DAT110
Methods for Electronic System Design and Verification

Per Larsson-Edefors
VLSI Research Group

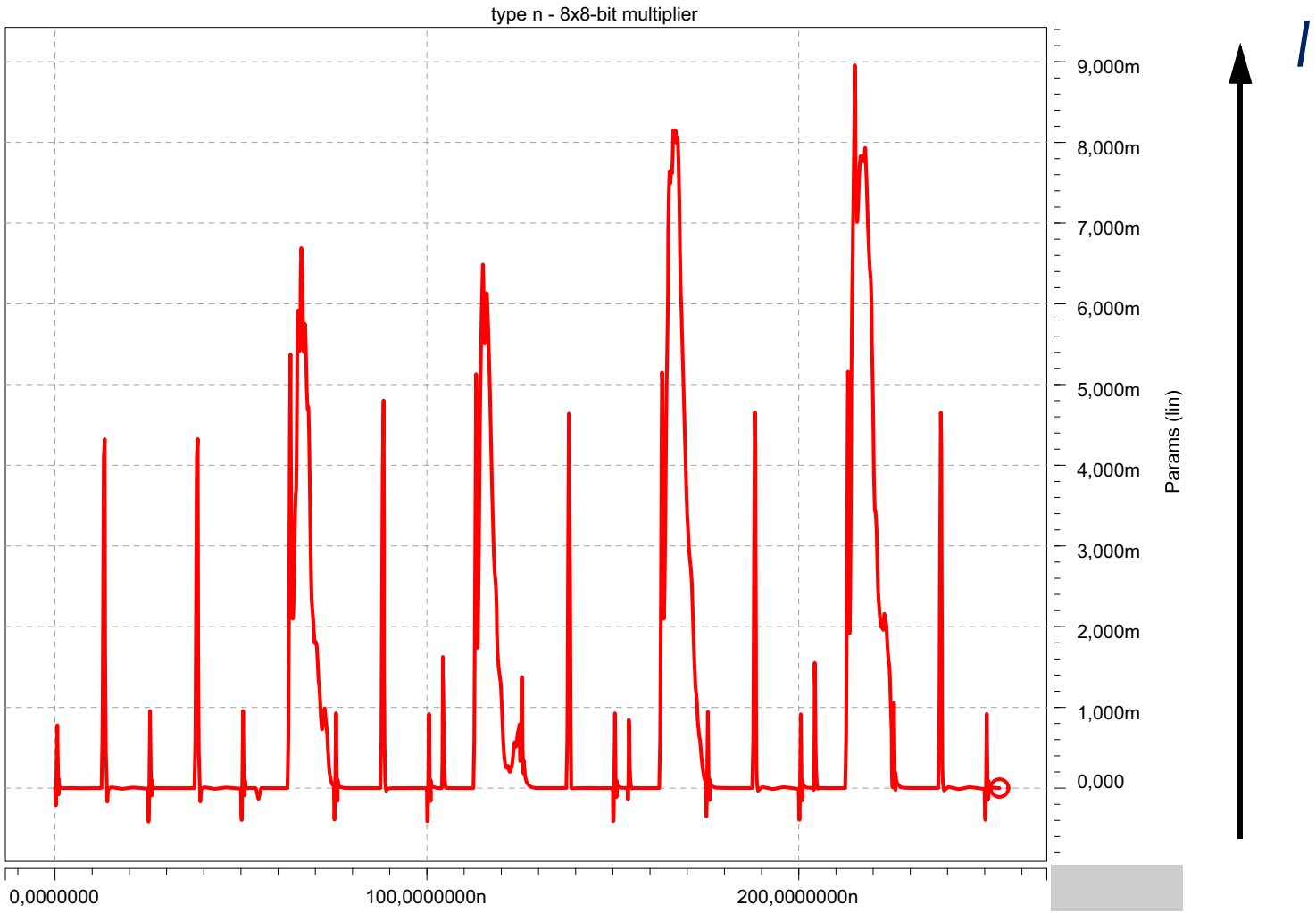
Lecture 6: Power and Energy. Variability.

Back to Basics - Power and Energy

- Charges $Q = I \cdot t$ are transported during time t, thanks to a current I (which is assumed to be constant!).
- Energy $E = Q \cdot V$ is associated with the charges Q that sit inside an electric field, at a potential V.
- Power dissipation $P = E/t$ is energy E expended during time t.
- $$P = \frac{E}{t} = \frac{(Q \cdot V)}{t} = \frac{((I \cdot t) \cdot V)}{t},$$

  which becomes $P = I \cdot V$, when current and voltage are constant.

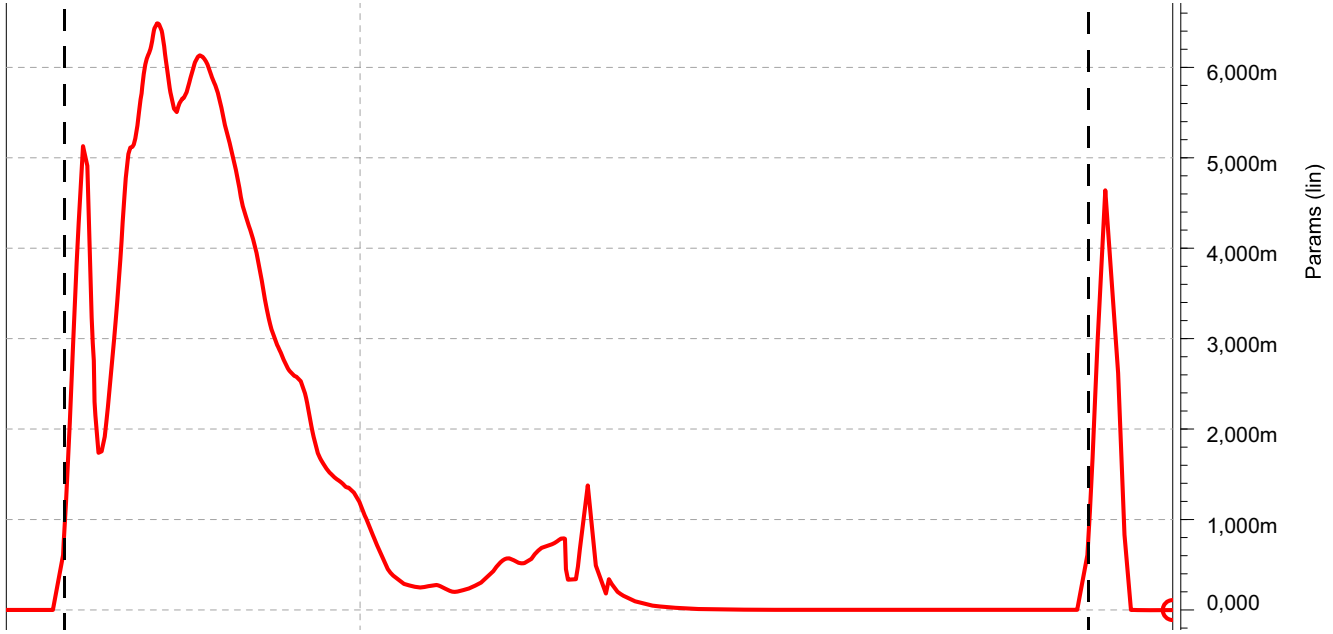
Current Varies (over 10 Cycles)



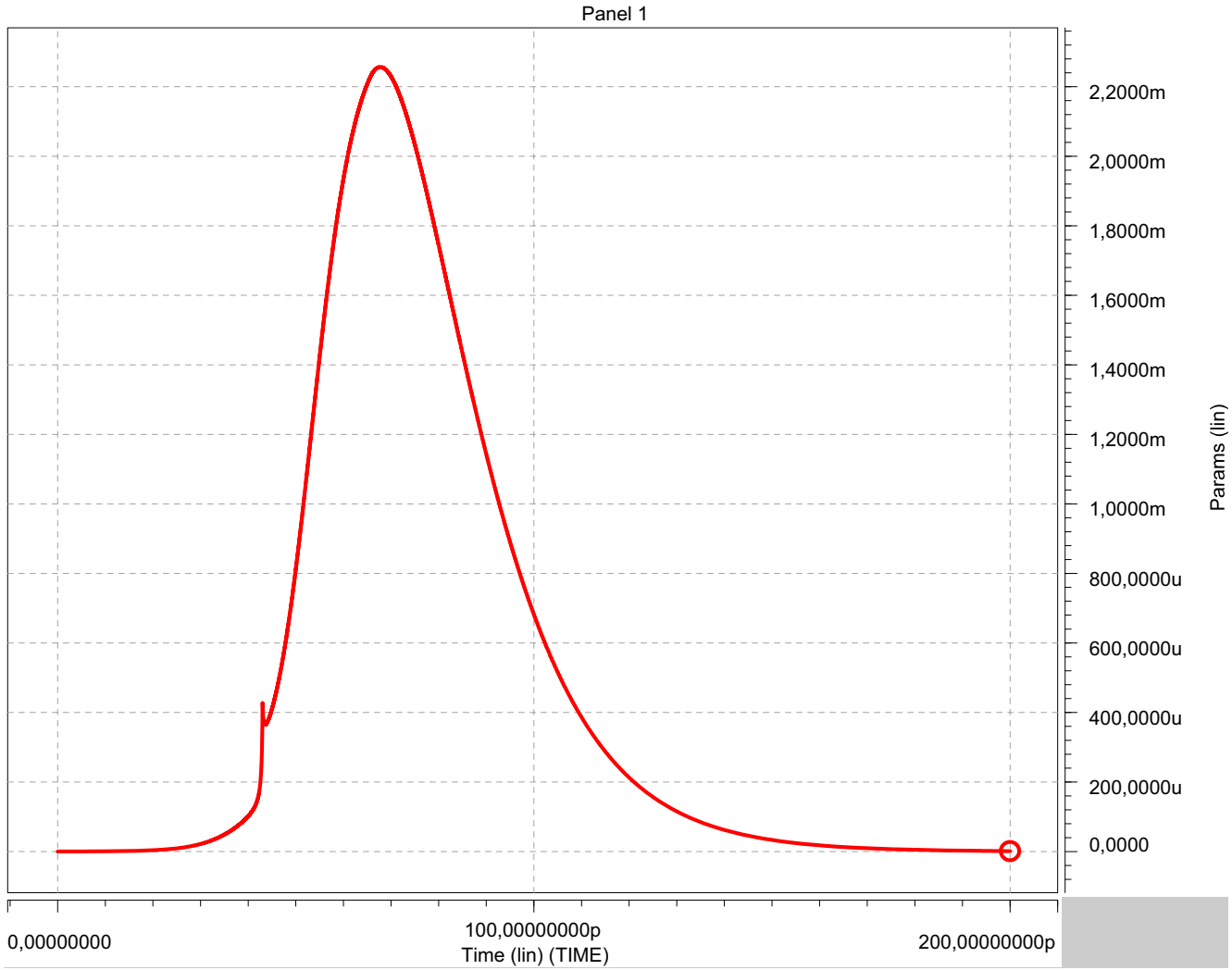
- Current drawn from V_{DD} for an 8x8-bit multiplier.

Close Up of One Clock Cycle

- In a larger circuit, containing many gates, the instantaneous current varies greatly with time.
- A large and diverse set of test vectors cause transitions to happen in a distributed fashion, over time and space.

