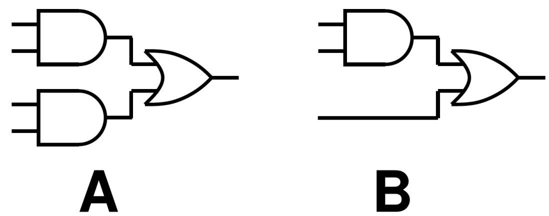
# ESE535:
# Electronic Design Automation

Day 23: April 10, 2013

Statistical Static Timing Analysis



## Delay PDFs? (2a)






## Today

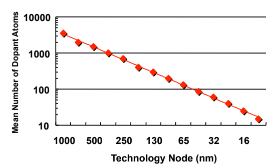
- Sources of Variation
- Limits of Worst Case
- Optimization for Parametric Yield
- Statistical Analysis

Behavioral
(C, MATLAB, …)
↓ Arch. Select, Schedule
RTL
↓ FSM assign, Two-level, Multilevel opt., Covering, Retiming
Gate Netlist
↓ Placement, Routing
Layout
↓
Masks



## Central Problem

- As our devices approach the atomic scale, we must deal with statistical effects governing the placement and behavior of individual atoms and electrons.




- Transistor critical dimensions
  - Atomic discreteness
  - Subwavelength litho
  - Etch/polish rates
  - Focus
- Number of dopants
- Dopant Placement



## Oxide Thickness

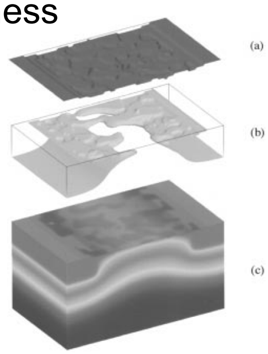


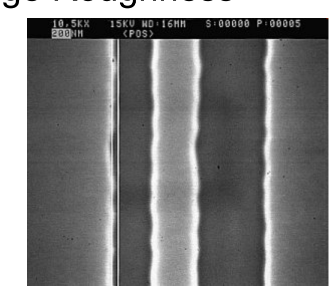

Fig. 1. (a) Typical profile of the random Si/SiO$_2$ interface in a 30 × 30 nm$^2$ MOSFET, followed by (b) an equiconcentration contour obtained from DG simulations, and (c) the potential distribution.

[Asenov et al. TRED 2002]



## Line Edge Roughness

- 1.2μm and 2.4μm lines



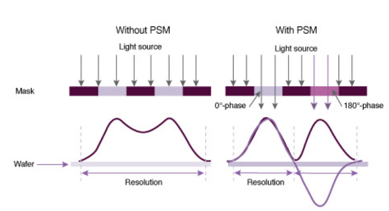

From: http://www.microtechweb.com/2d/lw_pict.htm

# Light

- What is wavelength of visible light?



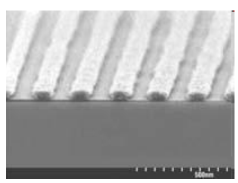
# Phase Shift Masking




Source
http://www.synopsys.com/Tools/Manufacturing/MaskSynthesis/PSMCreate/Pages/default.aspx



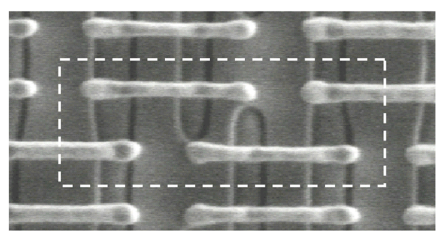
# Line Edges (PSM)



Source:
http://www.solid-state.com/display_article/122066/5/none/none/Feat/Developments-in-materials-for-157nm-photoresists



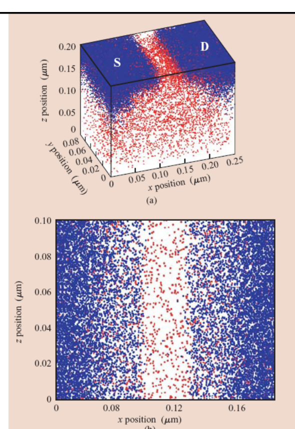
# Intel 65nm SRAM (PSM)



Source:
http://www.intel.com/technology/itj/2008/v12i2/5-design/figures/Figure_5_lg.gif



