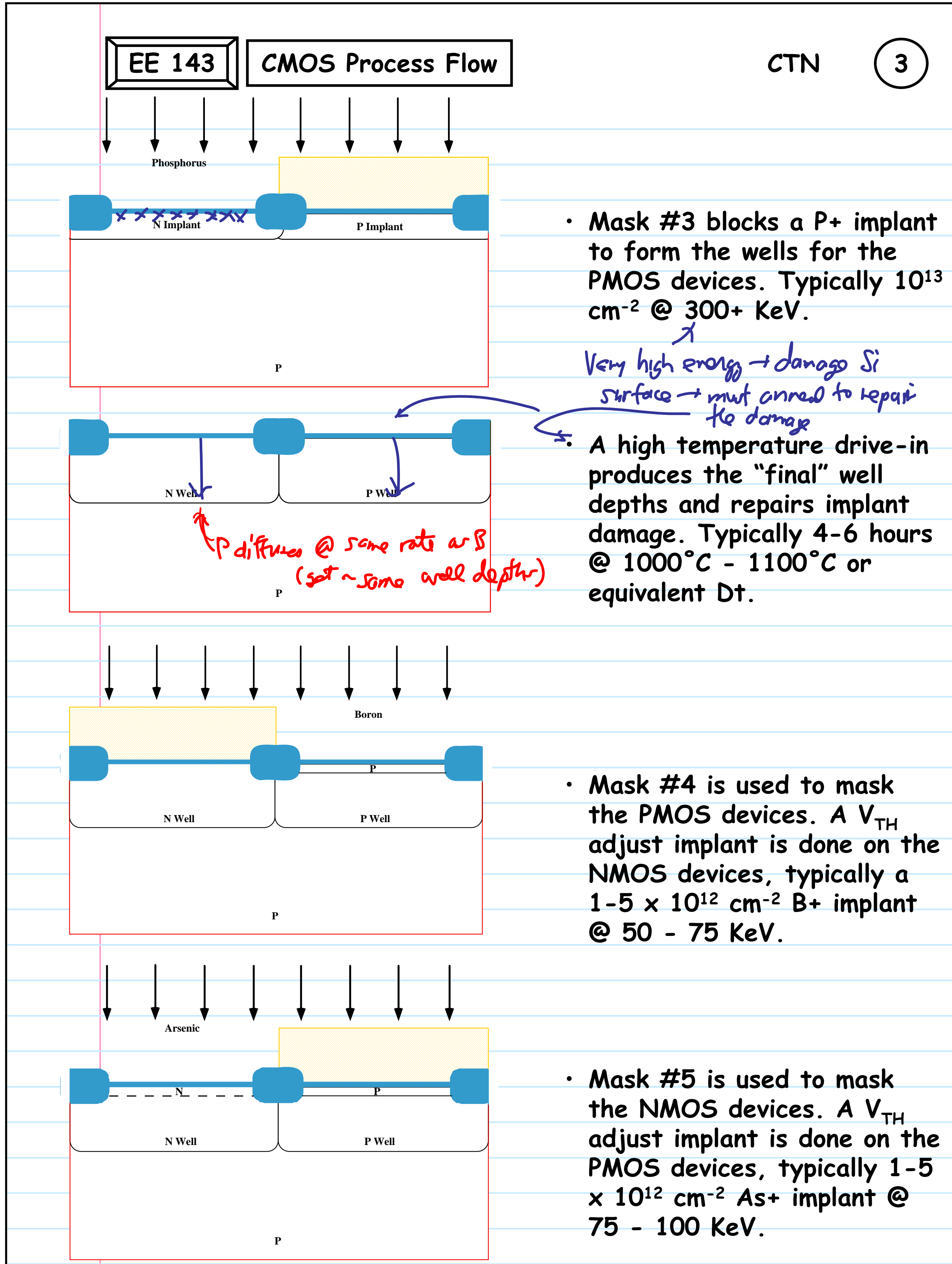
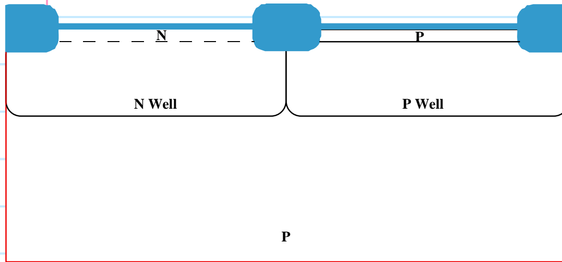