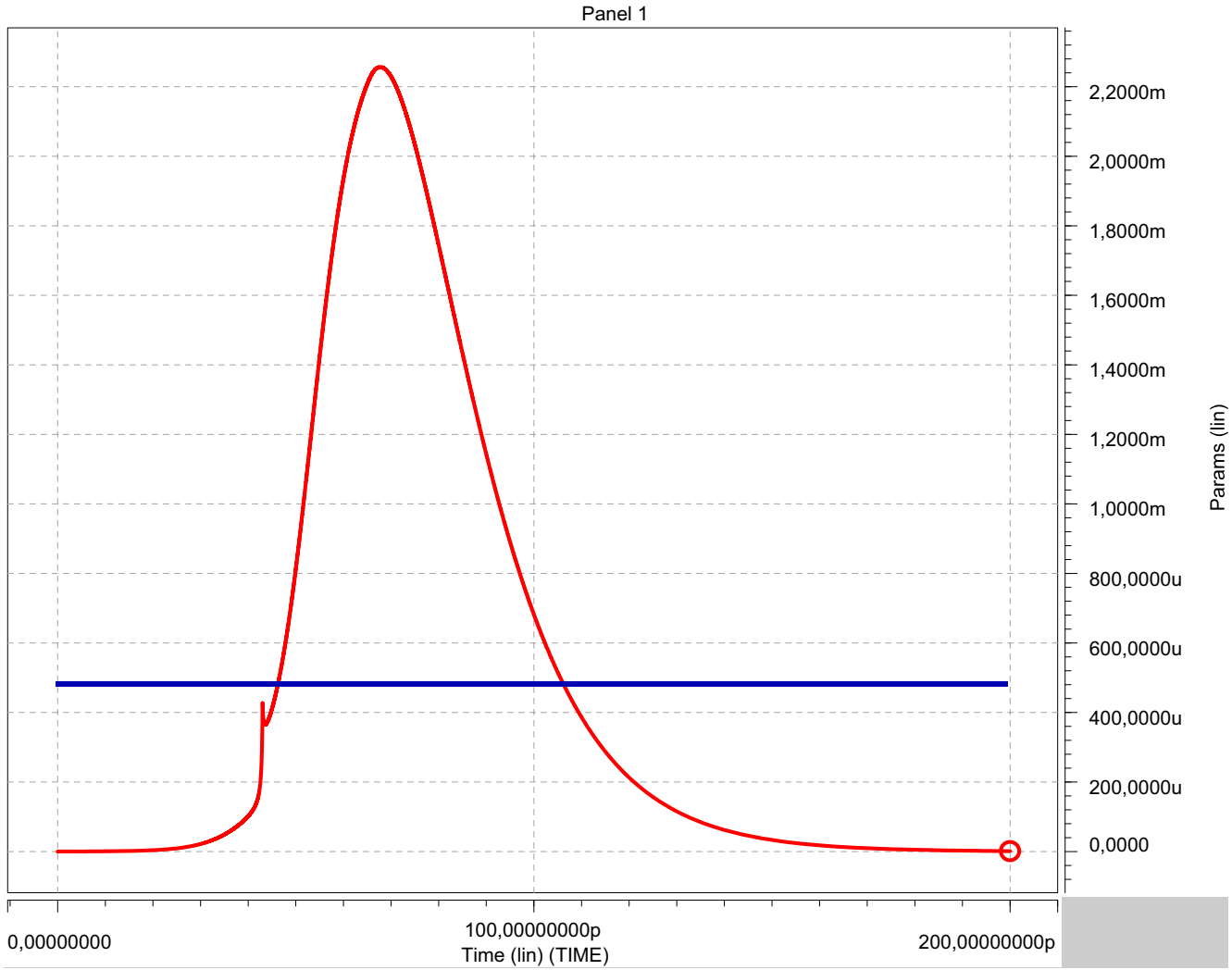


Charging an Inverter Output



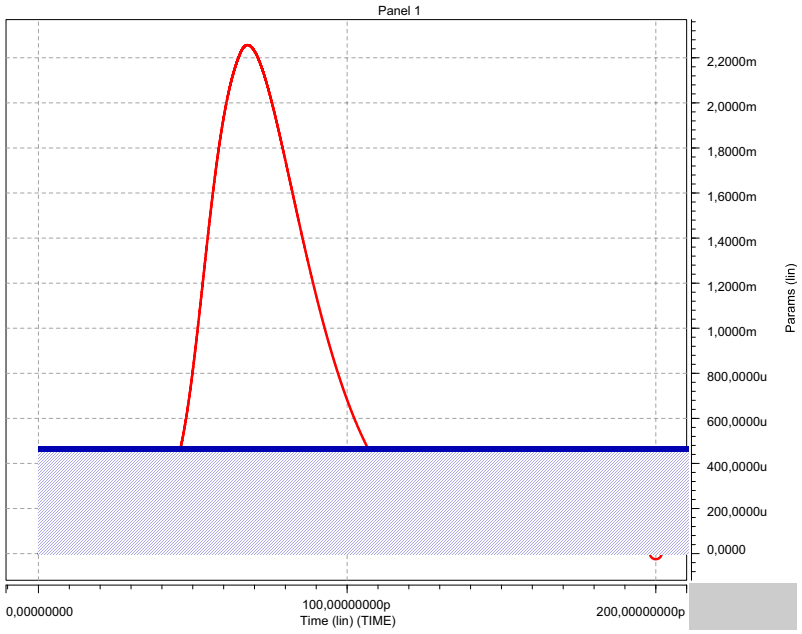
$$Q = \int I \, dt$$

The Average Current - The Blue Line



$$I_{avg} = \frac{\int I \, dt}{T}$$

Energy and Average Power 1(2)



Charge $Q = I_{avg} \cdot T$ drawn from V_{DD} to output node

$\Rightarrow$

electric energy $E = Q \cdot V_{DD}$ extracted from V_{DD}

$\Rightarrow$

an average power of

$$P = \frac{Q \cdot V_{DD}}{T} = I_{avg} \cdot V_{DD}$$

for the clock cycle.

Energy and Average Power 2(2)

- An adder draws an average of 1 mA of current at a supply of 1 V; thus, its power dissipation is 1 mW.
- The adder energy that is expended for an application is

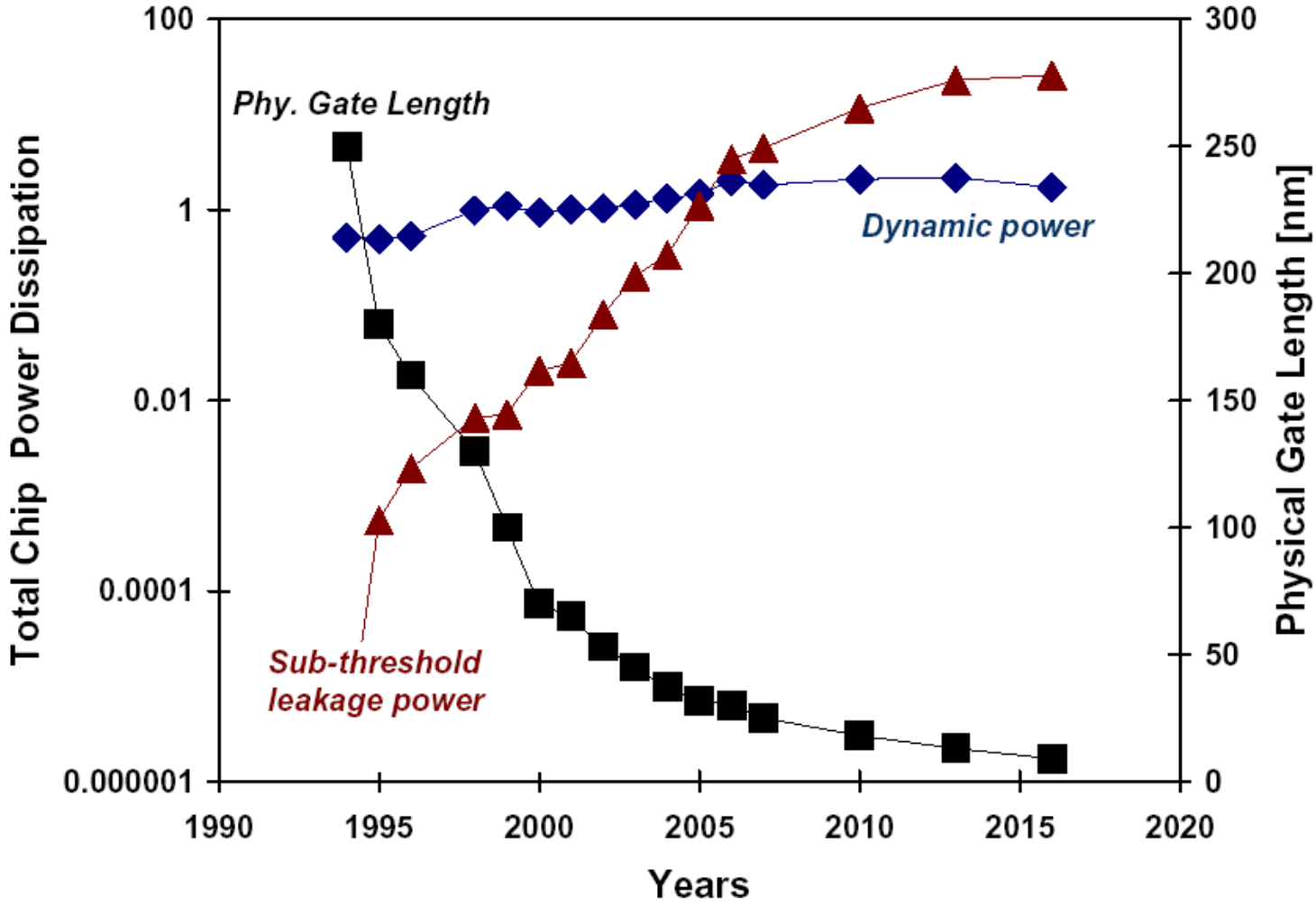
  $$\text{Energy} = 1\ \text{mW} \cdot \text{Execution time}.$$
- If the adder performs 1024 one-cycle additions for the application, the total execution time is

  $$\text{Execution time} = 1024 \cdot \text{Clock period}.$$
- Energy tradeoffs are complex; for example, an increasing circuit complexity (and power dissipation) can lead to faster execution and thus less total energy.

Different Mechanisms of Power Dissipation

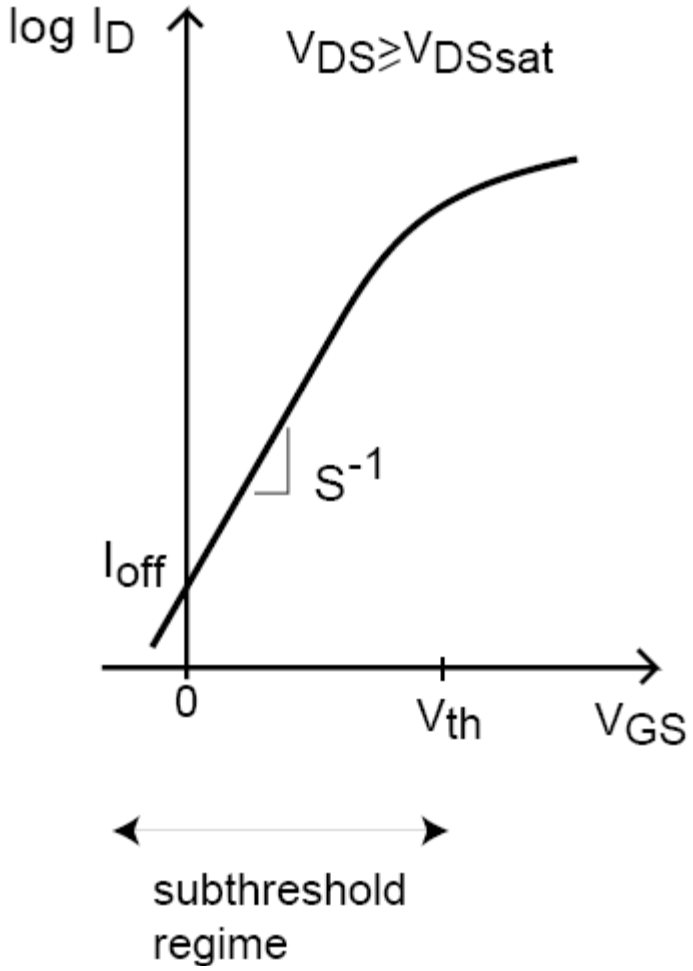
- Dynamic power dissipation, $P_{dynamic}$.
  - Charging/discharging = switching power (P_{sw}).
  - (Short-circuit power, negligible in low V_{DD} processes).
- Static power dissipation, P_{static}.
  - Subthreshold leakage power (P_{sub}).
  - Gate oxide tunneling leakage.
  - Junction leakage.
- The dominant portion has been, still is, and should be, switching power.

… But Static Power Has Been Increasing



Quest for Low Off Currents

- Subthreshold swing (or subthreshold slope factor, *S* below): Which change in V_{GS} results in 10X less current ?



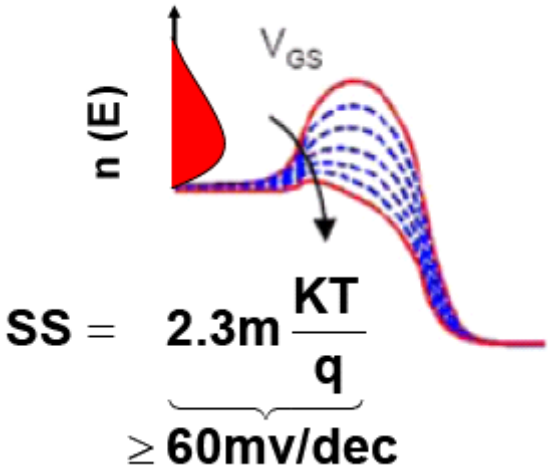
- The MOSFET's swing is limited to >60 mV/dec (in practice >70 mV/dec).
- Research on new devices with steeper subthreshold slope is ongoing.

