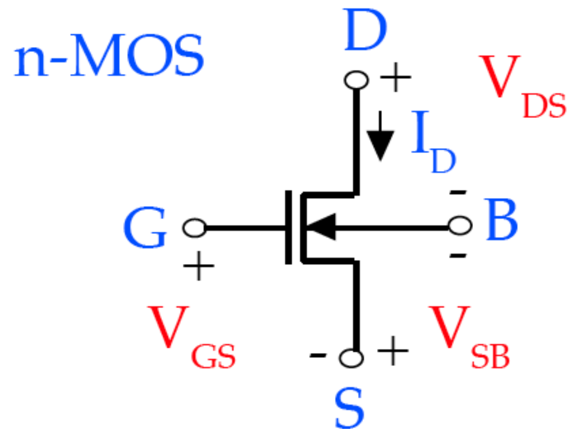
ESE370: Circuit-Level Modeling, Design, and Optimization for Digital Systems

Lec 10: September 27, 2021

MOS Transistor Variation

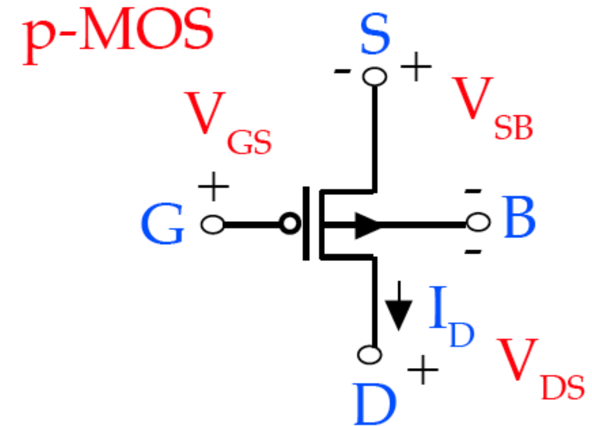
Penn

Review: nMOS IV Model



$$I_D = \begin{cases} I_S\left(\frac{W}{L}\right)e^{\left(\frac{V_{GS}-V_{Tn}}{nkT/q}\right)} & V_{GS} \le V_{Tn} & \text{Subthreshold} \\ \frac{\mu_n \cdot C_{ox}}{2}\left(\frac{W}{L}\right)\left(2(V_{GS}-V_{Tn})V_{DS}-V_{DS}^2\right) & V_{GS} > V_{Tn}, V_{DS} < V_{GS}-V_{Tn} & \text{Linear} \\ \frac{\mu_n \cdot C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS}-V_{Tn})^2 & V_{GS} > V_{Tn}, V_{DS} \ge V_{GS}-V_{Tn} & \text{Saturation} \\ \approx v_{sat}C_{ox}W\left(V_{GS}-V_{Tn}-\frac{V_{dsat}}{2}\right) & E_y > E_{cn} & \text{Velocity Saturation} \end{cases}$$

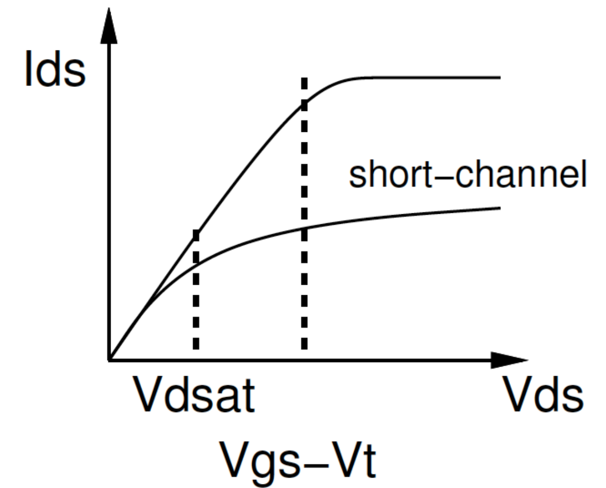
Review: pMOS IV Model

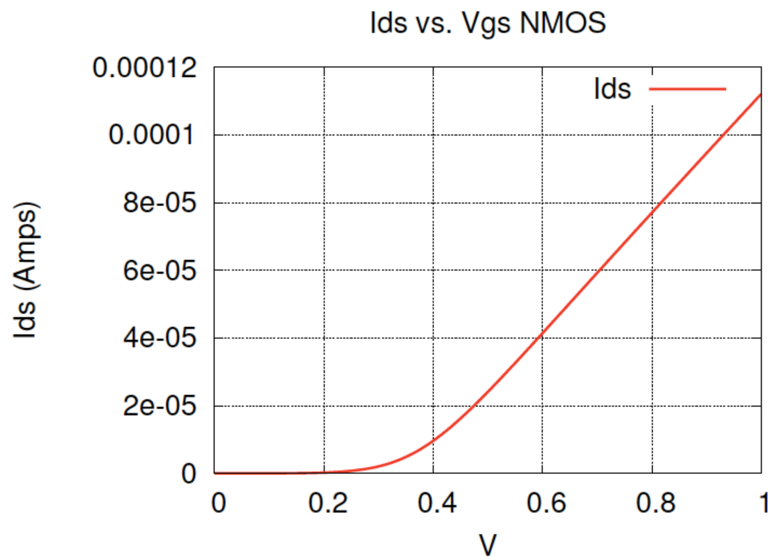


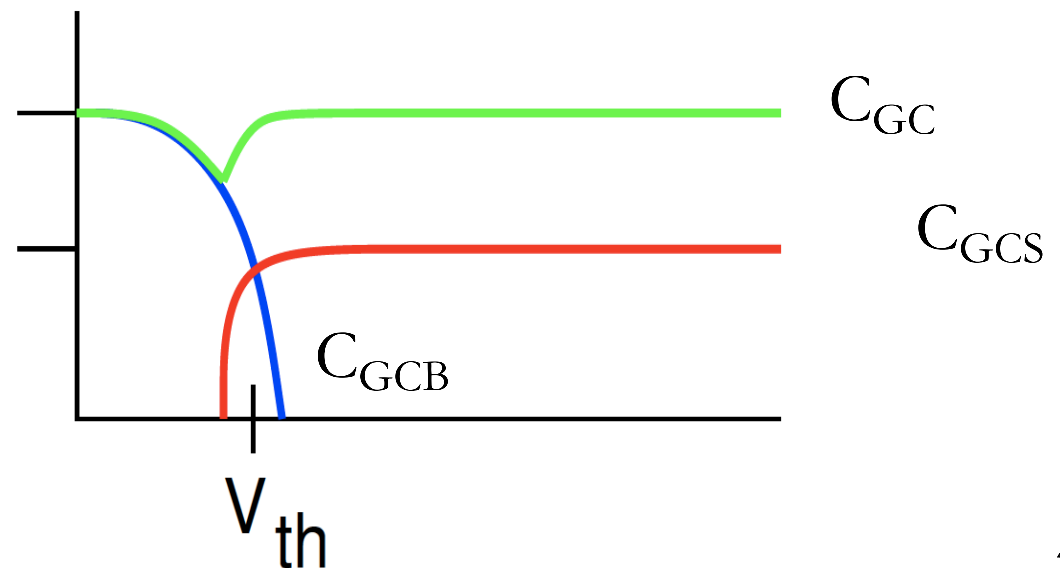
$$I_D = \begin{cases} I_S\left(\frac{W}{L}\right)e^{\left(\frac{V_{GS}-V_{Tp}}{nkT/q}\right)} & V_{GS} \geq V_{Tp} & \text{Subthreshold} \\ \frac{\mu_p \cdot C_{ox}}{2}\left(\frac{W}{L}\right)\left(2(V_{GS}-V_{Tp})V_{DS}-V_{DS}^2\right) & V_{GS} < V_{Tp}, V_{DS} > V_{GS}-V_{Tp} & \text{Linear} \\ \frac{\mu_p \cdot C_{ox}}{2}\left(\frac{W}{L}\right)\left(V_{GS}-V_{Tp}\right)^2 & V_{GS} < V_{Tp}, V_{DS} \leq V_{GS}-V_{Tp} & \text{Saturation} \\ \approx v_{sat}C_{ox}W\left(V_{GS}-V_{Tp}-\frac{V_{dsat}}{2}\right) & E_y > E_{cp} & \text{Velocity Saturation} \end{cases}$$

Thus far…

- Understand how to model transistor behavior
- Given that we know its parameters
  - V_{dd}, V_{th}, C_{OX}, W, L, μ …







But…

- We don't know its parameters (perfectly)
  - Fabrication parameters have nominal values and error range
  - Impact on I_{ds}?

- Identically drawn devices differ because of fabrication techniques (e.g. process mismatch)
- Parameters change with environment (*e.g.* Temperature)
- Parameters change with time (aging)

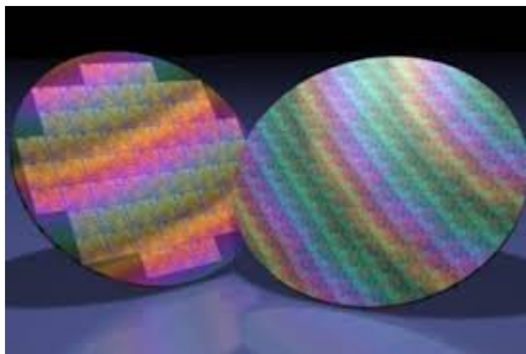
Today

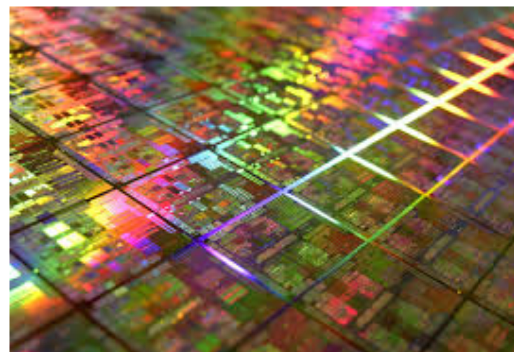
- Sources of Variation
  - Fabrication
  - Operation
  - Aging
- Designing to Account for Variation
  - Margin
  - Corners
  - Binning

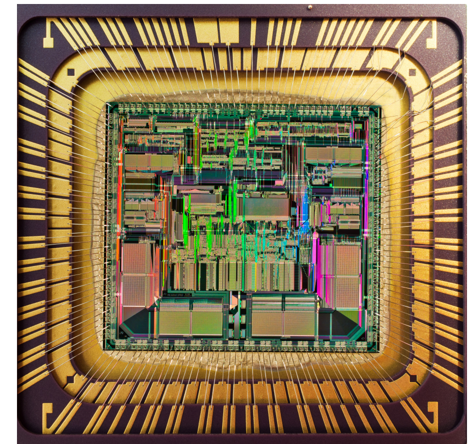
Fabrication

Variation Types

- Many reasons why variation occurs and shows up in different ways
- Scales of variation
  - Wafer-to-wafer, die-to-die, transistor-to-transistor
- Correlations of variation
  - Systematic, spatial, random (uncorrelated)