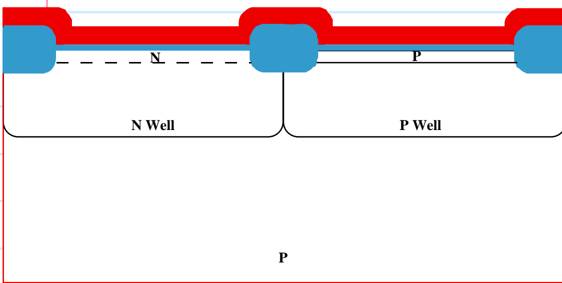
- Mask #3 blocks a P+ implant to form the wells for the PMOS devices. Typically 10^{13} cm^{-2} @ 300+ KeV.
- A high temperature drive-in produces the "final" well depths and repairs implant damage. Typically 4-6 hours @ 1000°C - 1100°C or equivalent Dt.
- Mask #4 is used to mask the PMOS devices. A V_{TH} adjust implant is done on the NMOS devices, typically a 1-5 x 10^{12} cm^{-2} B+ implant @ 50 - 75 KeV.
- Mask #5 is used to mask the NMOS devices. A V_{TH} adjust implant is done on the PMOS devices, typically 1-5 x 10^{12} cm^{-2} As+ implant @ 75 - 100 KeV.

Copyright @ 2010 Regents of the University of California at Berkeley

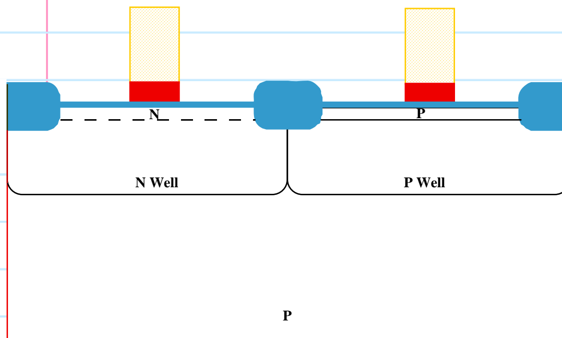
EE 143 | CMOS Process Flow | CTN | 4



- The thin oxide over the active regions is stripped and a new gate oxide grown, typically 3 - 5 nm, which could be grown in 0.5 - 1 hrs @ 800 °C in O2.
  → This is done because the previous oxide is damaged after numerous implants

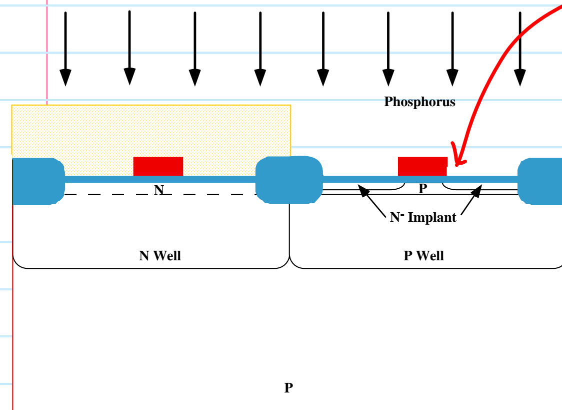


- Polysilicon is deposited by LPCVD (≈ 0.5 μm).



- Mask #6 is used to protect the MOS gates. The poly is plasma etched using an anisotropic etch.