MOSFET Current in the Subthreshold Regime

$$I_{sub} \propto e^{\frac{q(V_{GS}-V_T)}{kT}}\left(1-e^{-\frac{qV_{DS}}{kT}}\right), \text{ for } V_{GS} < V_T$$

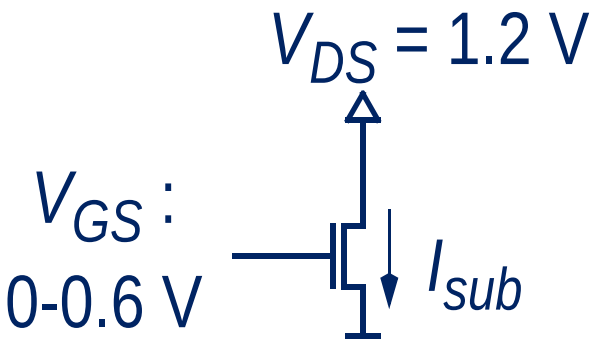
$$V_{thermal} = \frac{kT}{q} = 26\text{ mV (room temp)}$$

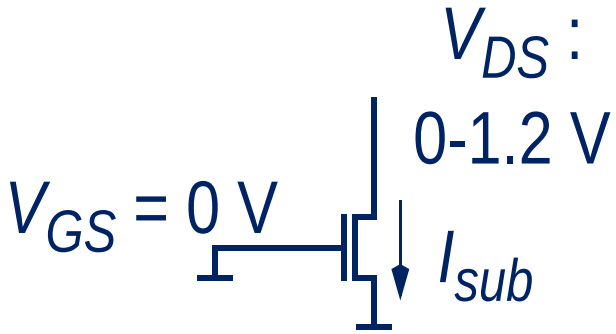
$$\ln(10) \approx 2.3$$



Source: “Band-to-Band Tunneling Field Effect Transistor for Low Power Logic and Memory Applications”, S. A. Mookerjea, PhD Thesis, Penn State Univ, 2010.

Consider Simulation of Single NMOSFET



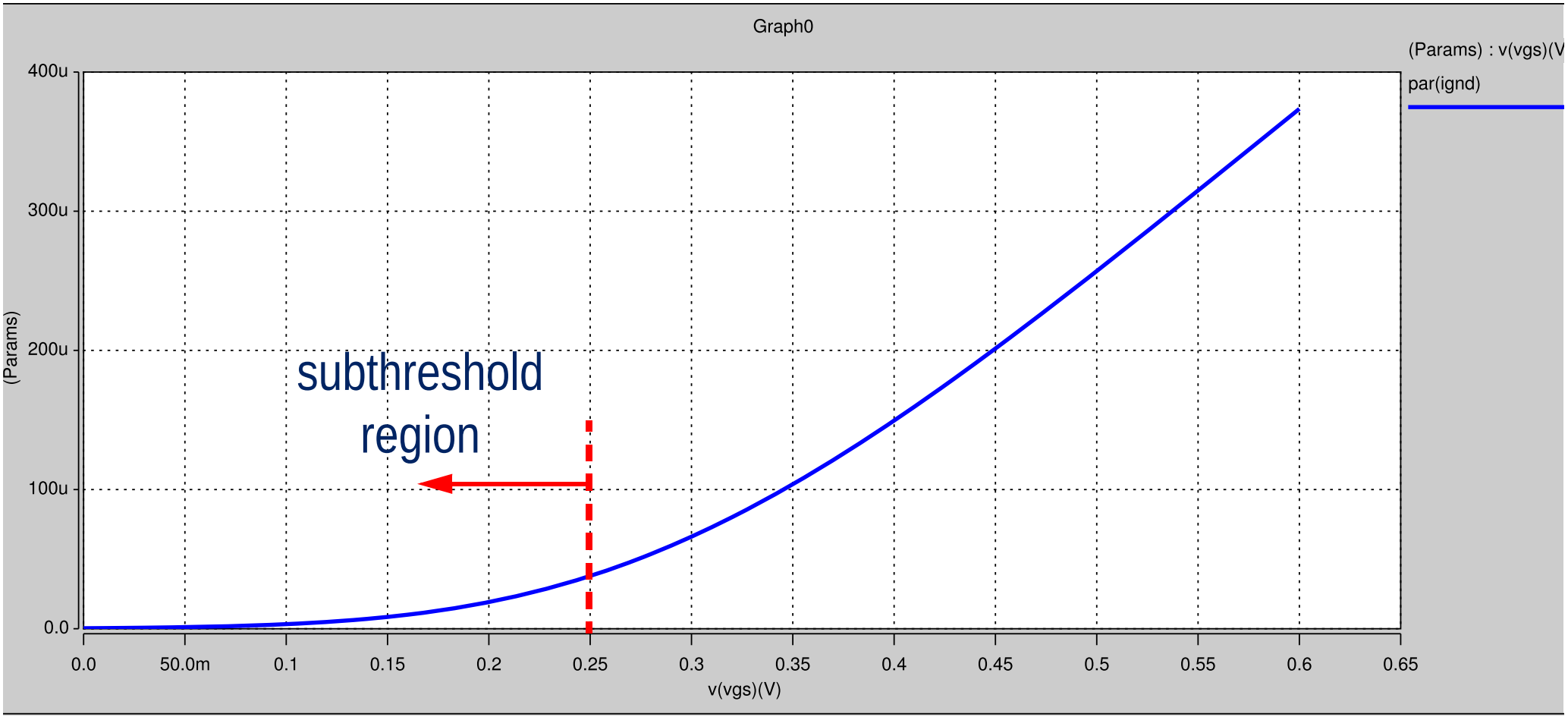


$$I_{sub} \propto e^{\frac{q(V_{GS}-V_T)}{kT}}$$

$$I_{sub} \propto e^{\frac{q(-V_T)}{kT}}\left(1-e^{-\frac{qV_{DS}}{kT}}\right)$$

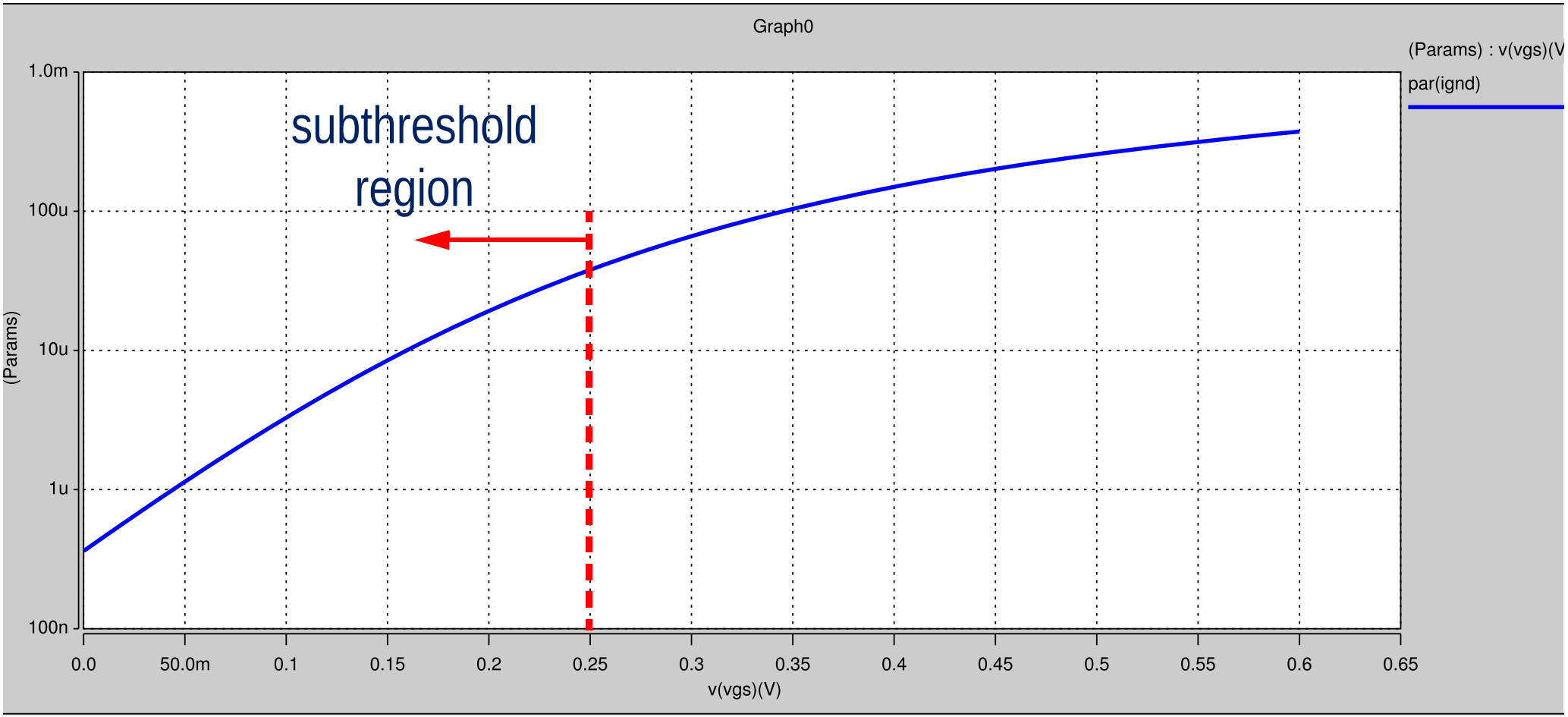
Isub (lin) vs Vgs

- Simulation of 65nm NMOSFET with V_{DS} = 1.2 V and V_T = 0.25 V:

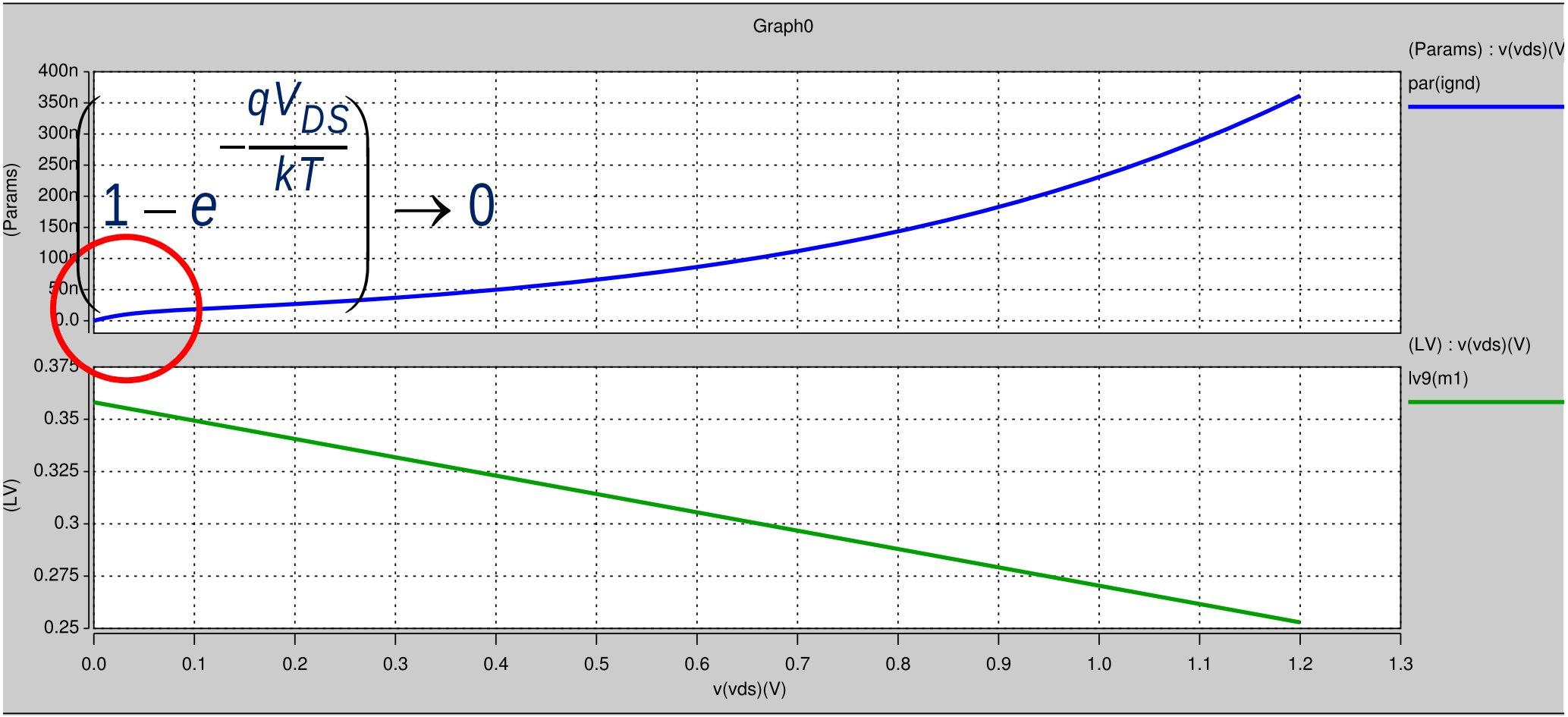


Isub (log) vs Vgs

- Simulation of 65nm NMOSFET with V_{DS} = 1.2 V and V_T = 0.25 V:



Isub vs Vds for Vgs = 0



V_T varies with V_{DS}, making I_{sub} vary with V_{DS}.

