse under the gate to make sure the gate electrode covers at least the depletion widths of each source and drain region (Fig 7.). If the diffusion time is too short, the gate will not overlap the drain and there will be almost zero drain current until the depletion width is extended under the gate due to an increasing reverse bias voltage of the drain. If the diffusion time is too long, the P-SOG can punch through the gate oxide and dope the substrate making normally on MOSFETS. The amount of overlap needed is based on a statical process control (SPC) report on alignment generated by the students earlier in the semester.

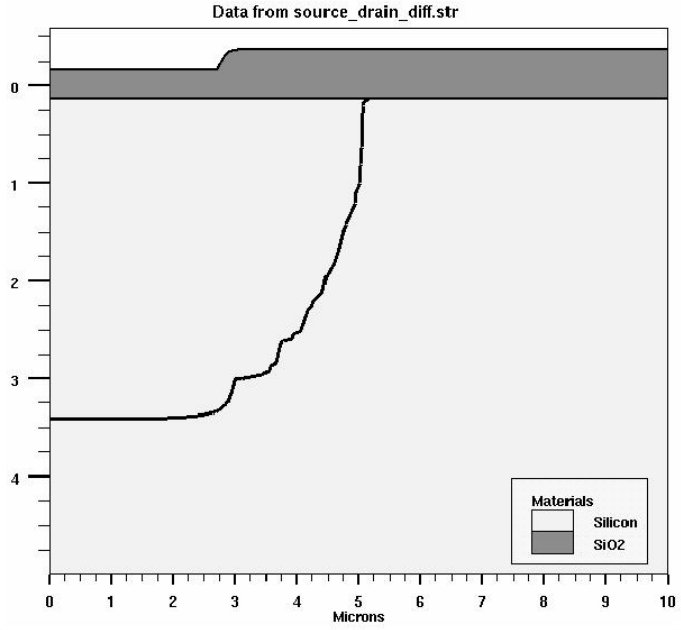

Fig. 7. Source/Drain Diffusion step.

Before the aluminum can be evaporated onto the wafers the P-SOG must be removed from the source and drain regions (Fig. 8.). It was thought at first that the selectivity of the P-SOG would be greater than that of the thermally grown oxide. This is true once the SOG has been removed form the Gate/Field oxide regions, however once the etching of the SOG near the surface of the source and drain regions begins the etch rate of the gate/field oxide remains around 400Å/minute but the SOG etch rate decreases dramatically. Since the SOG has to be removed in order to from a good ohmic contact, 1000Å of oxide maybe removed form the gate/field regions before the source and drain area read 20Å or below as measured with a nano-spec. In some cases the students etch of all the oxide, even after careful monitoring of the etch process.

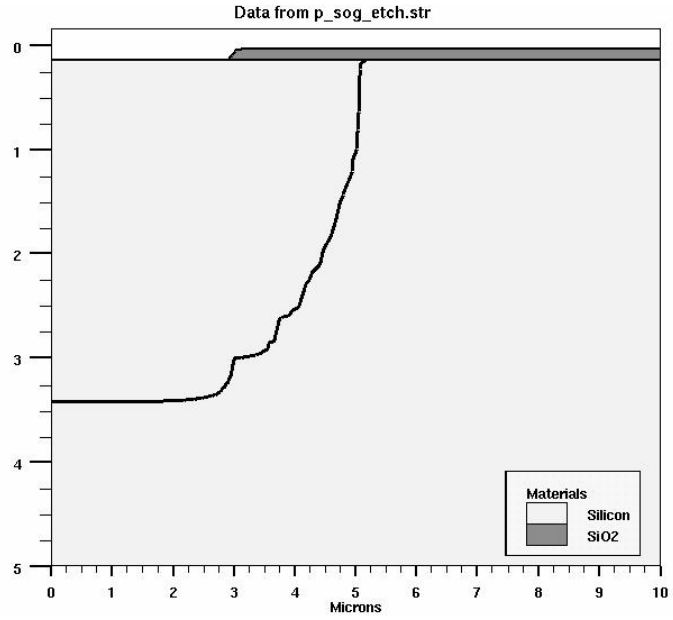

Fig. 8. P-SOG Etch step.

After the metal is evaporated the metal mask photolithography is performed. The metal is etched in Phosphoric acid at 42°C. The wafers are then annealed at 450°C for 30 minutes. The devices are now ready for testing (Fig. 9.). Notice that the junction diffuses almost 2.0 microns under the gate.