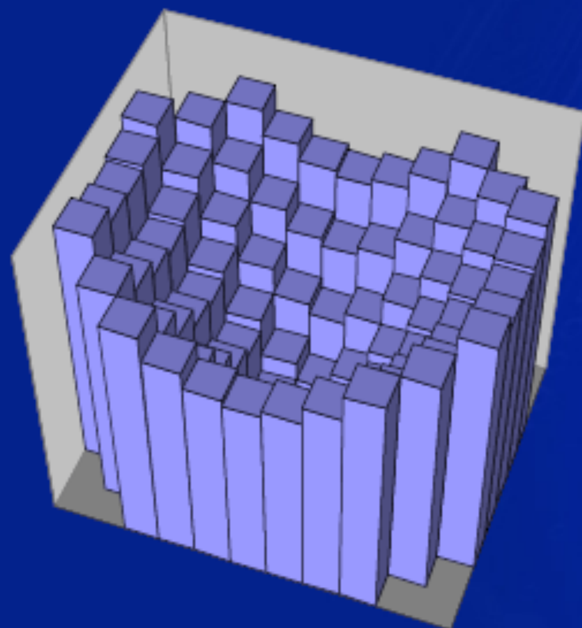
Scale of Variations

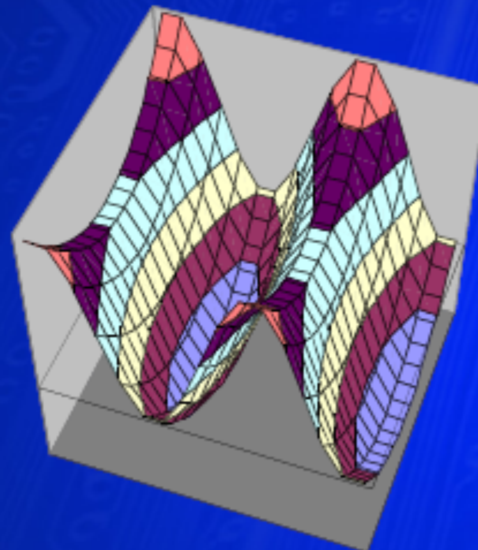
Die-to-Die (D2D) Variations

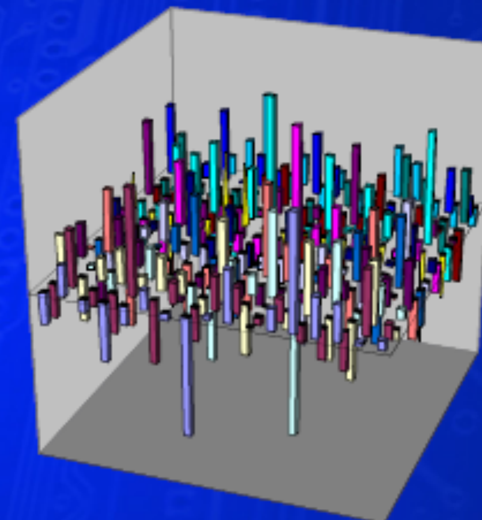
Within-Die (WID) Variations

Systematic

(Uncorrelated) Random







Wafer Scale

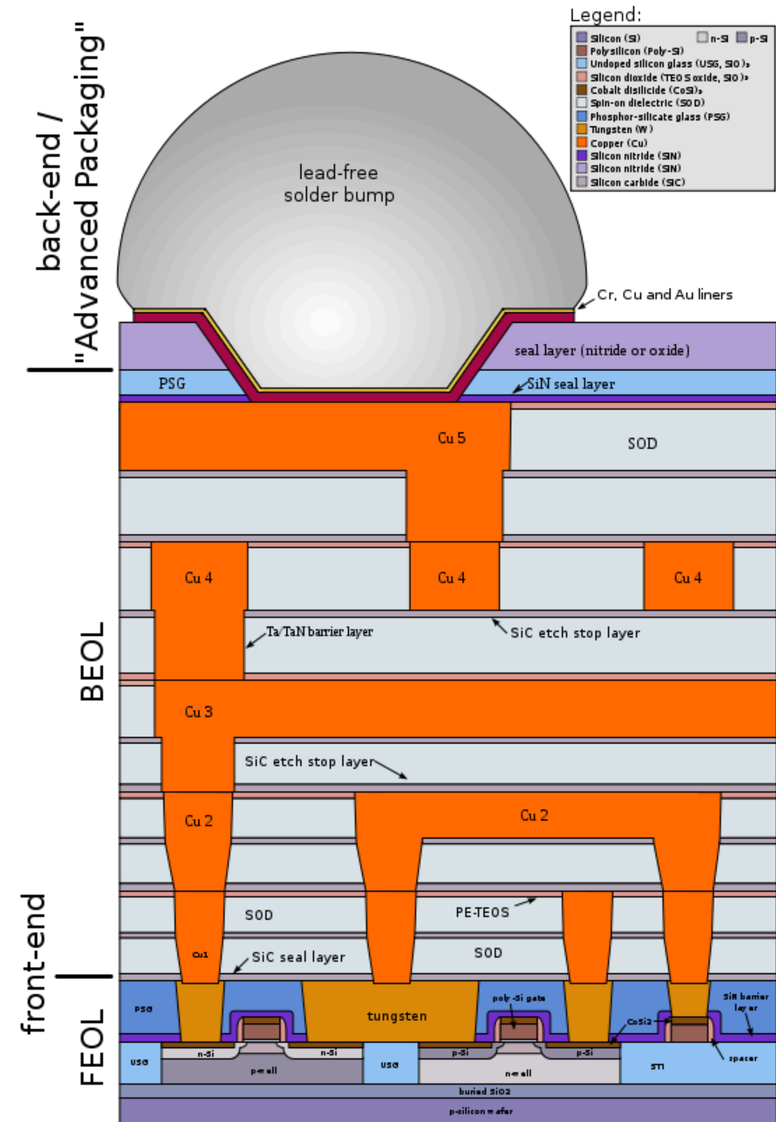
Die Scale

Feature Scale

Source: Noel Menezes, Intel ISPD2007

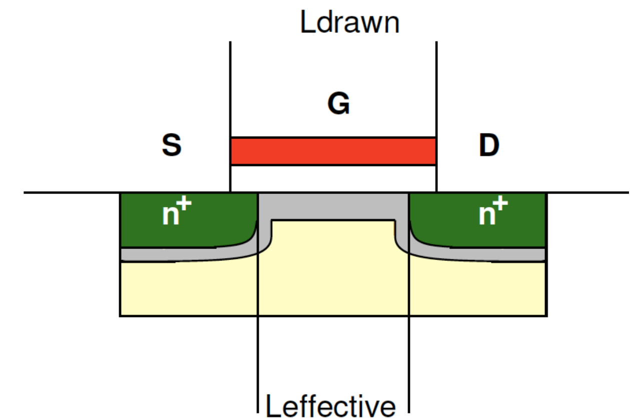
Basic Fabrication: Two Steps

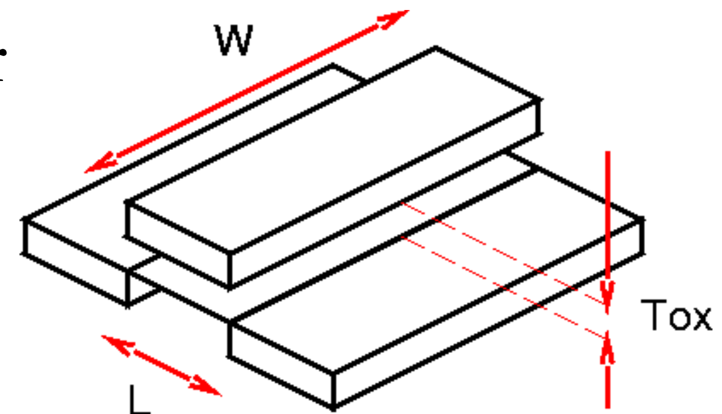
- (1) Transfer an image of the design to the wafer
- (2) Using that image (mask) as a guide, create the desired layers on silicon
  - Diffusion (add dopants to the silicon)
  - Oxide (create an insulating layer)
  - Metal (create a wire layer)



Wafer Scale: Process Shift

- Oxide thickness
- Doping level
- Layer alignment
- Growth and Etch rates and times
  - Depend on chemical concentrations
    - How precisely can we control those?
- Vary machine-to-machine, day-to-day
- Impact all transistors on wafer

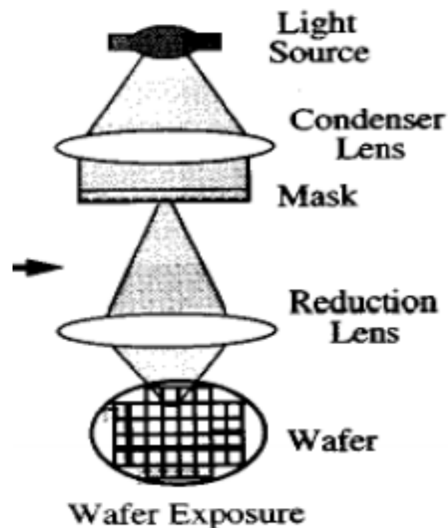


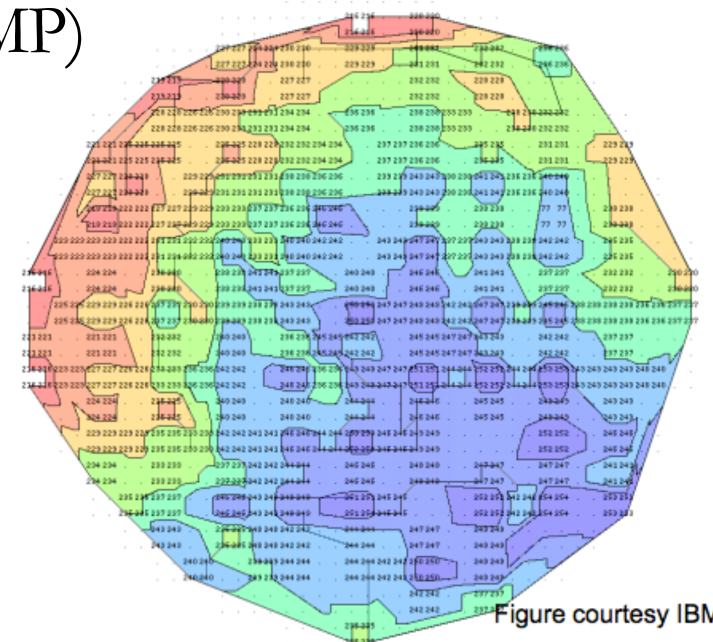


Systematic Spatial Variation

- Parameters change consistently across wafer or chip based on location

- Sources
  - Chemical-Mechanical Polishing (CMP)
    - Dishing
  - Lens distortion