DIBL 1(2)

- $I_{sub} \propto e^{\frac{q(V_{GS}-V_T)}{kT}}\left(1-e^{-\frac{qV_{DS}}{kT}}\right)$

  which for an offstate NMOSFET ($V_{GS} = 0$, $V_{DS} = V_{DD}$) becomes

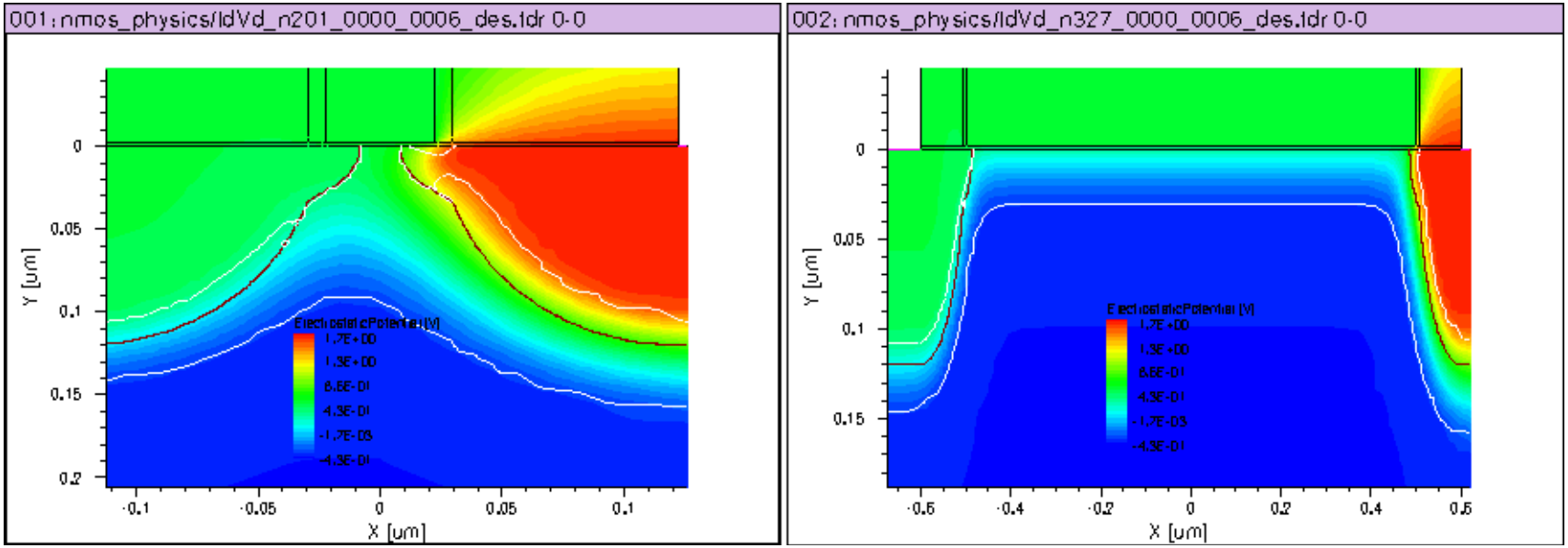
  $I_{sub} \propto e^{-\frac{qV_T}{kT}}$

  where V_T varies with V_{DS}. This is because of drain-induced barrier lowering (DIBL) which is a short-channel effect (SCE).

- Thus, $P_{static} = I_{sub} \cdot V_{DD}$ has a strong dependency on V_{DD}.

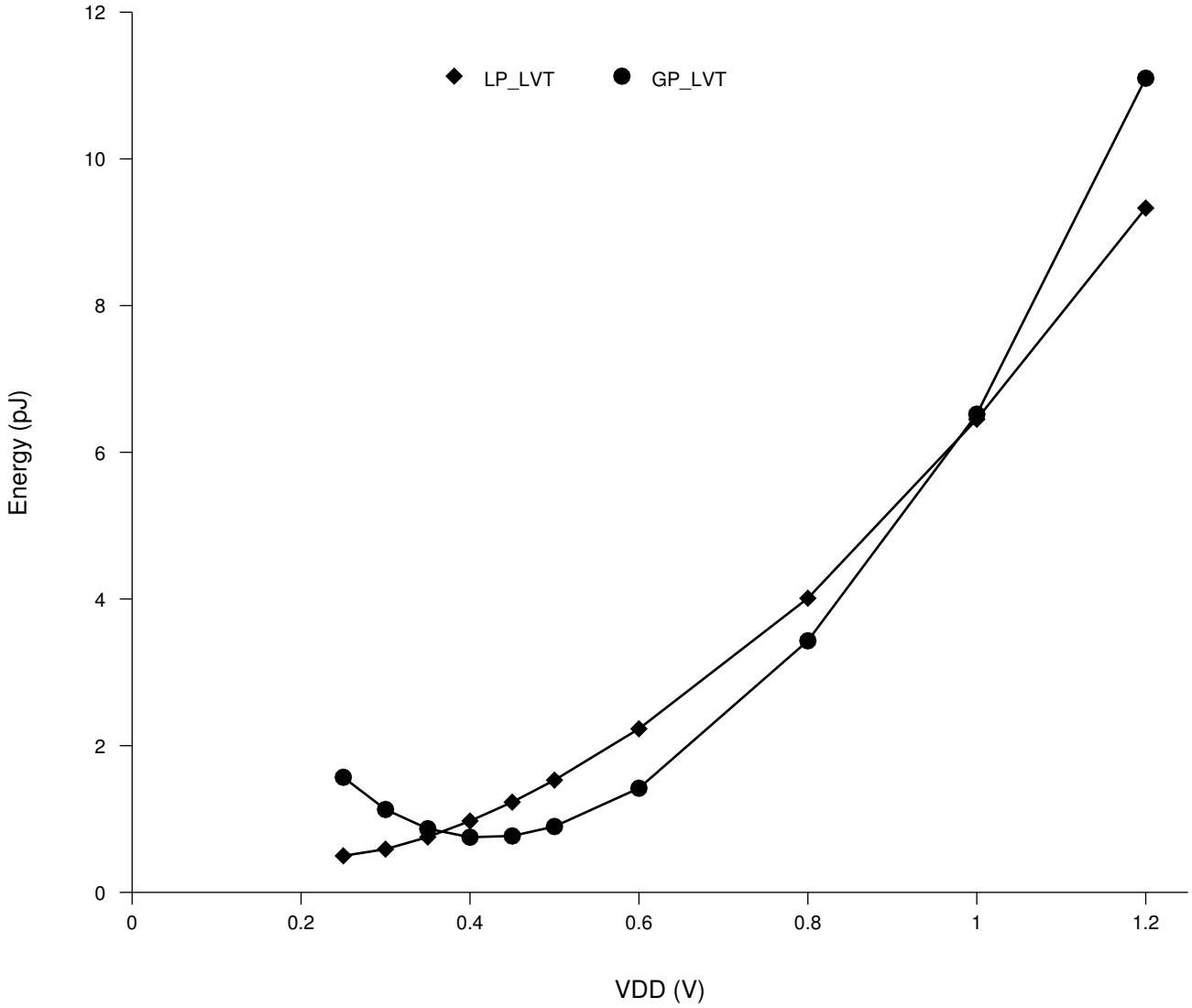
DIBL 2(2)

- The drain voltage, which is decided by V_{DD}, impacts channel electrostatic distribution. A higher V_{DD} will deplete relatively more of the channel when this is short, which in turn increases I_{sub}.

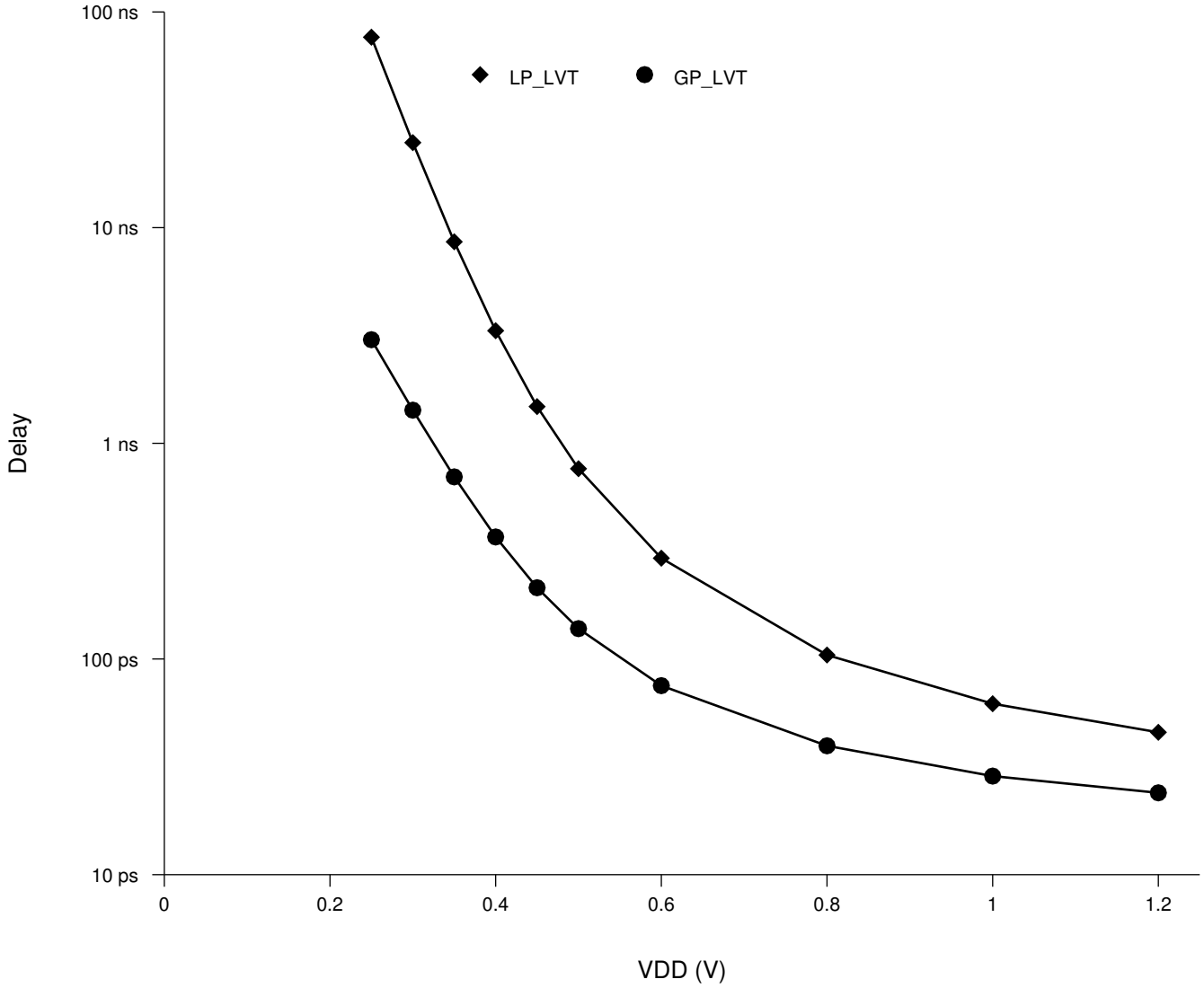


- $\Delta V_T = -\lambda_d V_{DS}$, where λ_d is the DIBL coefficient.

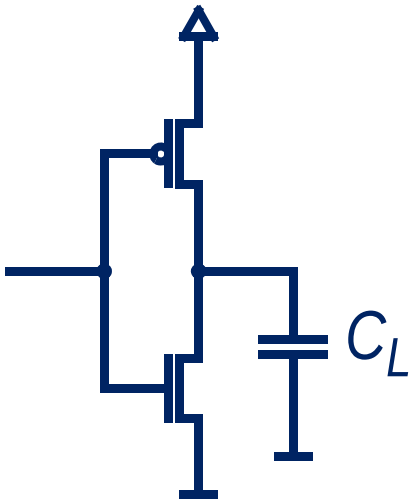
Low-Voltage Operation Is Energy Efficient...