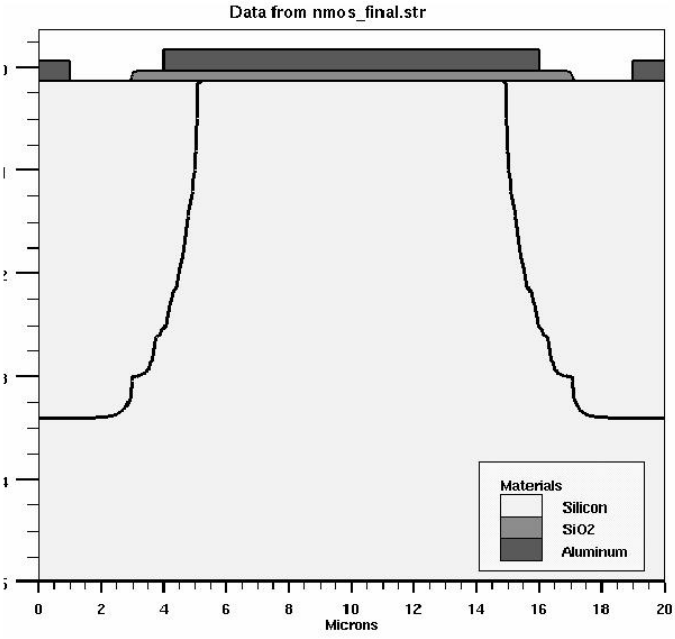

Fig. 9. Completed 2-MASK NMOS Device.

IV. Testing

The students use IC metrics in conjunction with an HP 4145 or 4156 semiconductor parameter analyzer. The plot set-ups

have all been saved previously. This has reduced the training time for testing.

The first test is an I_D vs. V_{DS} to see if transistors were created. Large step sizes are use to speed up the testing. If this test does not work and it can be determined that the gate is not shorted to the source or drain, and then the diodes are checked. If the gate is shorted then devices with a larger offset are chosen to test, as well as those with a minimum registration error. If the diodes work then more MOSFETS are tested with larger voltages. If the diodes are shorted, then usually the wrong dopant was used for the source and drains. (This has happened to me. Since then we have a double check system to make sure we use the correct dopant course.) If the diodes are shorted then a CV test is performed to see if MOS capacitors were formed.

If the process was found to produce working devices, then the students test diodes, capacitors, and try to extract spice level one parameters (Fig. 10.). They also try to find which gate to source/drain offset produces the highest yield of working devices.

Sometimes the process does not produce working devices. The students need extra help to find out exactly why the process did not work. As long as their final report includes a reasoned argument as to why the devices failed and a proposed solution to correct the problem, no points are deducted for the process not working. This is done to promote risk taking in process development. Sometime the students etched of all the gate oxide during the P-SOG removal step. Some times the students did not fully etch the P-SOG form the source and drain regions and there is no ohmic contact to the source and drain.

This past semester the students adopted a very thin screening oxide that was not modeled correctly in their TCAD simulations. Rather than the boron being screened to a substrate surface concentration of ~$10^{17}cm^{-3}$, the boron push through the screening oxide and doped the substrate to excess of $10^{20}cm^{-3}$. This resulted in very high threshold voltages. For those students with gate oxides of around 300Å measuring the V_T by either MOS CV or transistor measurement was impossible as the gate oxide broke down at 20V. One group was successful because their gate oxide was etched down only 1000Å and so a V_T of 60V could be measured with out shorting the structure. They did not fully clear the source drain widows for proper ohmic contact formation and one can see the results of a high source drain resistance (Fig. 11.). Since the source resistance due to the metal N+ contact resistance is extremely large the source and body are not shorted. This gives the appearance of an even higher V_T than one would expect due to the body effect.

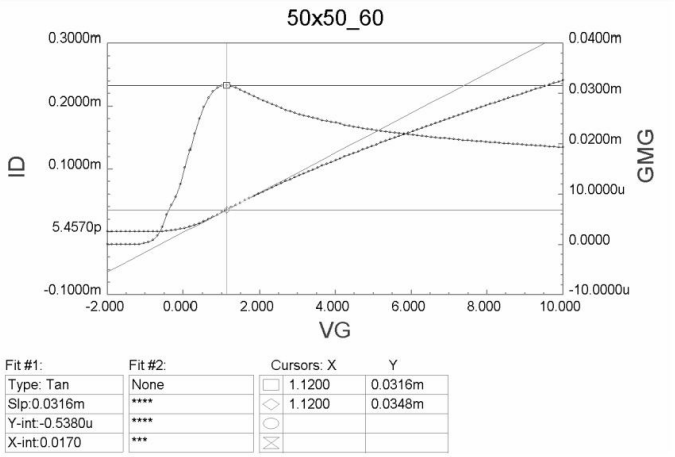

Fig. 10. I_D vs. V_G to extract V_T.