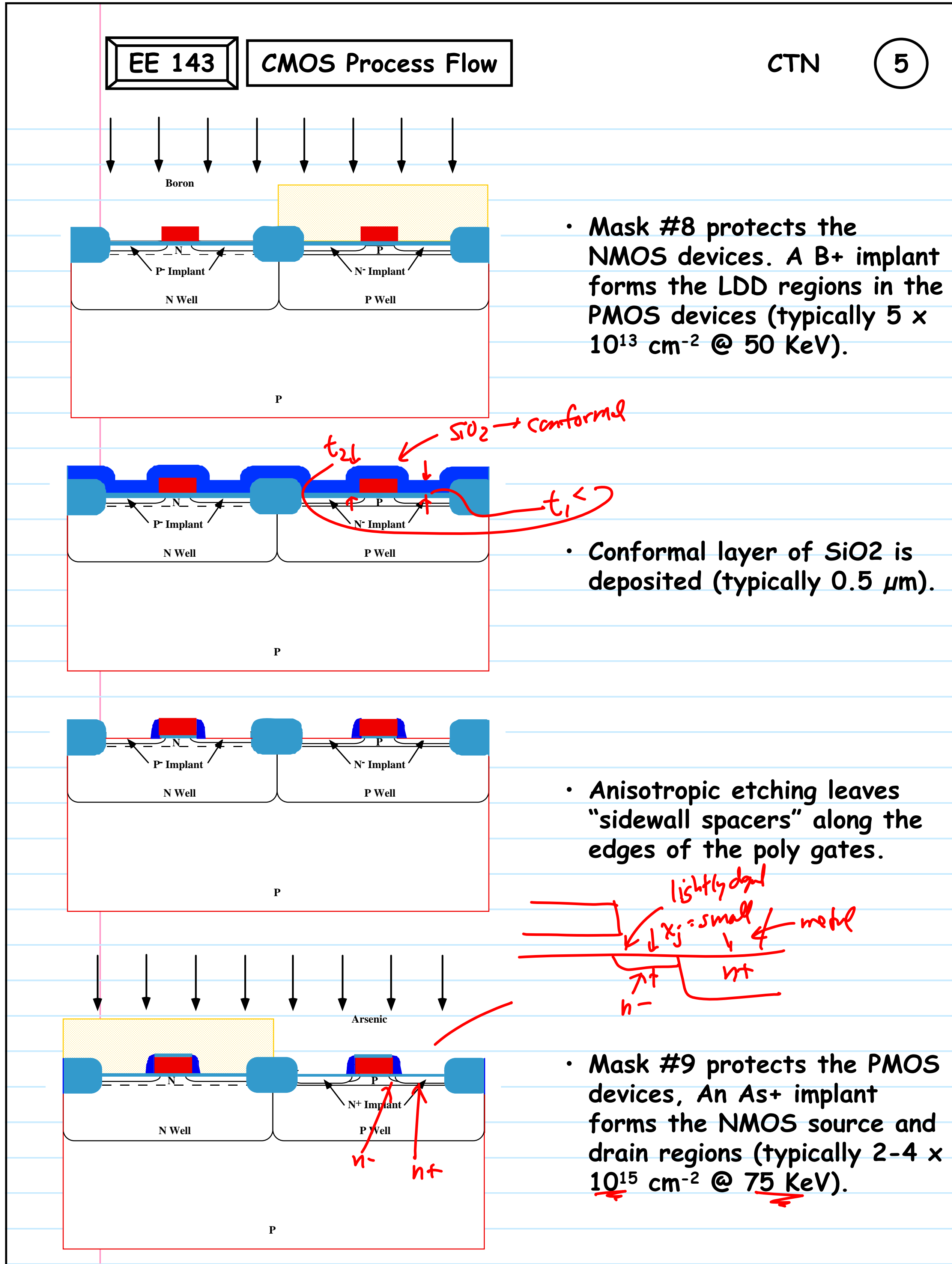
Want the S/D regions near the gate to be:
① shallow (i.e., x_j = small) to minimize punchthrough
② lightly-doped to minimize E-fields)



- Mask #7 protects the PMOS devices. A ~~P+~~ phosphorous implant forms the LDD regions in the NMOS devices (typically 5 x 10^{13} cm^{-2} @ 50 KeV). ← low

{ enhance device reliability

Copyright @ 2010 Regents of the University of California at Berkeley



Copyright @ 2010 Regents of the University of California at Berkeley

EE 143 CMOS Process Flow

CTN 6

Boron

P+ Implant N+ Implant N Well P Well P

does the PMOS gate p+ type

- Mask #10 protects the NMOS devices, A B+ implant forms the PMOS source and drain regions (typically 1-3 x 10^{15} cm^{-2} @ 50 KeV).