...Since Static Power Dominates Slow Circuits



Switching Power Dissipation



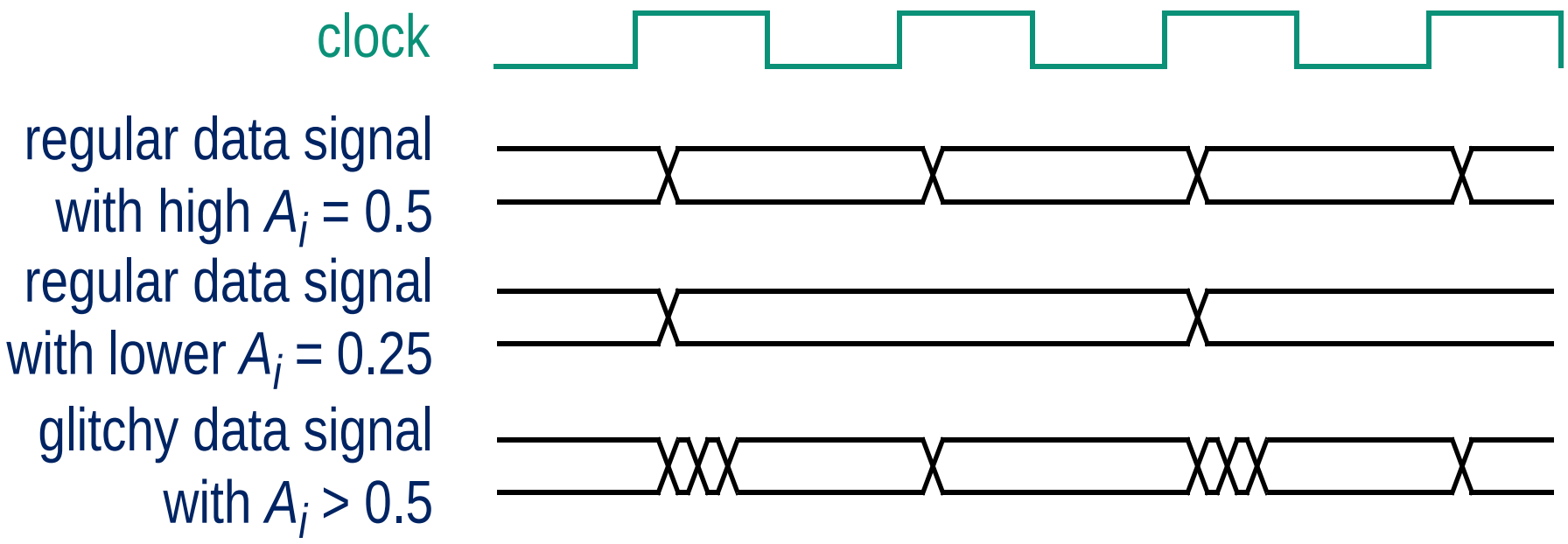
- It takes energy
  $E = Q \cdot V_{DD} = (C_L \cdot V_{DD}) \cdot V_{DD}$
  from V_{DD} to charge the output node.
- During a full transition (output $0 \rightarrow 1 \rightarrow 0$), the electric energy E turns into heat.

- Power dissipation = rate of energy conversion:

$$P_{sw} = \frac{1}{T} \cdot V_{DD}^2 \cdot \text{switched capacitance} = f \cdot V_{DD}^2 \cdot \sum_i A_i \cdot C_i,$$

where A_i is switching activity
on node i (switching activity for $0 \rightarrow 1$).

Switching Activity

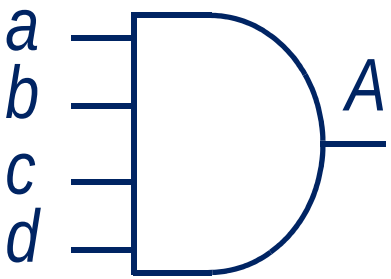


Switching Activity and Probability

- For a 3-bit truth table,
  the rightmost bit has a switching activity of 4/8 = 0.5,
  while the leftmost bit has an activity of 1/8 = 0.125.
  - 0.5 is the maximal activity of data signals.
  - 1 is the activity of non-gated clock.
- For a 3-bit truth table, $P(A) = P(B) = P(C) = 0.5$.
  This means the probability of the bits
  being logical 1s is 0.5.

```
000
001
010
011
100
101
110
111
```

Ex of Simplified Probability Propagation



$P(a{:}1)$ probability of a being logical 1.

$P_{0\to1}(a)$ probability of a switching from 0 to 1.

$P(a{:}1) = 0.2 \Rightarrow P_{0\to1}(a) = 0.16.$

$P(b{:}1) = 0.9 \Rightarrow P_{0\to1}(b) = 0.09.$

$P(c{:}1) = 0.7 \Rightarrow P_{0\to1}(c) = 0.21.$

$P(d{:}1) = 0.5 \Rightarrow P_{0\to1}(d) = 0.25.$

$P(A{:}1) = 0.2 \cdot 0.9 \cdot 0.7 \cdot 0.5 = 0.063 \Rightarrow$

$P_{0\to1}(A) = (1\text{-}0.063) \cdot 0.063 \approx \underline{0.059}.$

Switching Power Analysis 1(2)

- Simulation-based (dynamic and iterative) techniques.
  - Simulate the circuit and capture current/power.
  - Choose input vectors - few but representative?
  - What models, what simulator to be used?
- Probabilistic (static and deterministic) techniques.
  - Propagate signal switching activities at primary inputs, to find switching probabilities of all internal nodes of the circuit.
  - How are primary activities obtained - just guesses or use cases?
  - Correlations (temporal and spatial) are hard to capture.

Read [3.2.4/3.1.4 Logic-Level Power Estimation in Vol 2/Ch 3].

Switching Power Analysis 2(2)

- In the lab exercises you will get a chance to work with, first, a probabilistic technique for approximative analyses.
- Then, for higher accuracy, simulation-based techniques can be employed:
  1. logic simulation is used to establish actual switching activities for a particular set of test vectors (VCD/SAIF), by toggle counting.
  2. the switching activities obtained from simulation are fed to a gate-level power analysis tool that considers both switching and leakage power.

Working with Switching Information 1(3)

- As a result of a simulation, consider a TCF (toggle count format) file:

```
"Clk" : "0.50000 855432000";
```

  The clock is logic 1 for half of the simulation time.
  The second number gives 855,432,000 / 1e+9 =
  0.855 toggle / ns =
  2 toggles / actual clock period (here 2.338 ns).

- VCD captures the waveform, while SAIF accumulates toggling count only, to save space:
  - SAIF = 572 kB
  - VCD = 7.907 MB.

Working with Switching Information 2(3)

- 1,000 random vectors vs. 233,450 use-case vectors.
- TCF information for random:

```
"A[15]" : "0.51520 225271000";
"A[16]" : "0.50570 220151000";
```

- TCF information for use case:

```
"A[15]" : "0.04130 29906000";
"A[16]" : "0.54440 143117000";
```

Working with Switching Information 3(3)

- TCF information for random:
  - 225,271,000 / 1e+9 = 0.222527 toggle/ns = 0.5203 toggles / clock period
  - 220,151,000 / 1e+9 = 0.220151 toggle/ns = 0.5147 toggles / clock period
- TCF information for use case:
  - 29,906,000 / 1e+9 = 0.0699 toggles / actual clock period
  - 143,117,000 / 1e+9 = 0.3346 toggles / actual clock period

The Parameter Called Togg (RC)

- *togg* (toggling rate) describes how many times a signal toggles during 1 ns. Thus, *togg* fuses information on *f* and *Ai* as defined in P_{sw}: $f \cdot V_{DD}^2 \cdot \sum_i A_i \cdot C_i$
  - *togg*=1 could either mean that you have a signal with *f* = 500 MHz and *Ai* = 1, or *f* = 1 GHz and *Ai* = 0.5, or ...
  - *Ai* only counts 0→1 transitions, while *togg* counts both 0→1 and 1→0.
- *togg* = '*f* normalized to 1 GHz' * (*Ai**2), for example, (0.769 GHz/1 GHz) * (0.2*2) = 0.308.

Use Cases versus Random Data

- Power report for random data:

Instance	Cells	Leakage Power(nW)	Dynamic Power(nW)	Total Power(nW)
ALU	949	314,465	8,791,824	9,106,290

- Power report for use-case test vectors

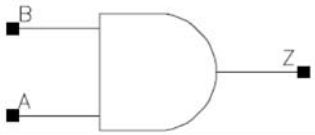
Instance	Cells	Leakage Power(nW)	Dynamic Power(nW)	Total Power(nW)
ALU	949	319,967	4,218,725	4,538,692

- Power depends strongly on input data.

Cell-Library Power Dissipation Properties

HS65_LS_AND2

Logical Symbol



Average Leakage Power (mW) at 25C, 1.20V Typ process

	vdd
X4	1.251e-07
X9	2.207e-07
X18	4.318e-07
X27	6.691e-07
X35	8.865e-07

Internal Energy (uW/MHz) at Minimum Output Load

Pin Cycle (vdd)	X4	X9
A (stable)	5.170e-05	9.591e
B (stable)	1.903e-04	3.556e
A to Z	5.500e-03	8.875e
B to Z	5.203e-03	8.359e

Pin Capacitance (pf) at 25C, 1.20V Typ process

A	0.0010	
B	0.0012	

Power Components in RTL Compiler

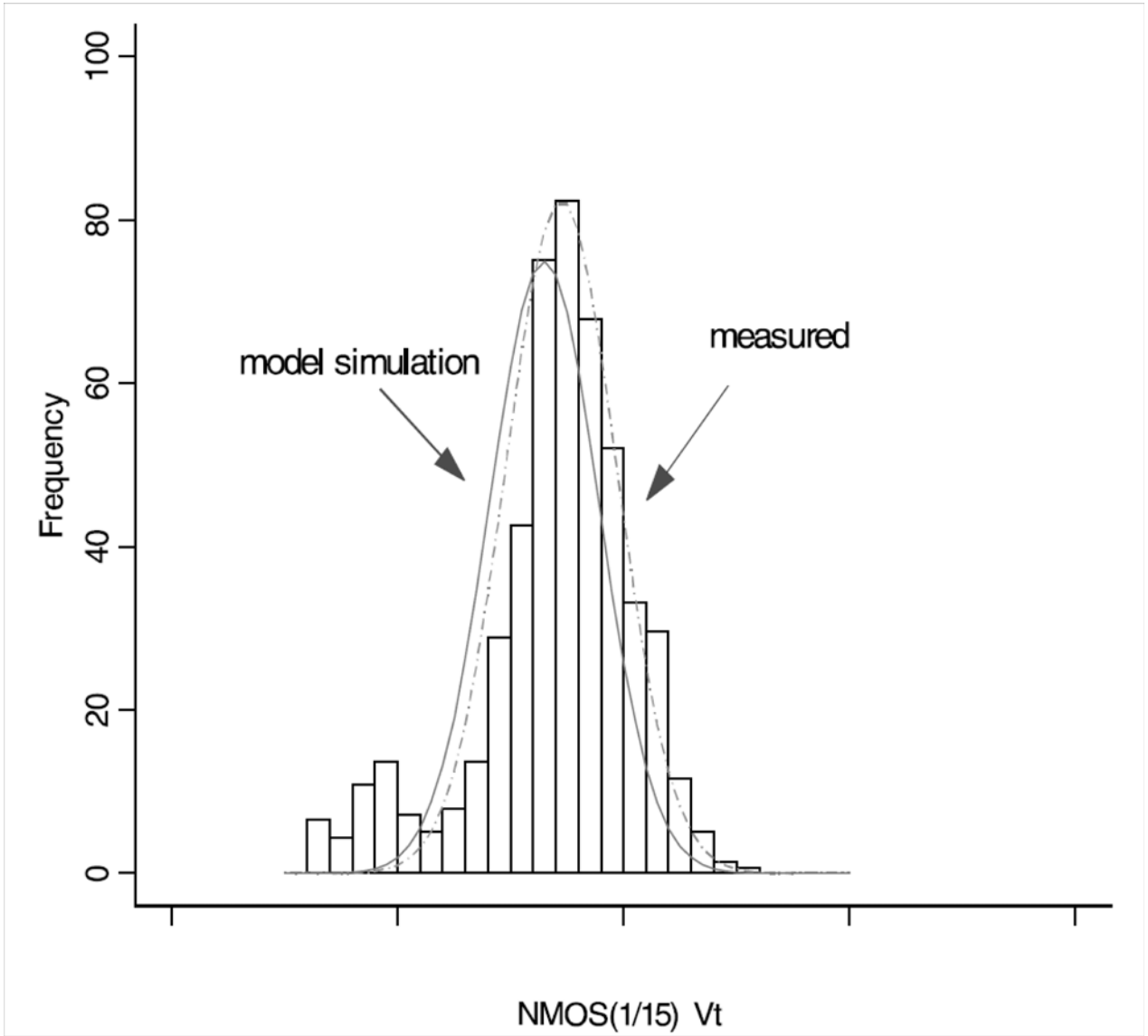
- Leakage power dissipation.
  - Leakage power annotation in the synthesis library.
  - Multi-V_T cells (for example, high $V_T \Rightarrow$ low leakage power, slow = poor timing performance) are supported in RTL Compiler.
- Dynamic power dissipation.
  - Includes switching and short-circuit power.
  - Both internal to cell and wires between cells.

Evaluating Power and Energy

- To get low power dissipation,
  simply reduce clock rate; UNFAIR!
- When comparing circuits, use a reasonable metric:
  - First, ensure the circuit fulfils the performance requirement.
  - Use power-delay-product *(this gives energy per clock cycle)*.
  - Or use energy-delay-product.