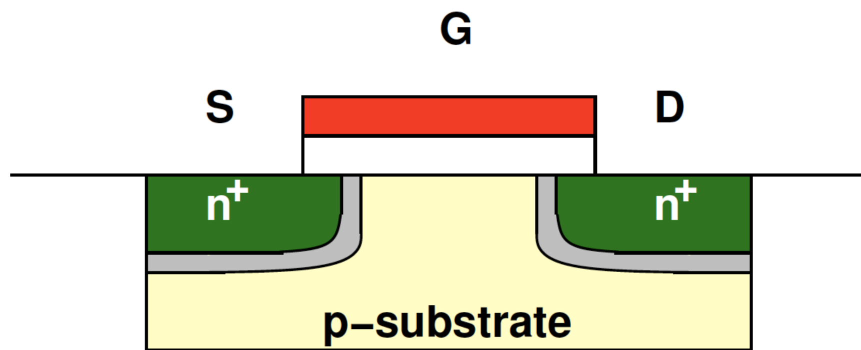


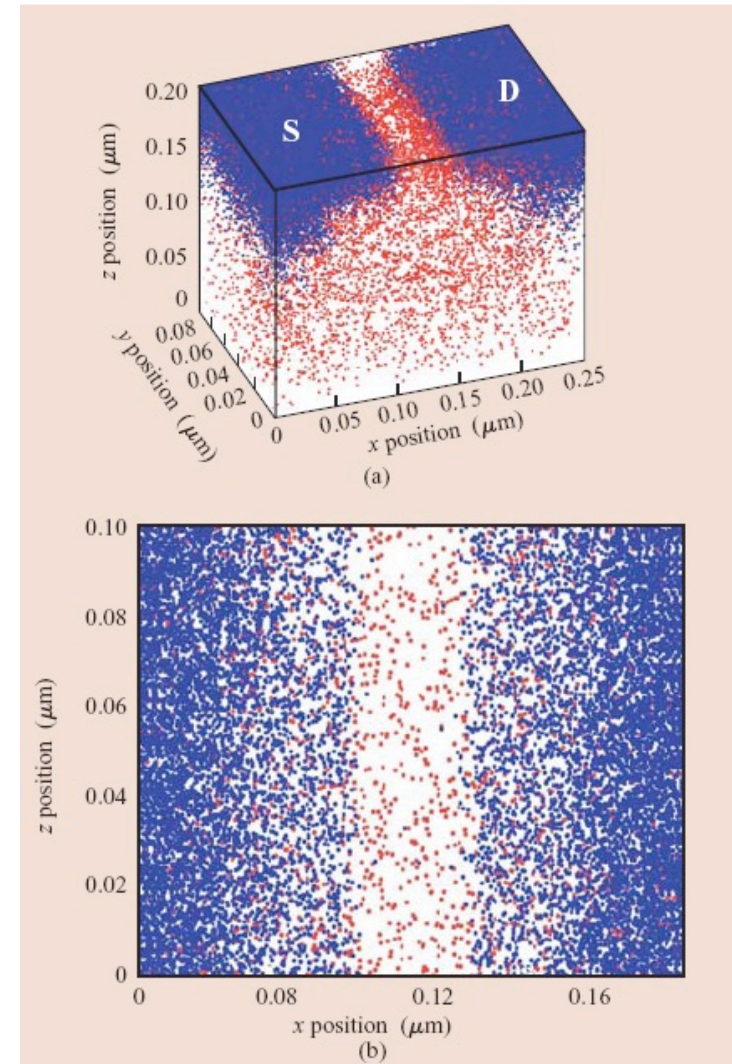


- Random dopant fluctuation
- Local oxide variation
- Line edge roughness
- Etch and growth rates

- Transistors differ from each other in random ways

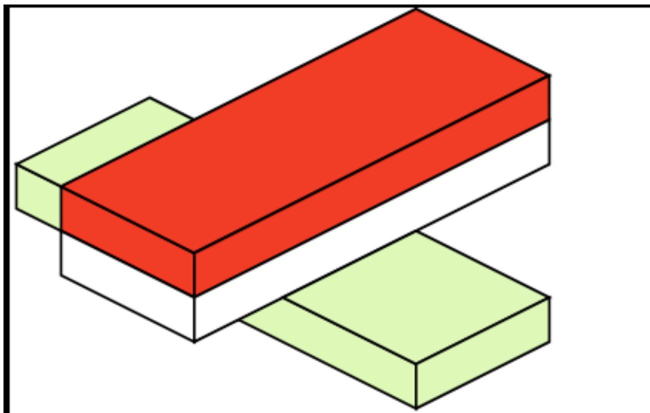
Statistical Dopant Placement

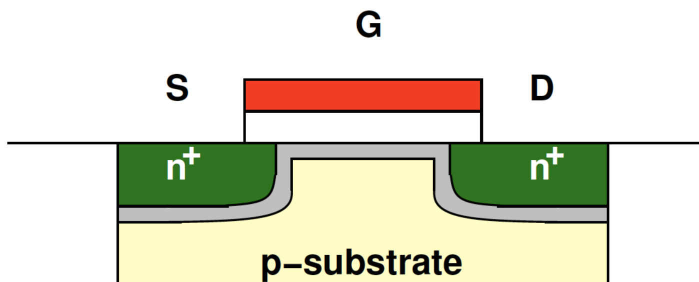




[Bernstein et al, IBM JRD 2006]

Oxide Thickness and Interface roughness





[Asenov et al. TRED 2002]

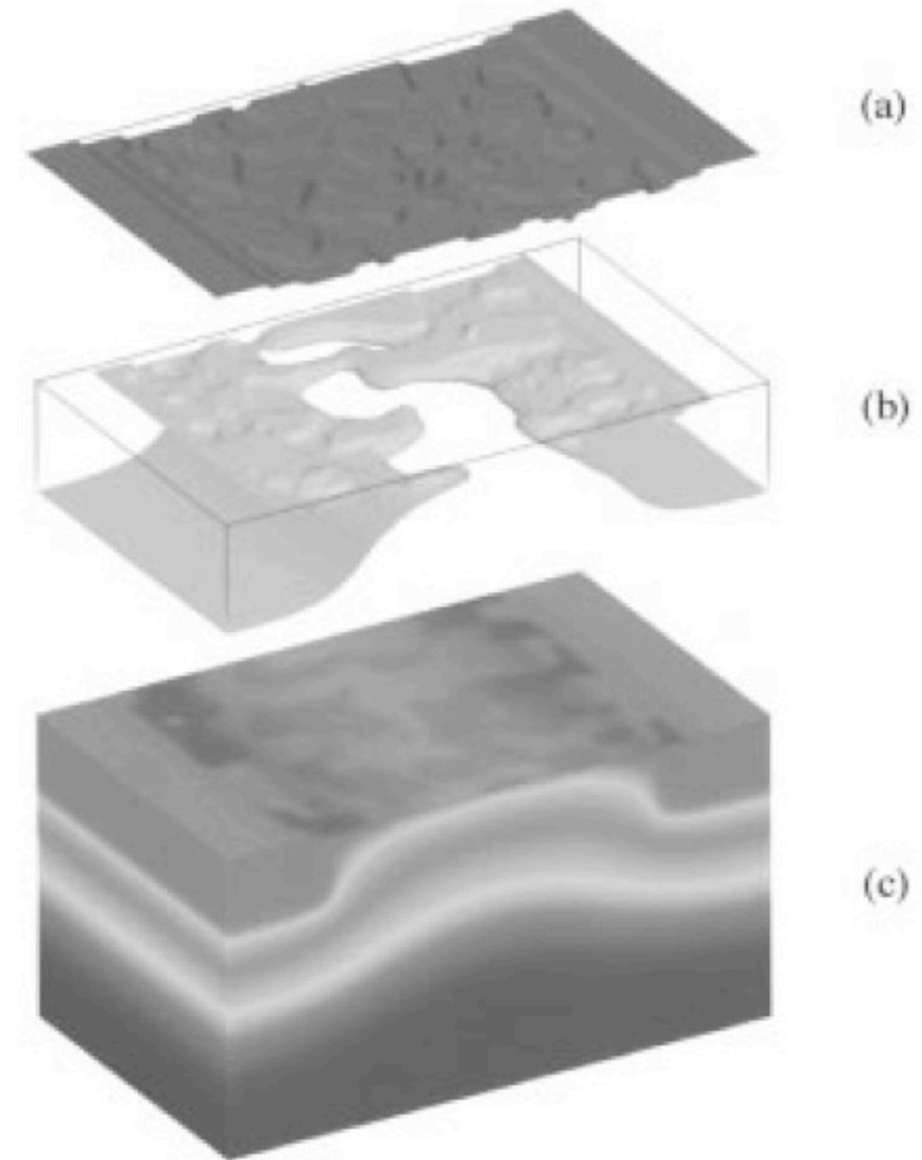
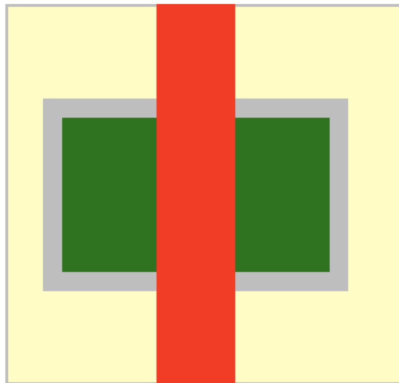
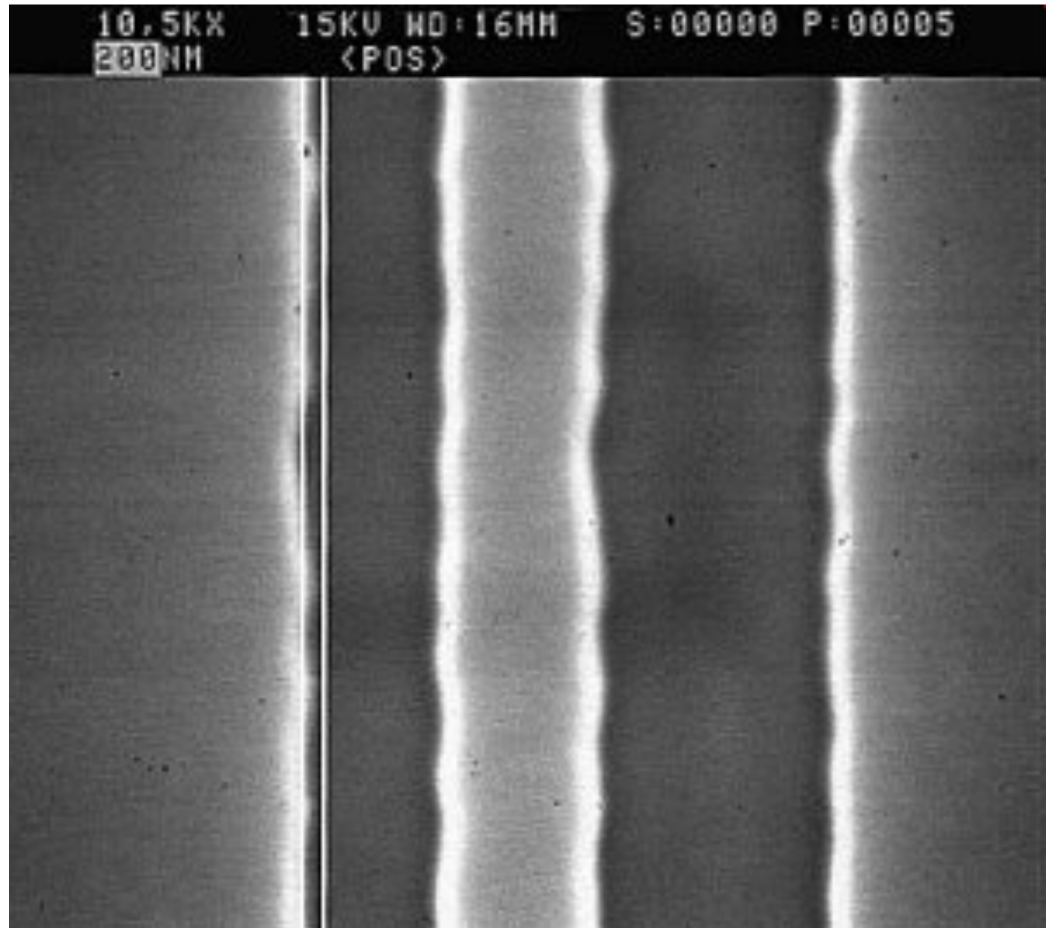


Fig. 1. (a) Typical profile of the random Si/SiO_2 interface in a $30 \times 30\ nm^2$ MOSFET, followed by (b) an equiconcentration contour obtained from DG simulations, and (c) the potential distribution.