P+ N P+ N+ P N+ N Well P Well P

- A final high temperature anneal drives-in the junctions and repairs implant damage (typically 30 min @ 900°C or 1 min RTA @ 1000°C).

rapid thermal anneal

oxide sidewall spacer

no oxide

P+ N P+ N+ P N+ N Well P Well P

- An unmasked oxide etch allows contacts to Si and poly regions.

Conventional Anneal vs. RTA

the black → should be orange = Ti

P+ N P+ N+ P N+ N Well P Well P

- Ti is deposited by sputtering (typically 100 nm).

Copyright @ 2010 Regents of the University of California at Berkeley

EE 143 | CMOS Process Flow | CTN | 7

orange = TiN
black = silicide
poly-side
reduce the sheet resistance of the poly-si gate & diffused junctions

- The Ti is reacted in an N_2 ambient, forming $TiSi_2$ and TiN (typically 1 min @ 600 - 700°C).

Allows interconnection between n+ & p+ polysi gates

- Mask #11 is used to etch the TiN, forming local interconnects.

soln! TiN
pmos gate
pn-junction
nmos gate

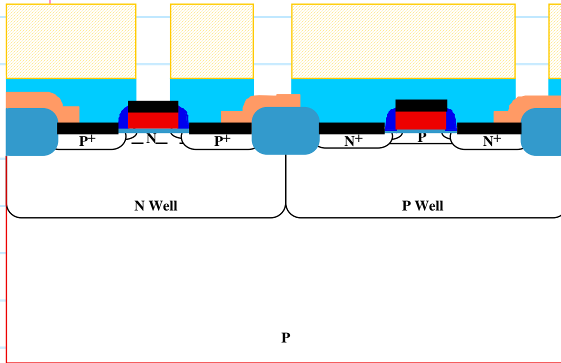
- A conformal layer of SiO_2 is deposited by LPCVD (typically 1 μm).

thick!

- CMP is used to planarize the wafer surface.

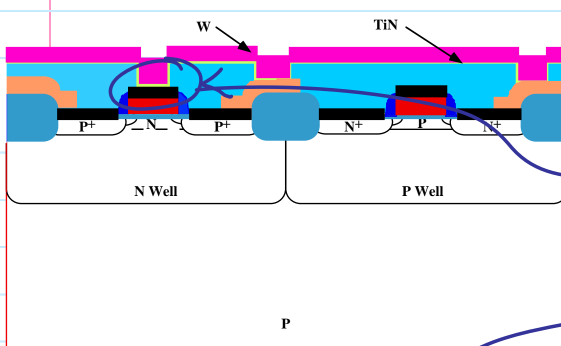
Copyright @ 2010 Regents of the University of California at Berkeley

EE 143 | CMOS Process Flow | CTN | 8



- Mask #12 is used to define the contact holes. The SiO_2 is etched.

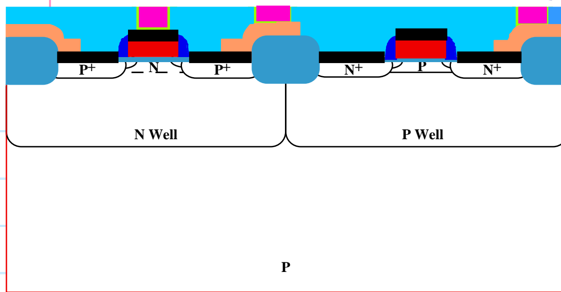
→ Why? ① Helps w/ W metal adhesion. ② Serves as barrier against pitting or spiking of metal into silicide



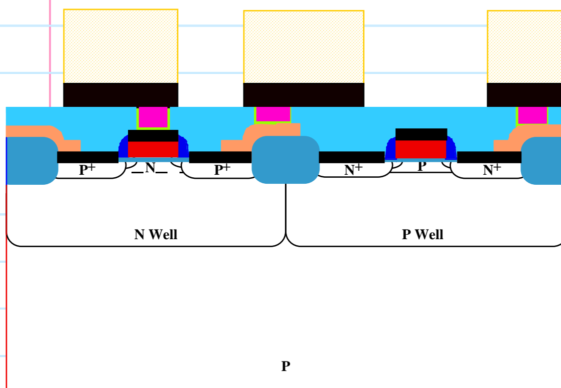
← the green is TiN.