Variability

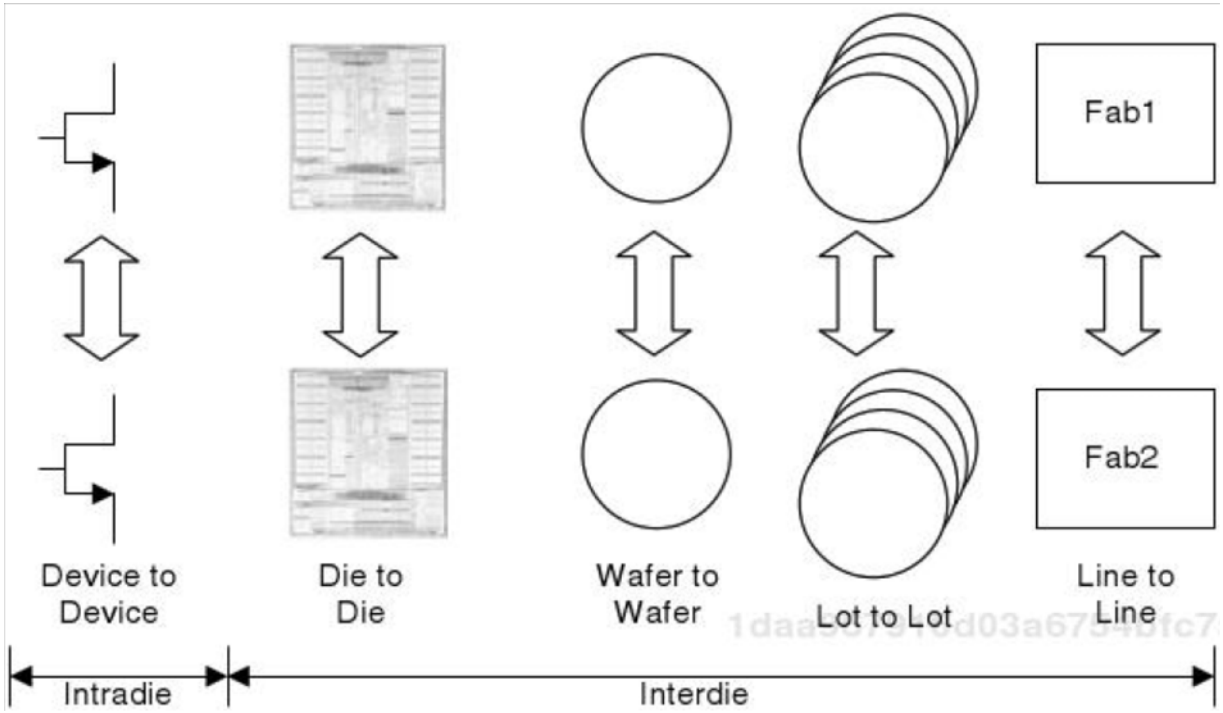
Device Parameters Vary



Source: Qiang Zhang et al., Solid-State Electronics 2001

Different Types of Variation

- Variations manifest in different ways.
  - Between dies on the same wafer - inter-die variations.
  - Between cells/blocks on the same die - intra-die variations.



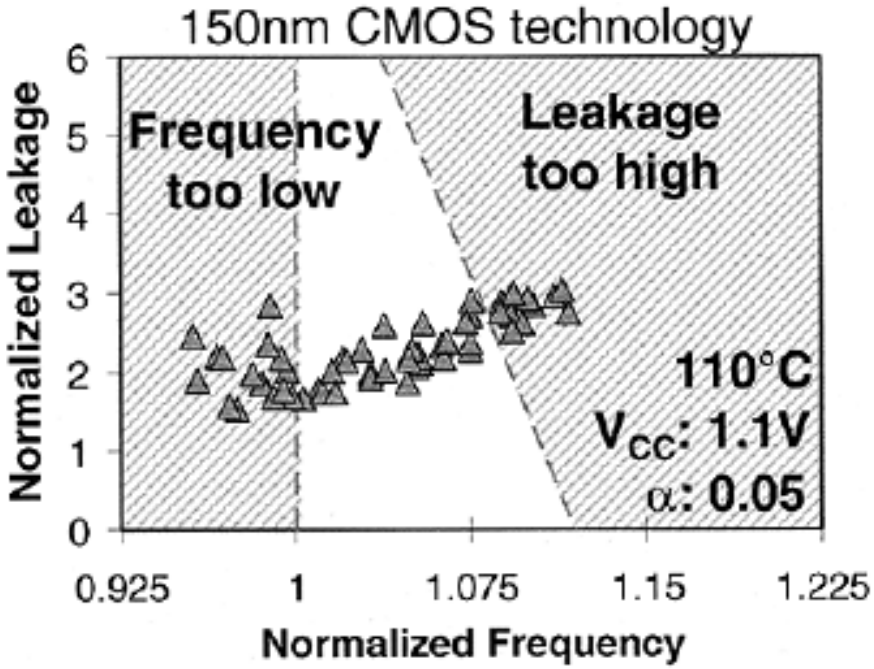
Source: Wang et al., Nano-CMOS Circuit and Physical Design, 2005

Combatting Variations

- Static power is a strong function of process and operational parameters. Thus, variations affect power dramatically.

$$I_{OFF} \propto e^{\frac{q(V_{GS} - V_T)}{kT}}$$

- Performance and power binning is common to qualify processors.
  - Body biasing can be used at production to fine tune threshold voltages.

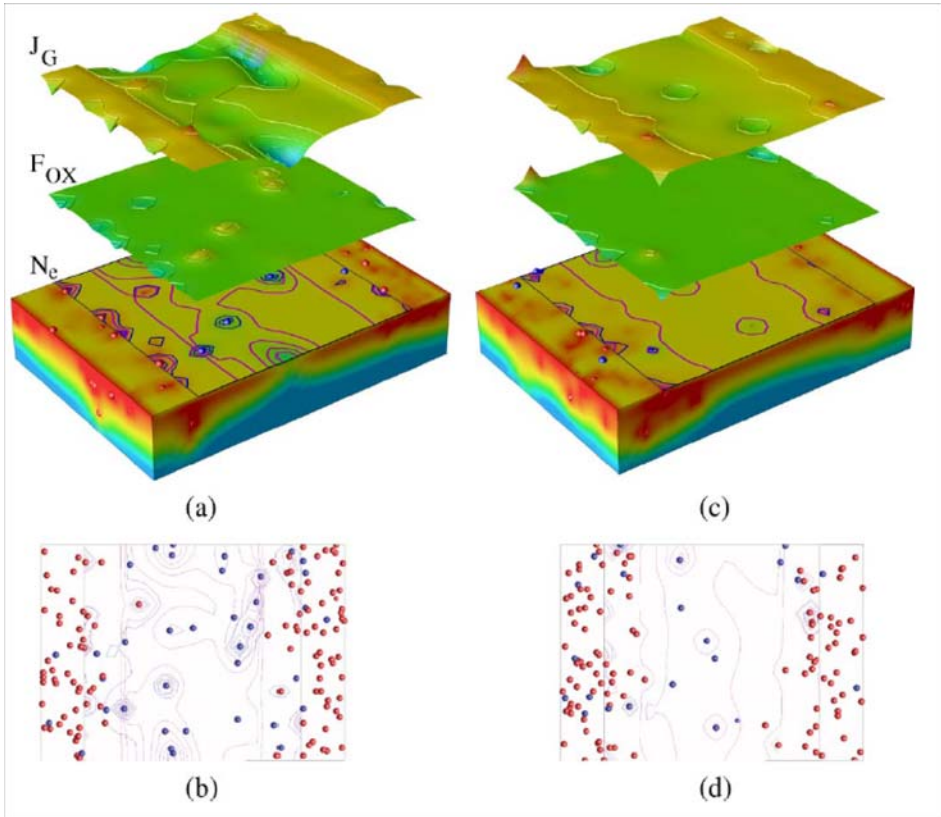


Source: Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage, IEEE JSSC, 2002.

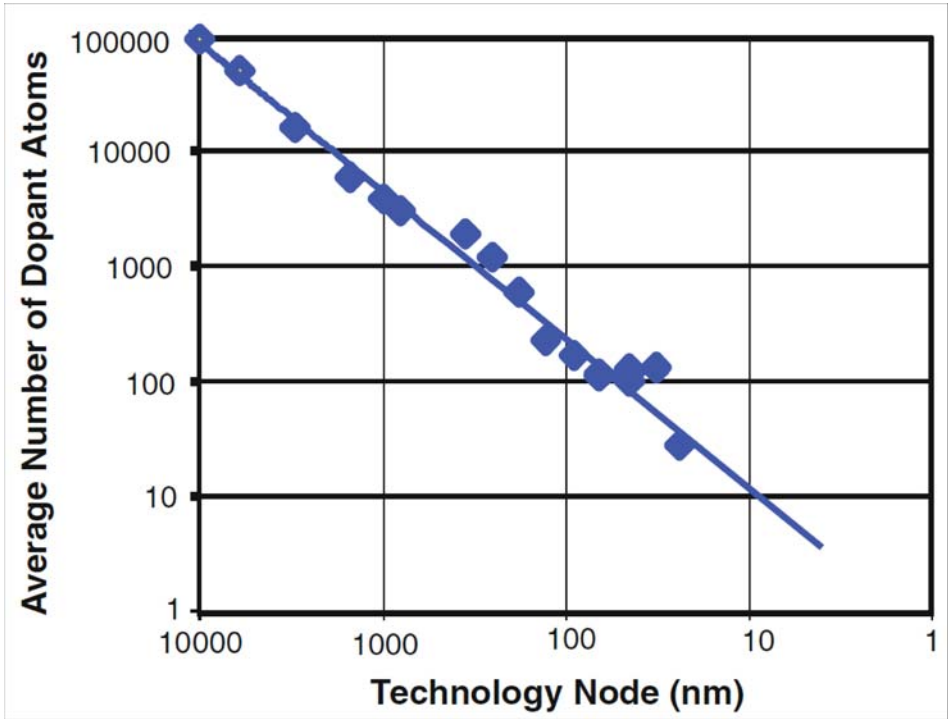
Scaling Increases Parameter Variations

- Static variations.
  - Imperfect manufacturing (CMP, lithography): Systematic variations.
  - Small devices means very few dopants: Random fluctuations.
- Dynamic variations.
  - Temperature changes.
  - Different operating modes (power on/off) and different test vectors are used.

Random Dopant Fluctuation



Source: Direct Tunnelling Gate Leakage Variability in Nano-CMOS Transistors, IEEE TED, 2010.

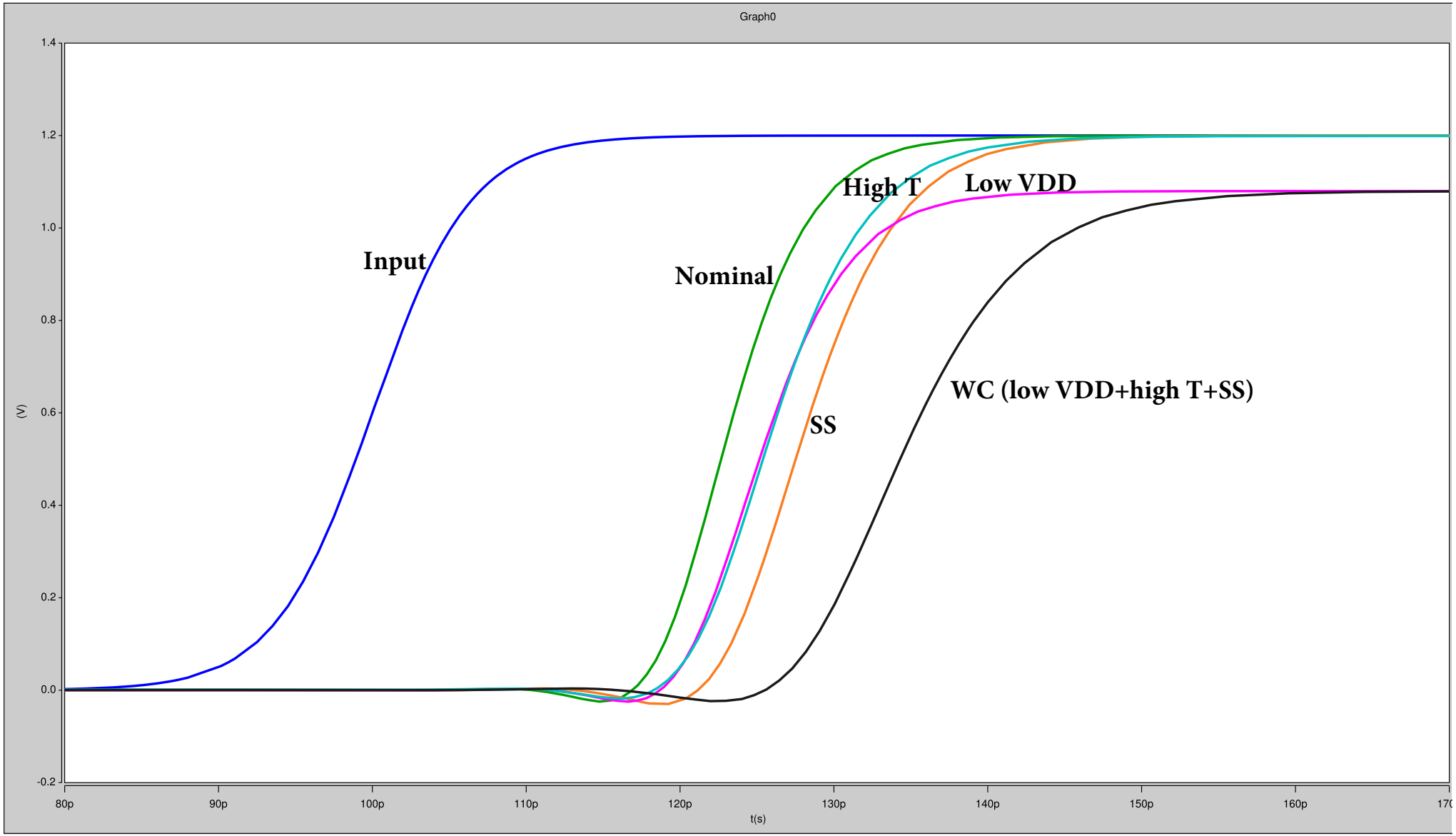


Source: Analog IC Reliability in Nanometer CMOS, Springer, 2013.