Line Edge Roughness

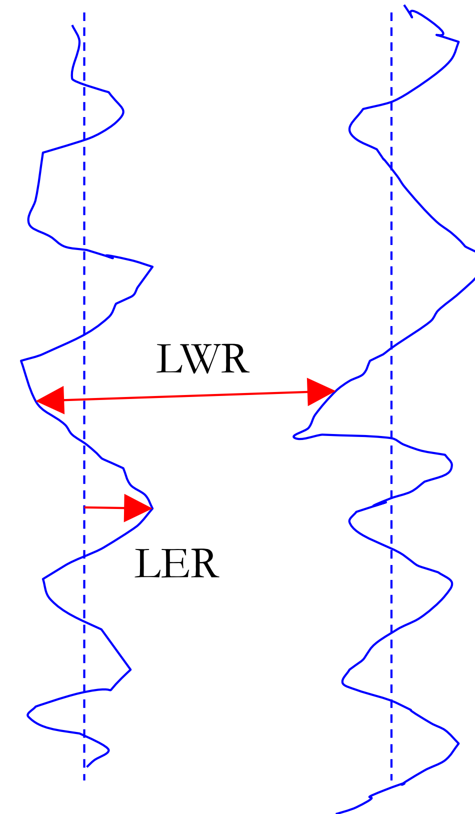




From:
http://www.microtechweb.com/2d/lw_pict.htm

Line Edge & Line-Width Roughness

- LER = lines of device features are not straight
- LWR = distance between lines is not uniform
- LER & LWR arise from the lithography and etching processes
- They are most pronounced in poly-gate patterning



Source: Kuhn et al.

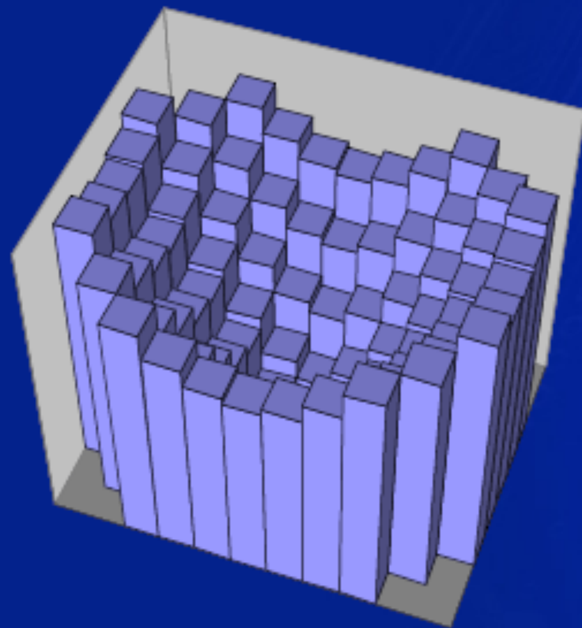
Scale of Variations

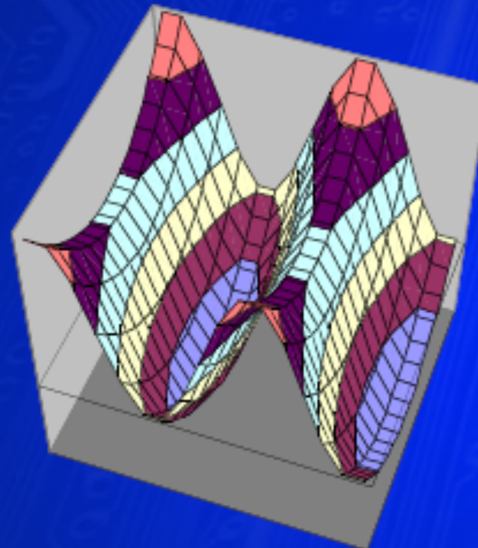
Die-to-Die (D2D) Variations

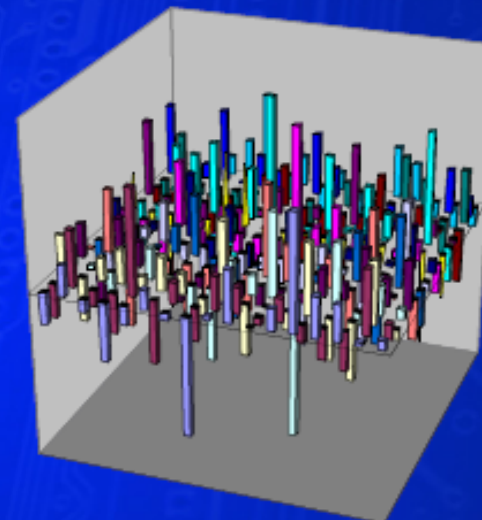
Within-Die (WID) Variations

Systematic

(Uncorrelated) Random







Wafer Scale **Die Scale** **Feature Scale**

Source: Noel Menezes, Intel ISPD2007

Impact on I_d (preclass 1)

- Changes parameters
  - W, L, t_{OX}, V_{th}, etc.
- Change transistor behavior
  - W increase?
  - L increase?
  - t_{OX} increase?
  - V_{th} increase?

$$I_{DS} \approx v_{sat} C_{OX} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right)$$

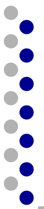
$$I_{DS} = \mu_n C_{OX} \left(\frac{W}{L} \right) \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

Example: V_{th}

- Many physical effects impact V_{th}
  - Doping, dimensions, roughness
- Behavior highly dependent on V_{th}

$$I_{DS} \approx \nu_{sat} C_{OX} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right)$$

$$I_{DS} = I_S' \left(\frac{W}{L} \right) e^{\left(\frac{V_{GS} - V_T}{nkT/q} \right)}$$



V_{th} Variability @ 65nm

[Bernstein et al, IBM JRD 2006]

Impact of V_{th} Variation?

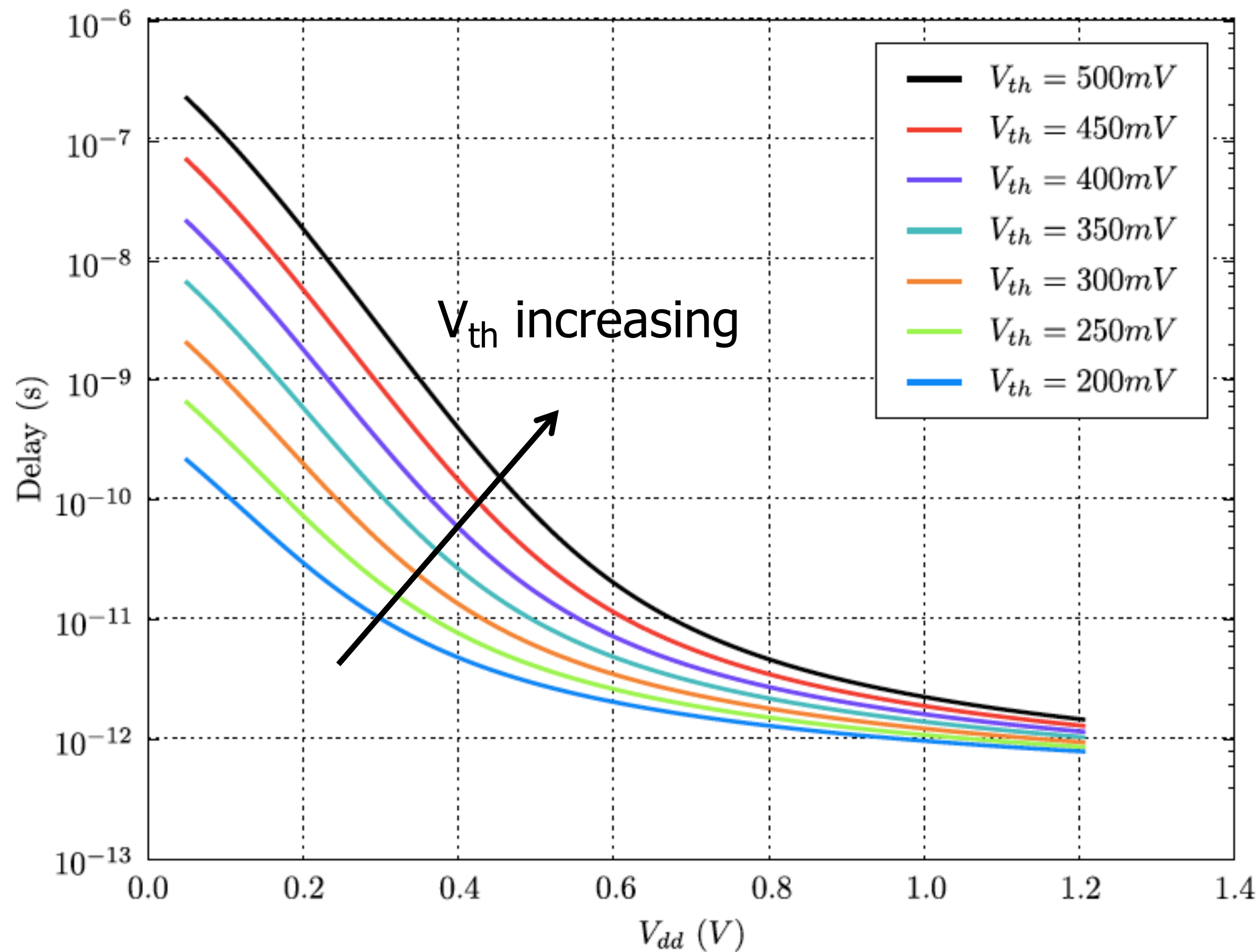
- Higher V_{th}?
  - Not drive as strongly
  - $I_{d,vsat} \propto (V_{gs} - V_{th})$
  - Performance?

Impact Performance

- Higher V_{th} → lower I_{ds} → Delay (R_{on} * C_{load})?