- A thin TiN barrier layer is deposited by sputtering (typically a few tens of nm), followed by W CVD deposition.

← CMP metal (W), stop on oxide



- CMP is used to planarize the wafer surface, completing the damascene process.

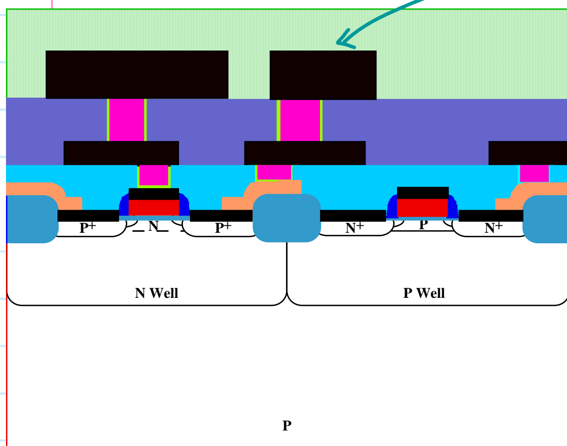


- Al is deposited on the wafer by sputtering. Mask #13 is used to pattern the Al and plasma etching is used to etch it.

Copyright @ 2010 Regents of the University of California at Berkeley

EE 143 | CMOS Process Flow | CTN | 9

Electroplated Cu for higher levels!



- Intermetal dielectric and second level metal are deposited and defined in the same way as level #1. Mask #14 is used to define contact vias and Mask #15 is used to define metal 2. A final passivation layer of Si_3N_4 is deposited by PECVD and patterned with Mask #16.

There will be more levels of metal → Cu. Use a dual-damascene process.

Differences from older long-channel CMOS:

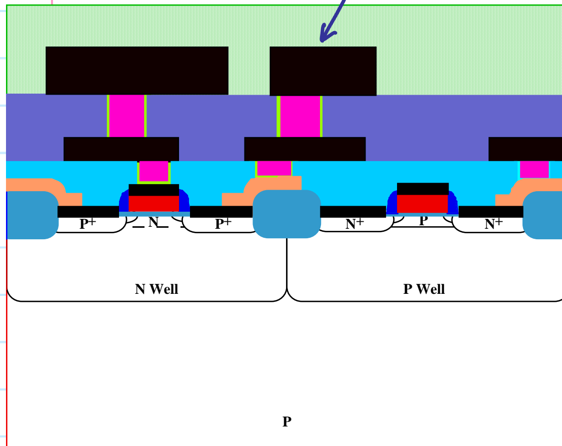
1. Very thin lightly doped drain (LDD) and source junctions; very small x_j
2. Very thin gate oxide
3. Much higher channel & substrate doping
   ↳ So why these differences? Answer: punchthrough!

Go to the regular lecture written format to discuss punchthrough.

Copyright @ 2010 Regents of the University of California at Berkeley

EE 143 | CMOS Process Flow | CTN | 10

Electroplated Cu for higher levels of metal.



- Intermetal dielectric and second level metal are deposited and defined in the same way as level #1. Mask #14 is used to define contact vias and Mask #15 is used to define metal 2. A final passivation layer of Si_3N_4 is deposited by PECVD and patterned with Mask #16.

There will be more levels of metal → Cu.
Use a dual-damascene process.

Differences from older long-channel CMOS:

1. Very thin lightly doped drain (LDD) and source junctions; very small x_j
2. Very thin gate oxide
3. Much higher channel & substrate doping
   - So why these differences? Answer: punchthrough!

Remarks:

- More advanced technologies also require a thinner gate oxide, for reasons explained above
  - SiO_2 can no longer satisfy, since the gate oxide thicknesses below 1 nm have leakage issues
  - Newest technologies use ALD high-k oxides (e.g., HfO_2)
- More advanced technologies must reduce as much as possible the interconnect capacitance ⇨ low-k inter-metal dielectrics
- The use of LOCOS in this process limits it to 0.18-0.25 μm channel lengths
  - Any smaller and the LOCOS encroachment it too much
  - Plus, the topography makes lithography for smaller channel lengths more difficult

Copyright @ 2010 Regents of the University of California at Berkeley