Variation-Aware Design

- Given the I-V relation for MOSFETs, vary any of tox ($\Rightarrow$ Cox), W, L or VT and think about impact on delay.
  - For example, a reduced W increases delay.
- Consider variations by using design corners (PVT) or only process corners (P).
- Use variation-aware simulation strategies.
  - Sigma-based simulation.
  - Monte-Carlo (random) simulation.

Impact of Design Corners on Delay

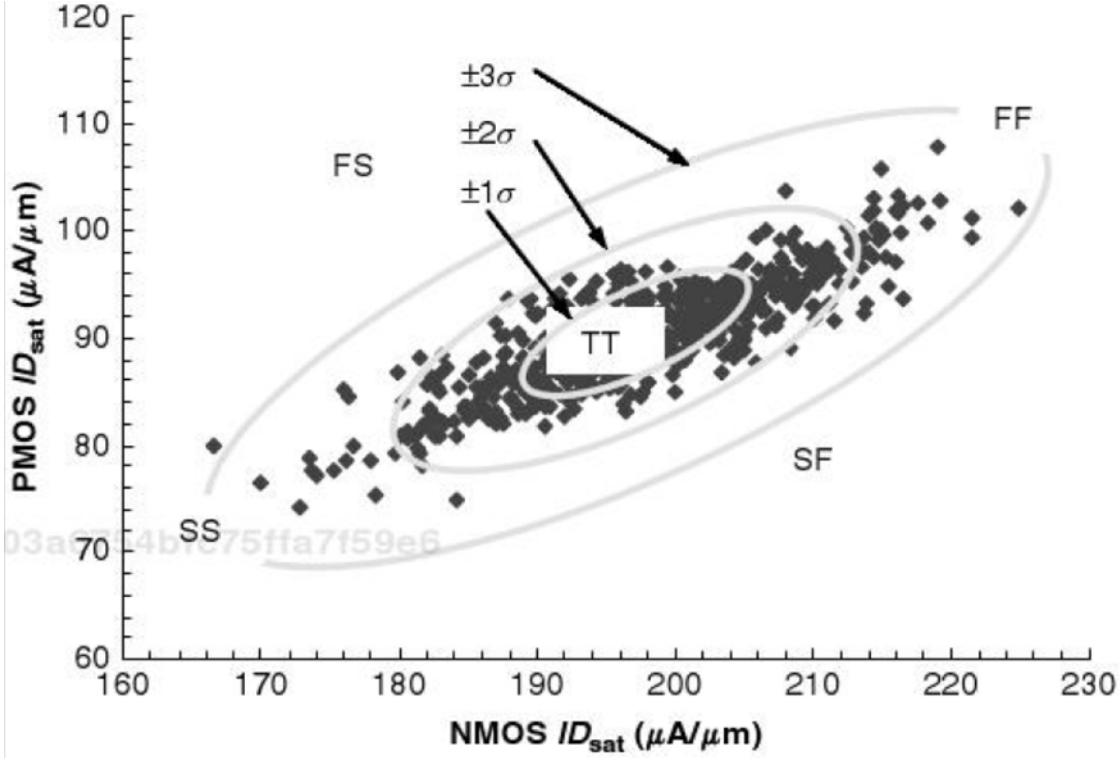


Process, Voltage and Temperature (PVT)

- Static variations (P - process).
  - Process variation captured in process corners.
  - Fast, Typical/Nominal, Slow.
- Dynamic variations (V - supply voltage, T - temperature).
  - Supply voltage changes with current computation (*IR drop + switching noise are treated later*).
  - Die temperature depends on ambient temperature, cooling system, and on current and past computation pattern.

Process Corners

- Process corners refer to fabricated transistors turning out to be fast, slow or nominal/typical.
- FS = Fast NMOS, Slow PMOS, while TT= Typical NMOS, Typical PMOS.

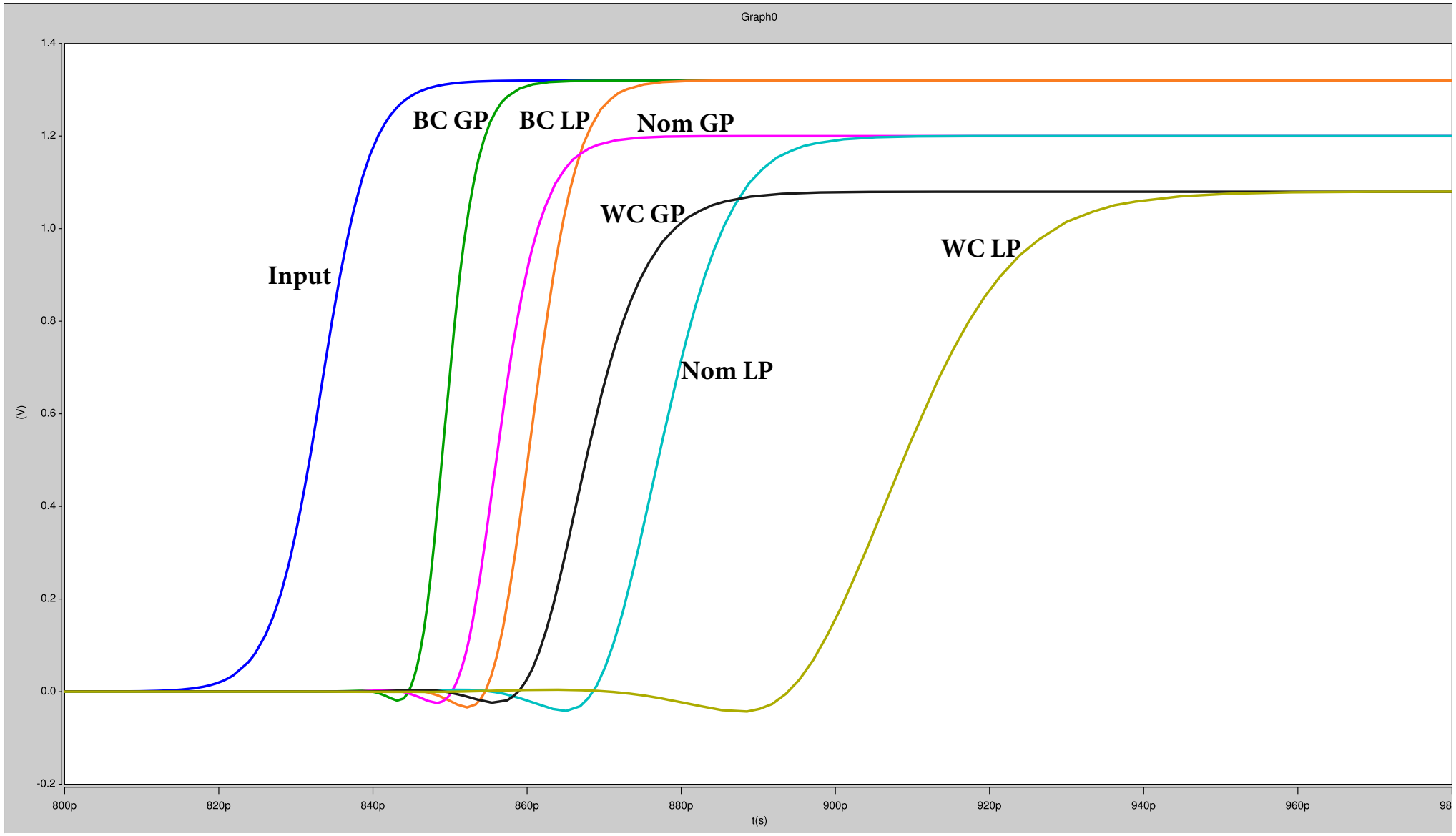


Source: Wang et al., Nano-CMOS Circuit and Physical Design, Wiley, 2005

Corner-Based Design

- Important to run setup time analysis
  for worst-case timing conditions, that is,
  Slow NMOS + Slow PMOS + High T + Low VDD.
- Conversely it is important to check
  for hold time violations under best-case timing conditions:
  Fast NMOS + Fast PMOS + Low T + High VDD.
- Power analysis can be done for typical conditions,
  while worst-case static power can be analyzed
  using best-case timing.

Impact of Worst- and Best Case Conditions



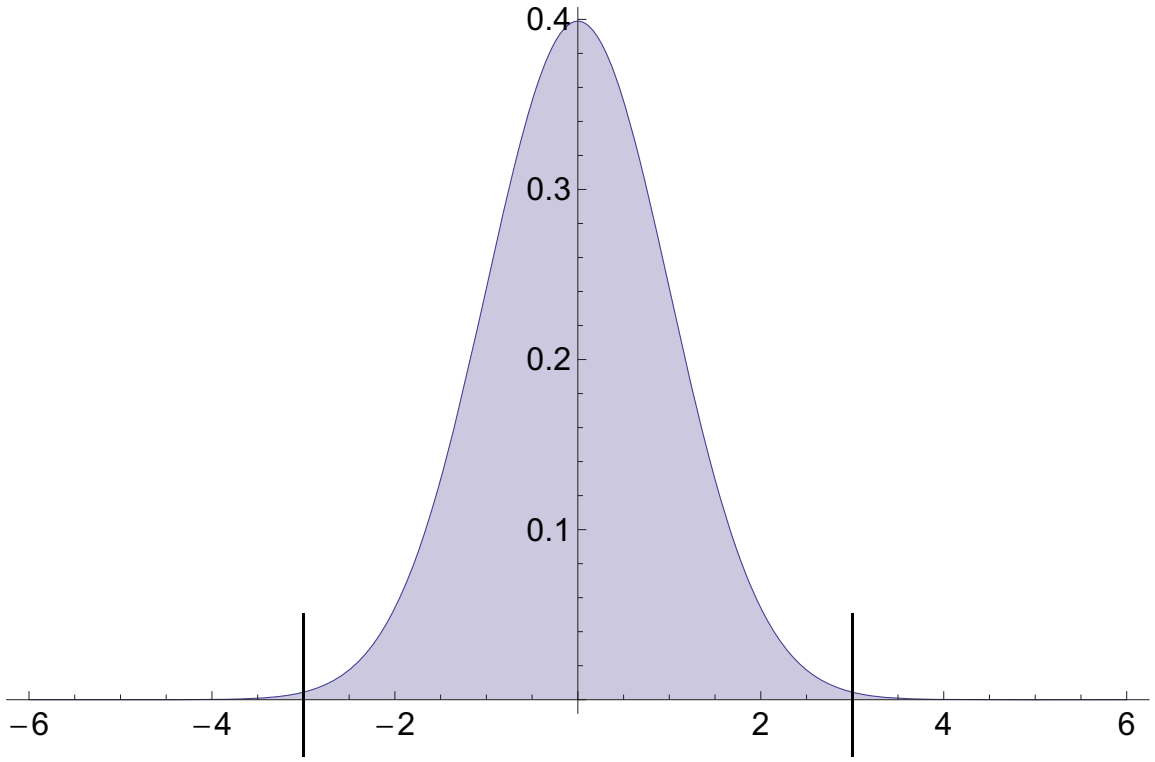
Methods Labs

- /usr/local/cad/stm-cmos065-5.4/CORE65LPSVT/5.2.c/libs/CORE65LPSVT_nom_1.20V_25C.lib
- [Assignment 3.3.2]: CORE65GPSVT_nom_1.10V_125C.lib
- mmmc.tcl:

```
# Define delay corners and analysis views.
create_delay_corner -name typical_corner \
    -library_set typical_lib \
    -rc_corner typical_rc
```

Standard Normal Distribution

- Random parameters are described with a probability density function (PDF) that follows a normal distribution.
- The standard deviation is denoted *sigma*, σ.