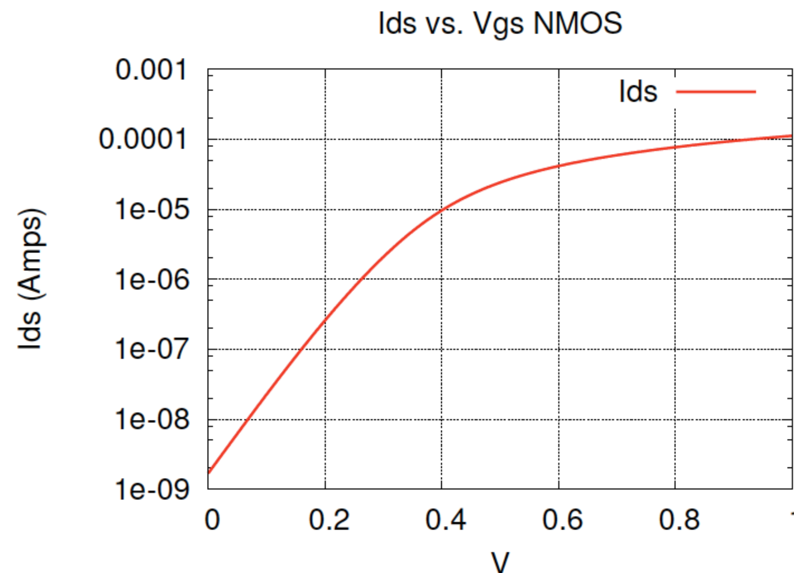
Impact Performance

- Higher V_{th} → lower I_{ds} → Delay (R_{on} * C_{load})?



Impact of V_{th} Variation?

- Lower V_{th}?
  - Not turn off as well → leaks more



$$I_{DS} = I_S'\left(\frac{W}{L}\right)e^{\left(\frac{V_{GS}-V_T}{nkT/q}\right)}$$

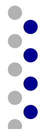
Operation

Temperature

Voltage

Temperature Changes

- Different ambient environments
  - January in Maine
  - July in Philly
  - Air conditioned machine room
- Self heat from activity of chip
- Quality of heat sink (attachment thereof)
  - E.g cooling fan



Thermal Profile for Processor

[Reda/IEEE Tr Emerging CAS v1n2 2011]

How does temperature impact on-current?

- High temperature
  - More free thermal energy
    - Easier to conduct
    - Lowers V_{th}
  - Increase rate of collision
    - Lower saturation velocity
    - Lower saturation voltage
    - Lower peak I_{ds} → slows down
- One reason don't want chips to run hot

Temperature and I_{ds}

Temperature Impact on Ids

V=1V-Ids*(1 Ohm)

1
0.99999
0.99998
0.99997
0.99996
0.99995
0.99994
0.99993

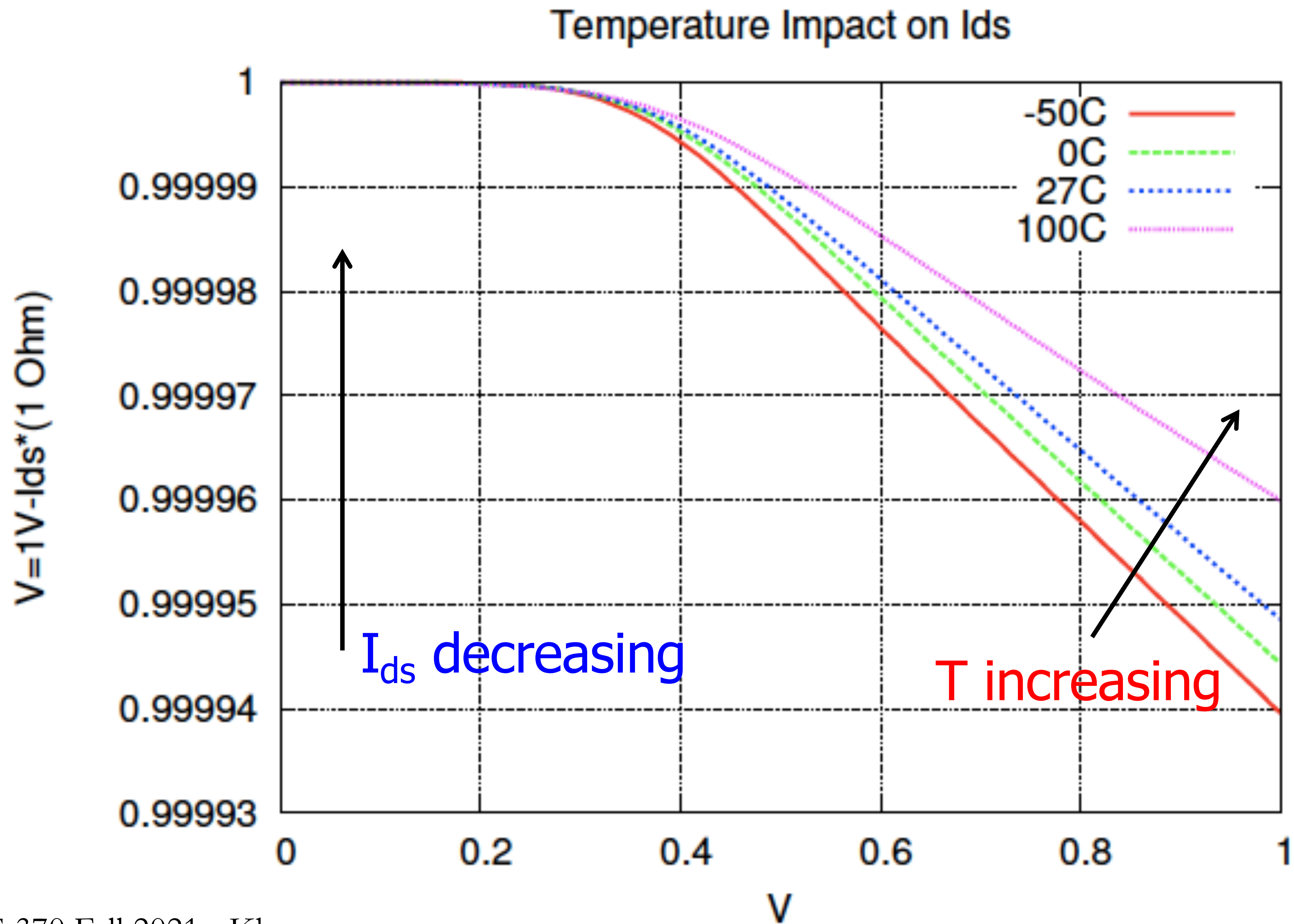
0 0.2 0.4 0.6 0.8 1

V

-50C
0C
27C
100C

I_{ds} decreasing

Temperature and I_{ds}



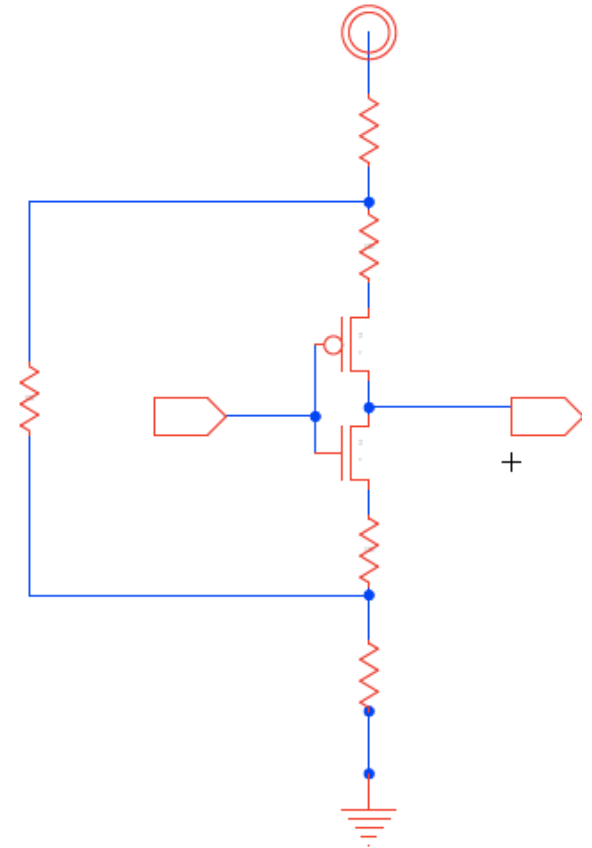
How does temp impact leakage current?

- High temperature lowers V_{th}

$$I_{DS} = I_S' \left(\frac{W}{L} \right) e^{\left(\frac{V_{GS} - V_T}{nkT/q} \right)}$$

Voltage

- Power supply isn't perfect
- Differs from design to design
  - Board to board?
  - How precise is regulator?
- IR-drop in distribution
- Bounce with current spikes



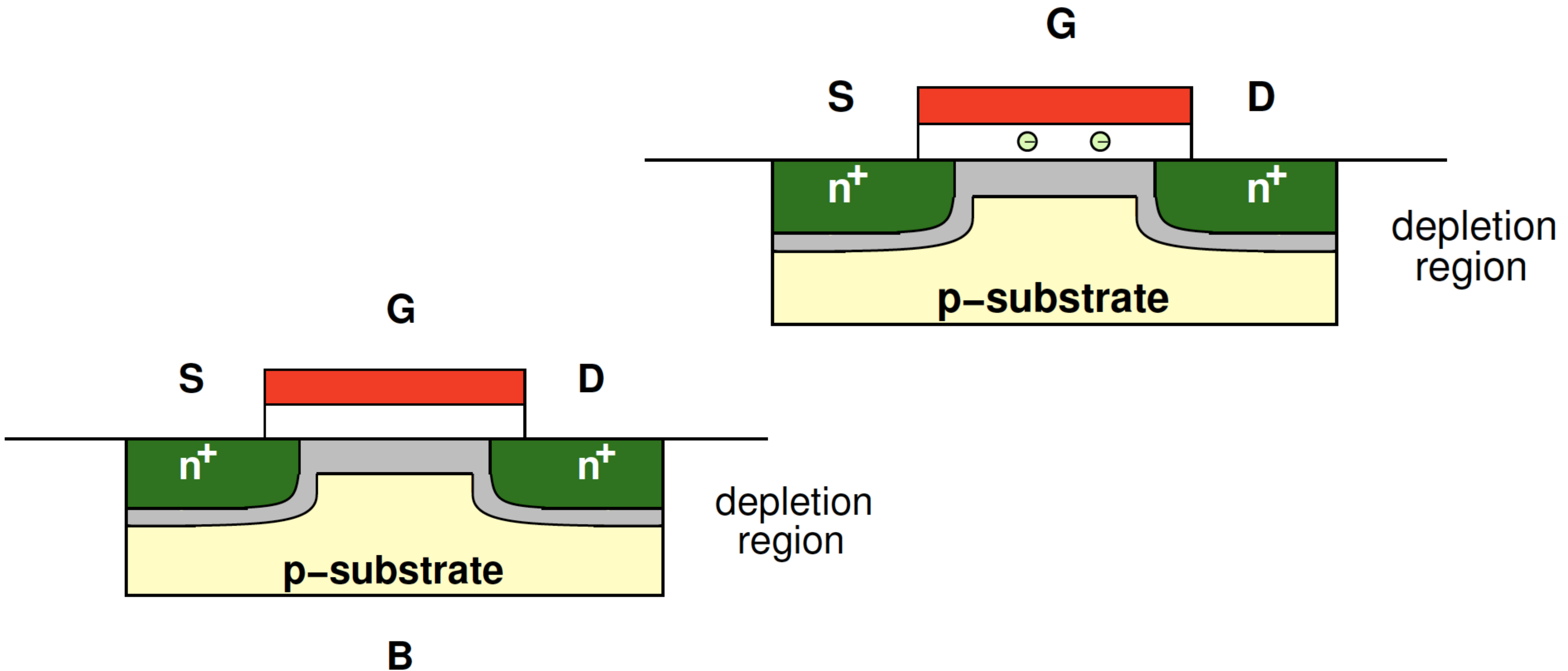
Aging

Hot Carrier Injection

Negative Bias Temperature Instability (NBTI)