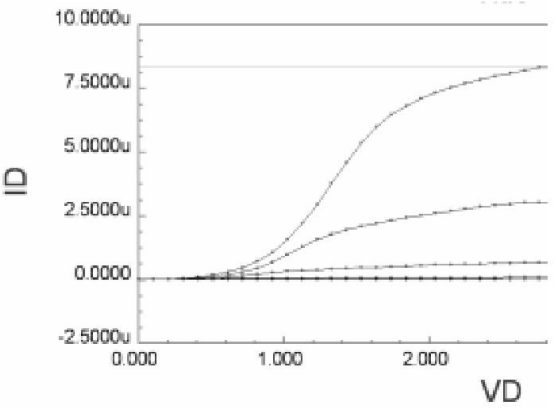

Fig. 11. I_D vs. V_D for a high drain resistance MOSFET.

V. Results/Discussion

The students enjoyed this module even though to complete it required a lot of time spent outside of normal laboratory meeting time to finish. Even though this module was designed as a method to reduce the cost to operate laboratory–based processing instruction, it still is resource intensive to teach. The improvement is that that time is better spent because the students are doing the design, and they have the opportunity to make and correct mistakes.

Students need help interpreting the testing results. The modeling problem of this past spring semester was especially difficult for them to explain as it was the model that was wrong, not the processing. Even when student devices work, they need help to explain how substrate-doping can affects the break down voltage, λ, and even mobility. They do understand the affect of substrate doping on V_T.

After running this module for three semesters, it was observed that the students who made incorrect processing decisions and corrected them seemed to learn more. The students who made a mistake were in general trying to improve the process flow

by aggressively scaling back diffusion time, or oxide thickness. The ability to determine what went wrong in a process flow and correct the problem in one semester is quite an achievement. Personally my first BJT's did not work and it took me years to find out what had a happened (Bad P diffusion modeling). It was this kind of experience that made me a better process engineer.

The two-mask process has also been useful in process development for other more complex processes. We were able to reduce the ramp down time for wafers at oxidation/diffusion temperatures by 60%, because of an experiment that we were able to conduct quickly by using only the two-mask process rather than a four-mask process. We probably would not have invested the time to see if we could reduce than ramp time if we had to use a four-mask process in our experimental plan. We were also able to solve a high source/drain resistance problem in our 4-mask NMOS teaching process that was causing the I_D/V_{DS} curves to be concave downwards in the linear regime instead of concave upwards (Fig. 11.). It turned out the surface concentration required for a good ohmic contact was not being maintained. The contact resistance is not modeled well and it was the fact the source drain doping step came after the gate oxide step (unlike our 4-mask NMOS process) that pointed in the direction of how to fix the contact resistance problem. The next studies we will conduct is how much oxide can be left in the source and drain regions and still get a proper source drain region, and a good ohmic contact.

VI. Conclusions

The 2-mask NMOS process can be used to teach semiconductor design and processing. The reduced the number of processing steps saves time in not only processing the wafers but in the TCAD simulations. The 2-mask process is also helpful in process development. One could use this process to rapidly evaluate high K dielectrics, by simply changing the gate oxide and using implants instead of diffusion.

In addition, this kind of module prepares students for future graduate study in process development because the student will have designed fabricated, and improve a process.

Acknowledgment

We would like to thank Neil Peters for his efforts in managing the MPE laboratory, Kindness Israel for extensive IT support, Linda Shnell of Cadence Design Systems for access to the Virtuoso tool set, and Jerry Kissinger of Intel for the donation of testing equipment and Linux stations.

References

[1] D. W. Parent, E. J. Basham,Y. Dessouky, S. Gleixner, G. Young, E. Allen, "Improvements to a Microelectronic Design and Fabrication Course", IEEE Transactions on Education, Vol. 48, No. 3, pp. 497-502 (2005).

[2] D. W. Parent, Y. Dessouky, S. Gleixner, G. Young, E. Allen, " Microelectronics Process Engineering Program at SJSU," *Proceedings of the 15th Biennial IEEE University/Government/Industry Microelectronics Symposium,"* Richmond , VA pp. 128-134 (2001).

[3] D. W. Parent, http://www.engr.sjsu.edu/MatE129/Process%20Handbook.htm Accessed 12 June 2006.

[4] http://jas.eng.buffalo.edu/ Assessed 12 June 2006.

[5] D. W. Parent, "Silvaco Tutorial", http://www.engr.sjsu.edu/dparent/Silvaco/silvaco.pdf, Accessed 12 June 2006.

[6] G. H. Bernstein, R. J. Minniti, X, Huang, "Advanced IC processing laboratory at the University of Notre Dame" IEEE Transactions on Education, Vol. 37, No. 4, pp. 334-340 (1994).

[7] http://www.emulsitone.com/, Accessed 12 June 2006.