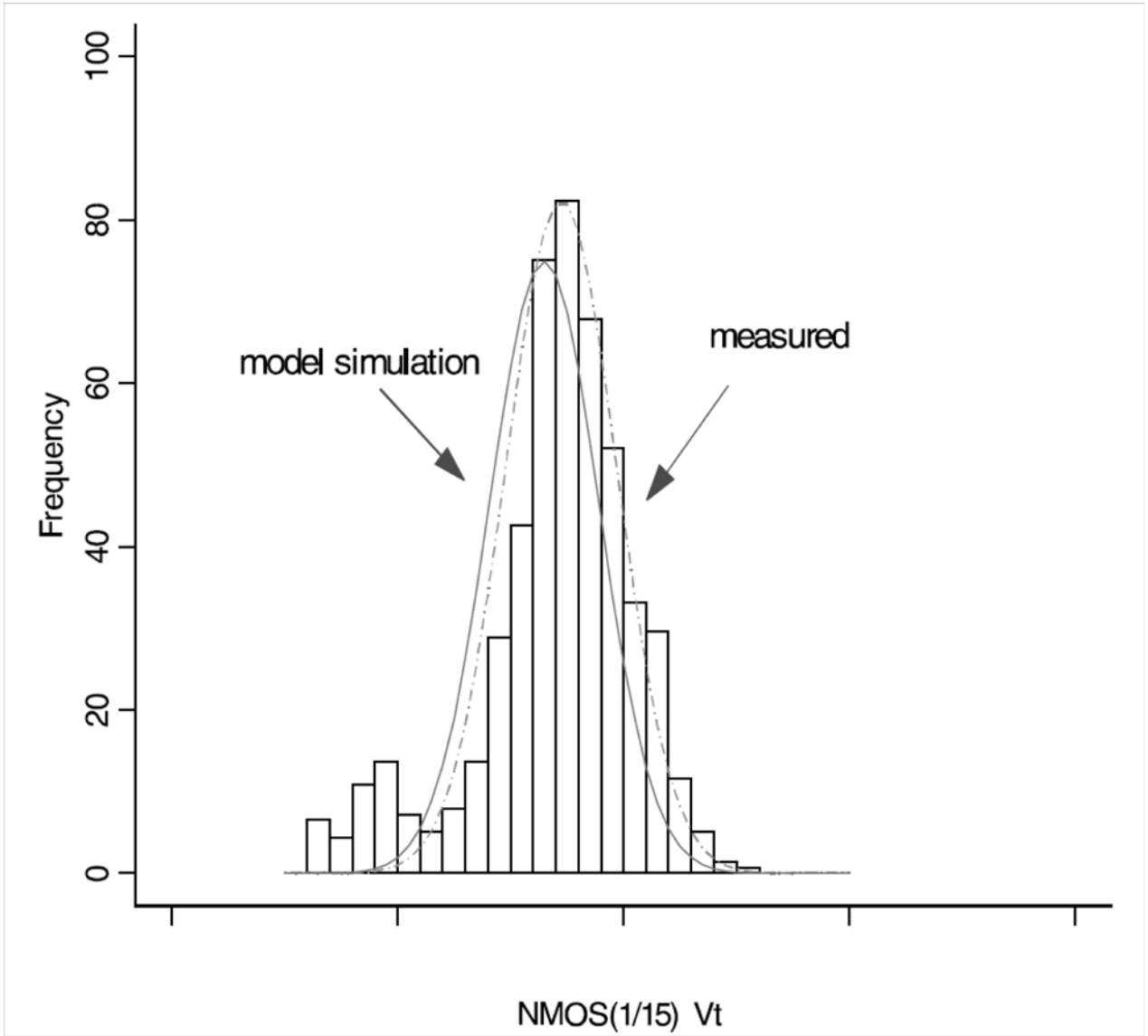
$$\frac{1}{\sqrt{2\,\pi}\,\sigma}\,e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}$$

Left graph:

$\sigma = 1$ and $\mu = 0$.

$x \in [-3, 3]$ captures 99.73% of the total area of curve.

Measurements vs Model



Source: Qiang Zhang et al., Solid-State Electronics 2001

Applying Statistics to Measurements

1. Standard deviation known a priori (experience) as 2 nm.
2. Make 15 measurement samples $\Rightarrow$ sample mean = 20 nm
3. Assume 95% confidence is sought for $\Rightarrow$ $\int_{-1.96}^{1.96} \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

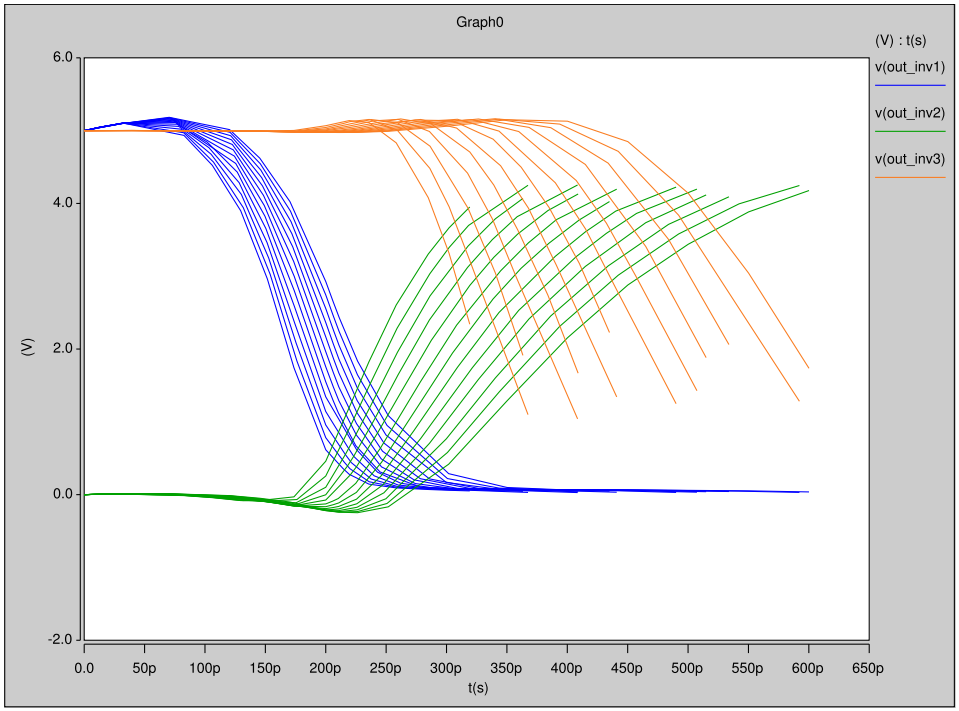
   since $x \in [-1.96\,\sigma, 1.96\,\sigma] \Rightarrow$ 95% of standard normal curve area.
4. The true mean is: $20 \pm 1.96 \frac{2}{\sqrt{15}}$.

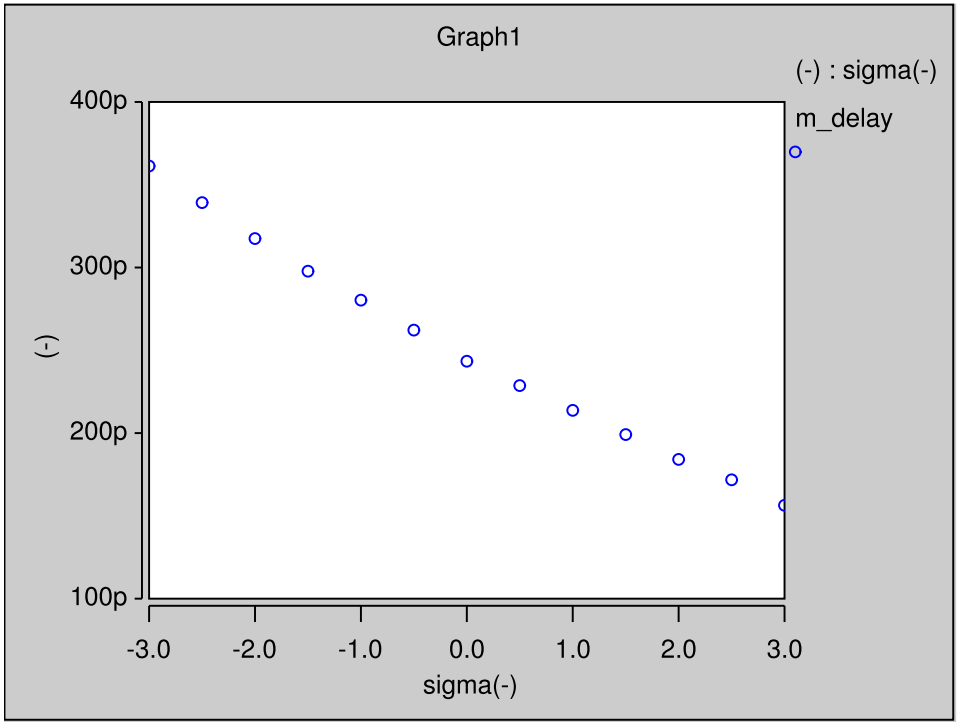
Some Sigma-Based Simulations Follow

```
$ Polysilicon Critical Dimensions
+ polycd=agauss(0,0.06u,1) xl='polycd-sigma*0.06u'

$ Active area Critical Dimensions
+ nactcd=agauss(0,0.3u,1) xwn='nactcd+sigma*0.3u'
+ pactcd=agauss(0,0.3u,1) xwp='pactcd+sigma*0.3u'

$ Gate Oxide Critical Dimensions
+ toxcd=agauss(200,10,1) tox='toxcd-sigma*10'

$ Threshold voltage variation
+ vtoncd=agauss(0,0.05v,1) delvton='vtoncd-sigma*0.05'
+ vtopcd=agauss(0,0.05v,1) delvtop='vtopcd+sigma*0.05'

$ Active layer resistivity variation
+ rshncd=agauss(50,8,1) rshn='rshncd-sigma*8'
+ rshpcd=agauss(150,20,1) rshp='rshpcd-sigma*20'
```

Inverter Chain Delay as Function of Sigma

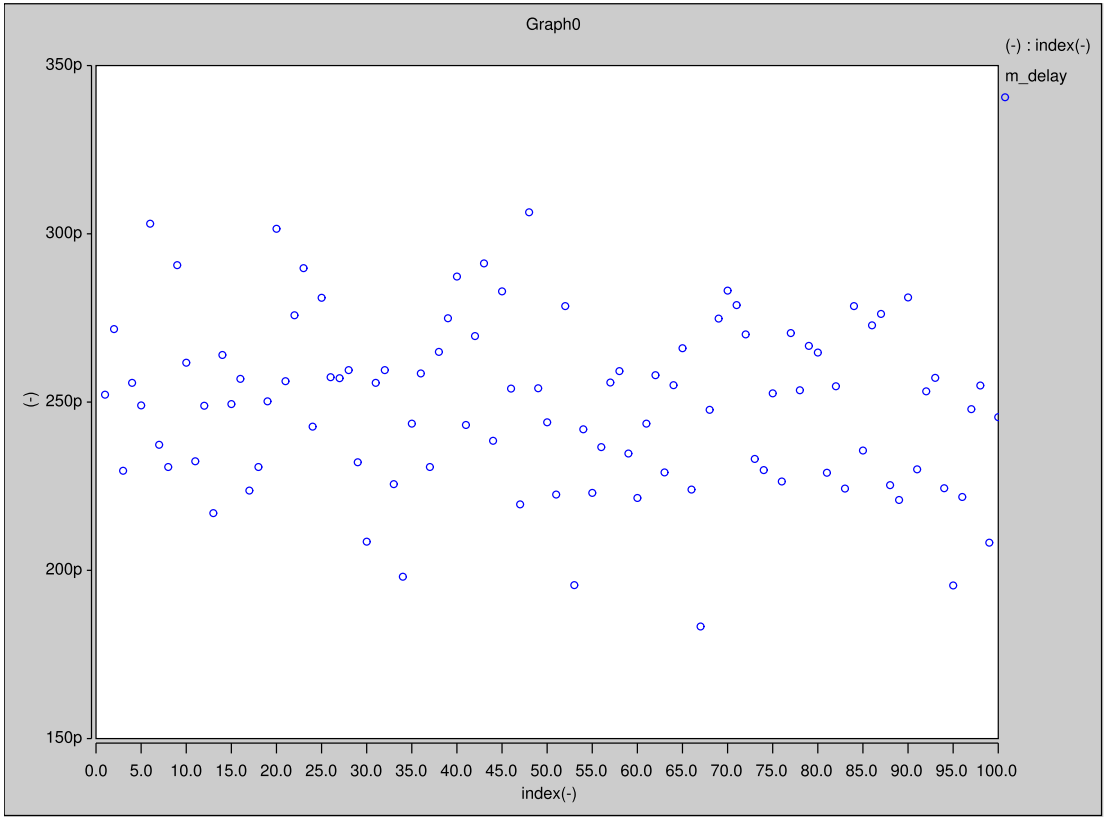
- Varying sigma in all parameters, -3σ, -2.5σ, ..., 2.5σ, 3σ.





Monte-Carlo Simulations

- In statistical simulations, each variable that follows the standard normal distribution is sampled randomly.



- Notice how the delay span reduces from 160ps-360ps.

Corners, Sigma, and Monte Carlo

- Simulations using process corners give pessimistic view of performance and power, but they have low run time.
- Monte Carlo simulation is the most accurate method, but it is very time consuming.

Power, Energy and Variability: Conclusion

- Power and energy analysis.
  - Depends on implementation, operating conditions and test vectors, thus, very complex.
  - Combination of simulation and static analysis; identify where is the weakness in terms of accuracy. For example, do you need detailed implementation data if you have no use case information?
- Variability.
  - Impacts delay and power dissipation.
  - Simulation methods like Monte Carlo work only for small circuits.