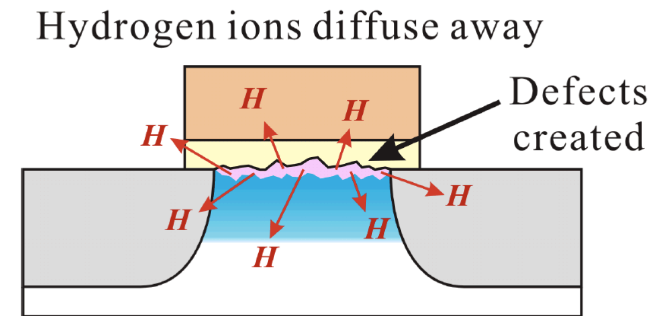
Hot Carrier Injection

- Trap electrons in oxide
  - increases V_{th}



NBTI

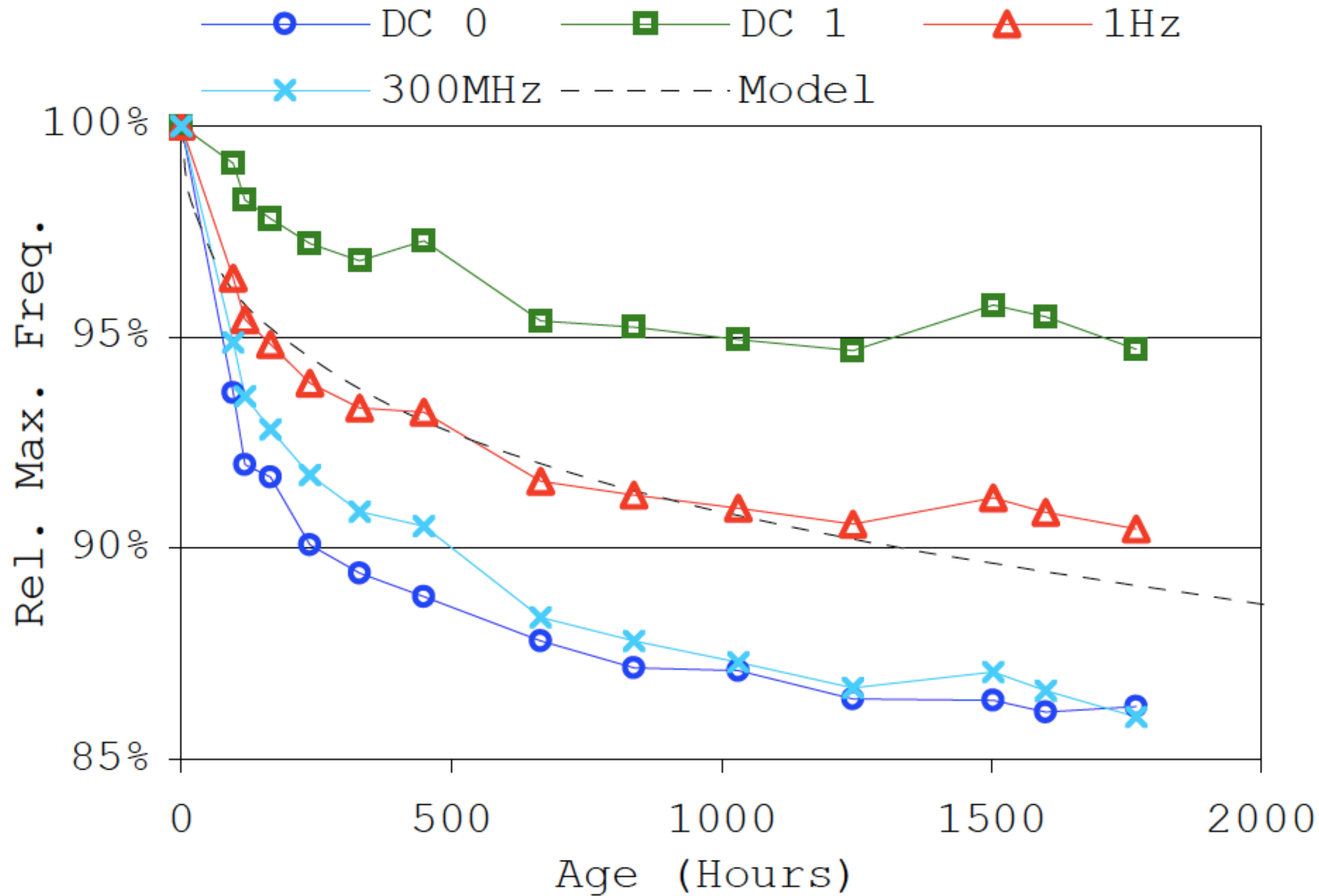
- Negative Bias Temperature Instability
  - Interface traps, Holes
- Long-term negative gate-source voltage
  - Affects PFET most
- Increase V_{th}
- Temperature dependent



[Stott, FPGA2010]

$$\Delta V_t(t) \propto \exp(-\beta V_G)\exp(-\frac{E_a}{\kappa T})t^n$$

Measured Accelerated Aging (Cyclone III, 65nm FPGA)

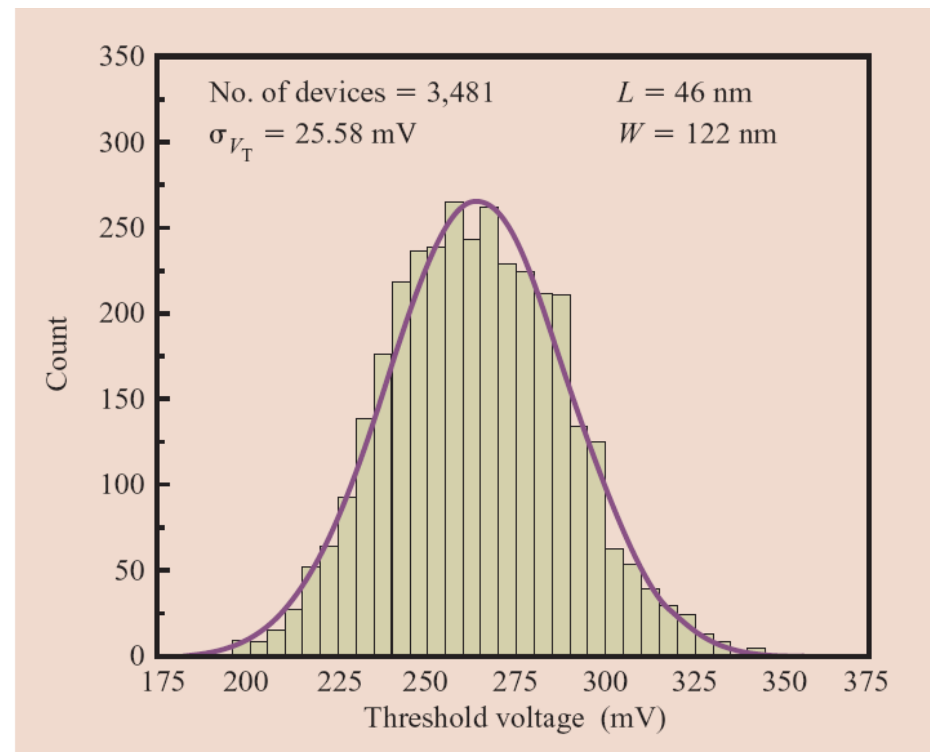


[Stott, FPGA2010]

Coping with Variation

Variation

- See a range of parameters
  - L: $L_{min} - L_{max}$
  - V_{th}: $V_{th,min} - V_{th,max}$



Impact of V_{th} Variation

- Higher V_{th}
  - Not drive as strongly
  - $I_{d,vsat} \propto (V_{gs} - V_{TH})$

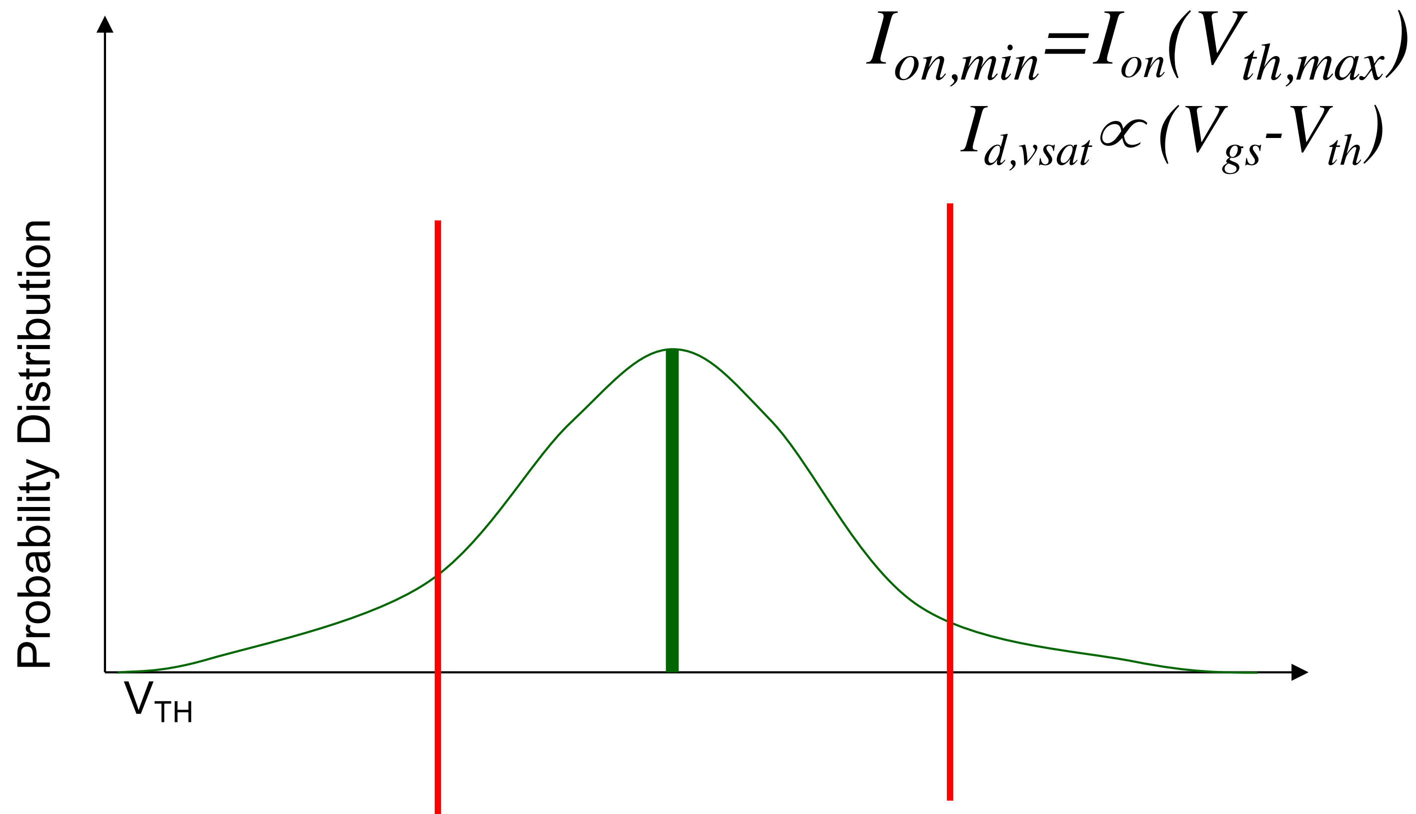
$$I_{DS} \approx \nu_{sat} C_{OX} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right)$$

- Lower V_{th}
  - Not turn off as well → leaks more

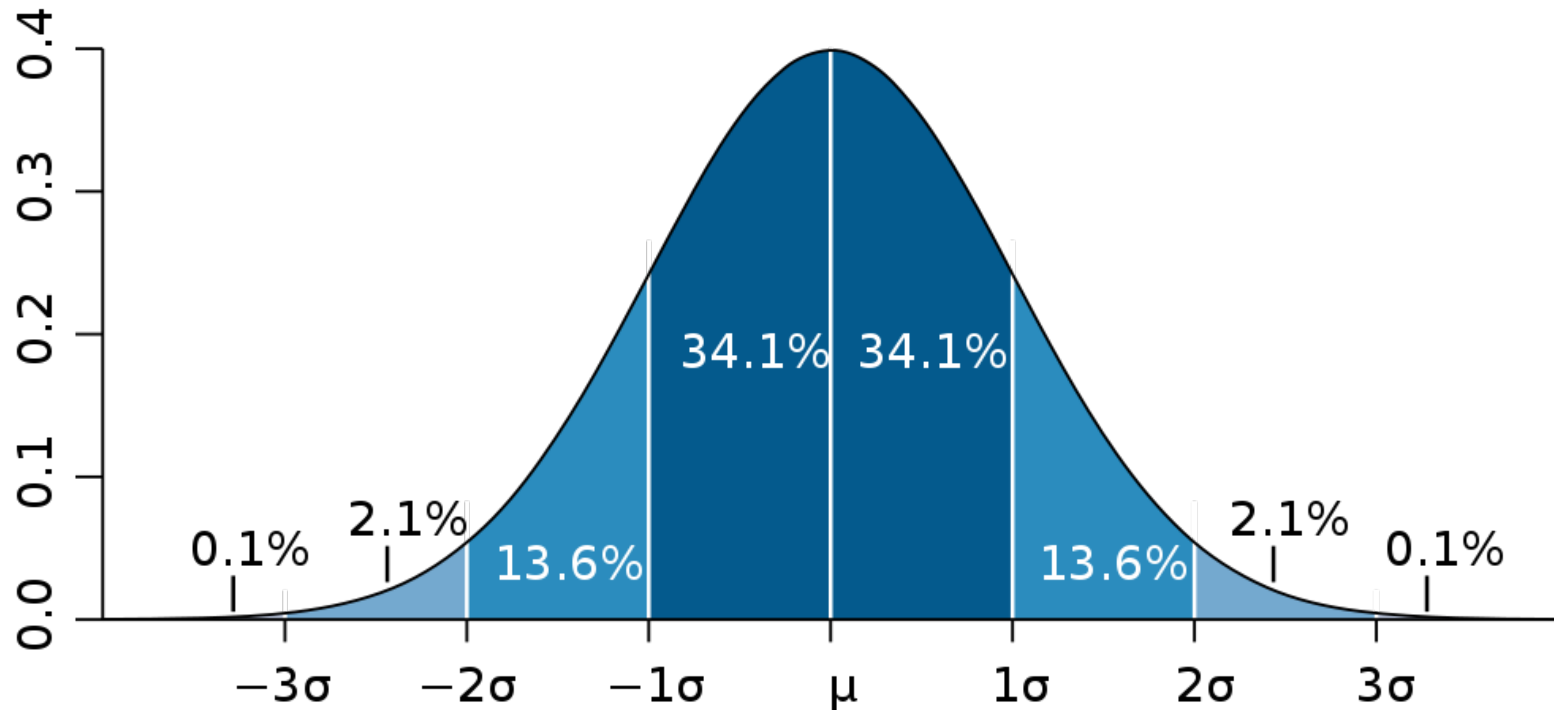
$$I_{DS} = I_S' \left(\frac{W}{L} \right) e^{\left(\frac{V_{GS} - V_T}{nkT/q} \right)}$$

Variation

- Margin for expected variation
- Must assume V_{th} can be any value in range
  - Speed → assume V_{th} slowest value



Gaussian Distribution



From: http://en.wikipedia.org/wiki/File:Standard_deviation_diagram.svg

Impact (preclass 2&3)

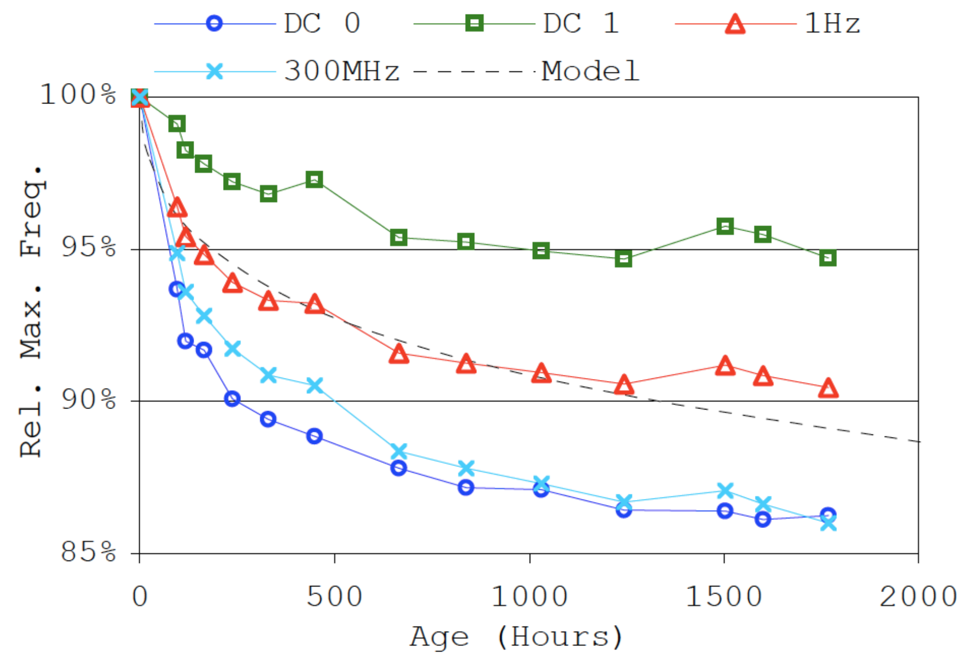
- Given
  - $V_{th,nom}$ = 250mV
  - Standard deviation: σ= 25mV
- Probability of 100 transistor circuit having all transistors with threshold in range 200mV<V_{th}<300mV
  - When each transistors has 96% prob of being in range?
  - When each has 99.8% probability?

Variation

- See a range of parameters
  - L: $L_{min} – L_{max}$
  - V_{th}: $V_{th,min} – V_{th,max}$
- Validate design at extremes
  - Work for both $V_{th,min}$ and $V_{th,max}$?
  - Design for worst-case scenario

Margining

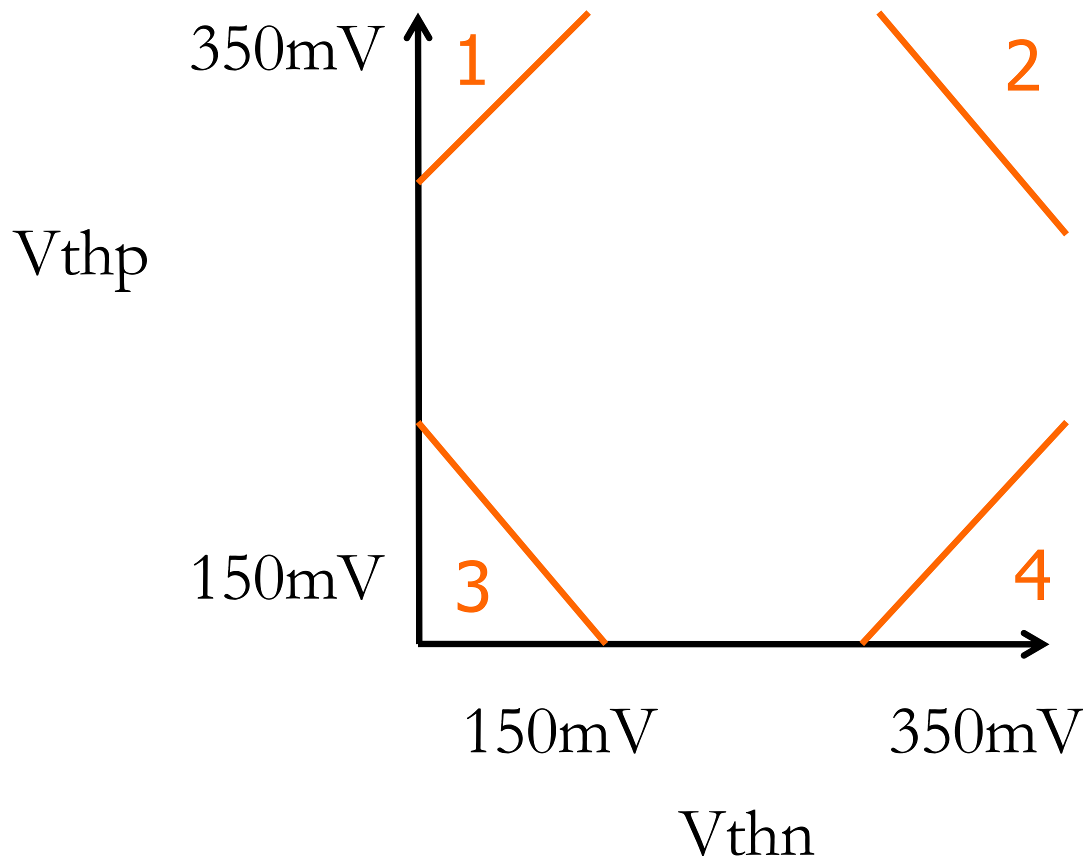
- Also margin for
  - Temperature
  - Voltage
  - Aging: end-of-life



Process Corners (preclass 4)

- Many effects independent
- Many parameters
- With N parameters,
  - Look only at extreme ends (low, high)
  - How many cases?
- Try to identify the {worst,best} set of parameters
  - Slow corner of design space, fast corner
- Use corners to bracket behavior

Simple Corner Example



What happens at various corners?

Process Corners

- Many effects independent
- Many parameters
- Try to identify the {worst,best} set of parameters
  - E.g. Lump together things that make slow
    - Vtn, Vtp, temperature, Voltage
    - Try to reduce number of unique corners
  - Slow corner of design space
- Use corners to bracket behavior

Worst-case Corner Model

- corners for analog applications
  - For modeling worst-case speed
    - Slow NMOS and slow PMOS(SS) corner
  - For modeling worst-case power
    - Fast NMOS and fast PMOS(FF) corner
- corners for digital applications
  - For modeling worst-case 1
    - Fast NMOS and slow PMOS(FS) corner
  - For modeling worst-case 0
    - Slow NMOS and fast PMOS(SF) corner

Worst-case Corner Model

- Advantages
  - Worst case corner models give designers the capability to simulate the pass/fail results of a typical design and are usually pessimistic.

- Disadvantages
  - The fixed-corner method is too wide
  - Some valid designs can not be accepted in worst-case corner model
  - The correlations between the device parameters are ignored

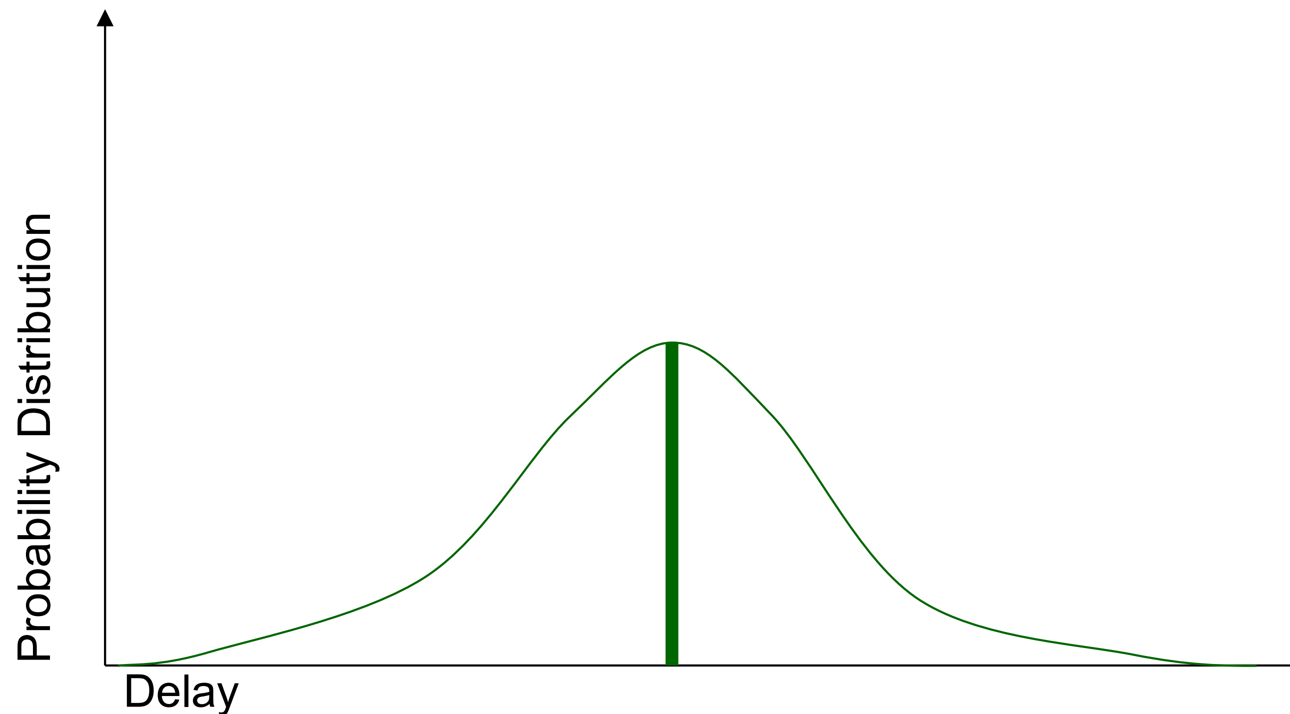
Statistical Corner Model

- For more realistic modeling for process variability than worst-case corner model.
  - Using data from different dies, wafers, and wafer lots collected over a long enough period of time to represents realistic process variability of the target technology

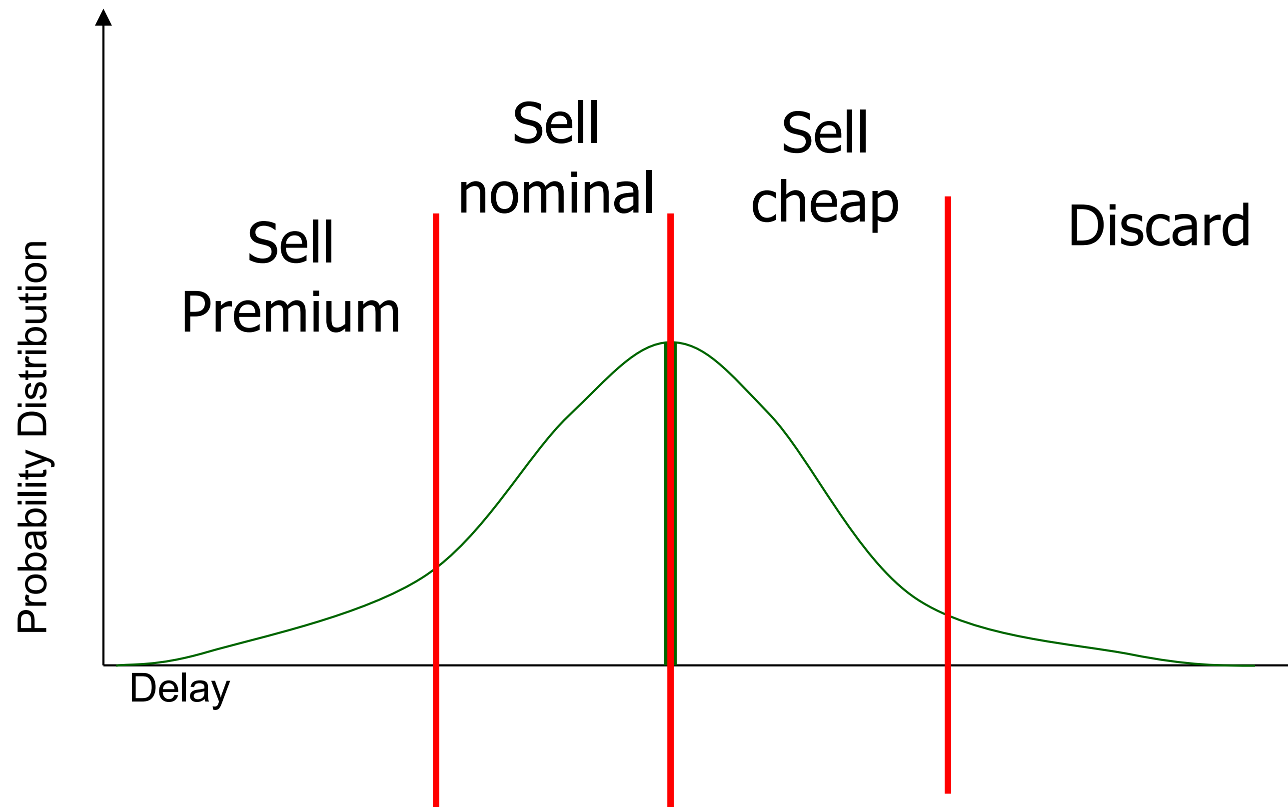
- The difference between statistical corner model and worst-case corner-model
  - Statistical corner model use the realistic PDF of the corresponding model parameter of its typical model
    - PDF is obtained from the distribution of a large set of production data
  - Statistical models can pass a valid design, which were rejected in worst-corner model

Range of Behavior

- Still get range of performances
- Any way to exploit the fact some are faster?

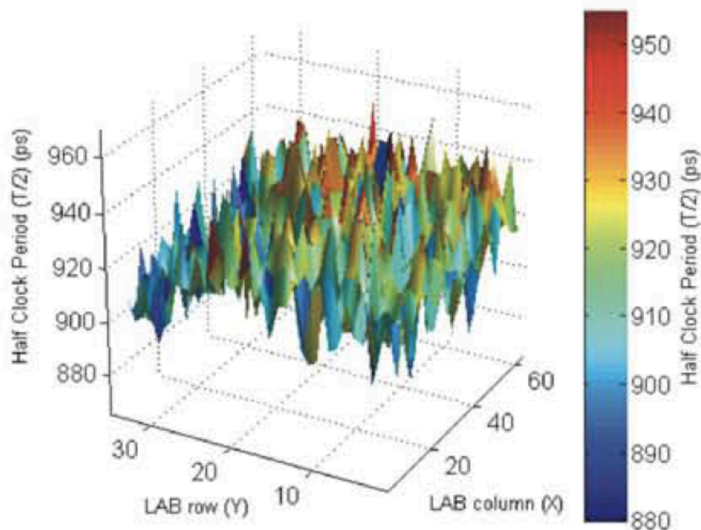


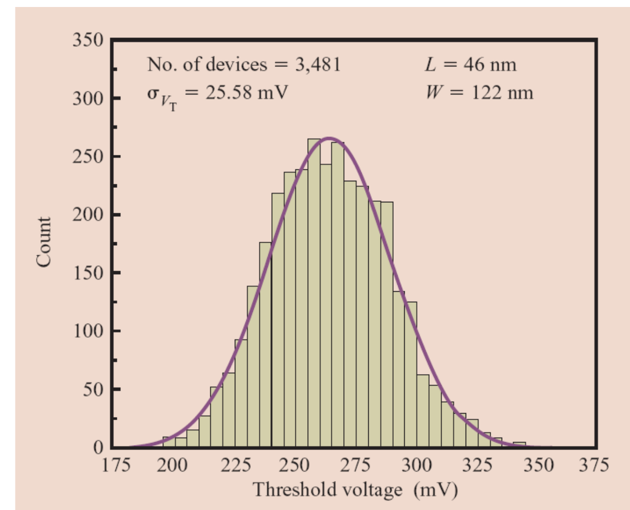
Speed Binning



Big Idea

- Parameters Approximate
- Differ
  - Chip-to-chip, transistor-to-transistor, over time
- Robust design accommodates
  - Tolerance and Margins
  - Doesn't depend on precise behavior





Admin

- HW 3 due tonight @midnight
- Midterm 1 Friday 10/1 (next week)
  - 7-9pm in Towne 309
  - No Lecture, virtual office hours
  - Virtual review session on Wednesday, will be recorded
    - See Piazza for updates
- HW 4 posted Friday 10/1 after midterm
  - Due following Friday 10/8 @midnight