# Statistical Dopant Placement

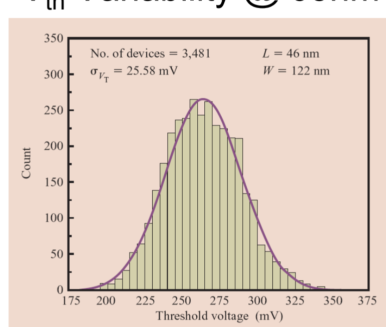





[Bernstein et al, IBM JRD 2006]



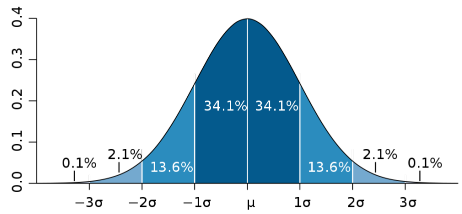
# $V_{th}$ Variability @ 65nm






[Bernstein et al, IBM JRD 2006]

# Gaussian Distribution




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



# ITRS 2005 Variation (3σ)

*Table 18a Design-for-Manufacturability—Near-term Years*

| *Year of Production* | *2005* | *2006* | *2007* | *2008* | *2009* | *2010* | *2011* | *2012* | *2013* | *Driver* |
|---|---|---|---|---|---|---|---|---|---|---|
| *DRAM ½ Pitch (nm) (contacted)* | *80* | *70* | *65* | *57* | *50* | *45* | *40* | *36* | *32* | |
| Mask cost (\$m) from publicly available data | **1.5** | **2.2** | **3.0** | **4.5** | **6.0** | **9.0** | **12.0** | **18.0** | **24.0** | *SOC* |
| % $V_{dd}$ Variability % variability seen at on-chip circuits | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | *SOC* |
| % $V_{th}$ va… Do… | | | | | | | | | | |
| % $V_{th}$ va… Inc… | | | | | | | | | | |
| % CD va… CD… | | | | | | | | | | |
| % circui… circ… | | | | | | | | | | |
| % circui… circ… | | | | | | | | | | |

*Table 18b Design-for-Manufacturability—Long-term Years*

| *Year of Production* | *2014* | *2015* | *2016* | *2017* | *2018* | *2019* | *2020* | *Driver* |
|---|---|---|---|---|---|---|---|---|
| *DRAM ½ Pitch (nm)(contacted)* | *28* | *25* | *22* | *20* | *18* | *16* | *14* | |
| Mask cost (\$m) from publicly available data | **36.0** | **48.0** | **72.0** | **96.0** | **144.0** | **192.0** | **288.0** | *SOC* |
| % $V_{dd}$ Variability % variability seen at on-chip circuits | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | *SOC* |
| % $V_{th}$ variability Doping Variability impact on VTH | **81%** | **81%** | **81%** | **81%** | **112%** | **112%** | **112%** | *SOC* |
| % $V_{th}$ variability Includes all sources | **81%** | **81%** | **81%** | **81%** | **112%** | **112%** | **112%** | *SOC* |
| % CD variability CD for now, might add doping later | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | **10%** | *SOC* |
| % circuit performance variability circuit comprising gates and wires | **58%** | **61%** | **62%** | **65%** | **66%** | **69%** | **69%** | *SOC* |
| % circuit power variability circuit comprising gates and wires | **59%** | **60%** | **60%** | **61%** | **61%** | **62%** | **62%** | *SOC* |



# Example: $V_{th}$

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

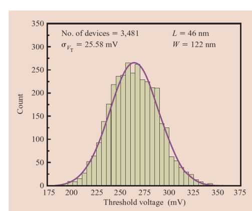
$$I_{DS} = \frac{\mu_n C_{OX}}{2}\left(\frac{W}{L}\right)\left[\left(V_{GS} - V_T\right)^2\right]$$

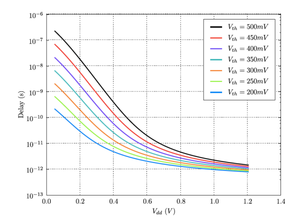
$$I_{DS} = I_S\left(\frac{W}{L}\right)e^{\left(\frac{V_{GS}}{nkT/q}\right)}\left(1 - e^{-\left(\frac{V_{DS}}{kT/q}\right)}\right)\left(1 + \lambda V_{DS}\right)$$



# Impact Performance

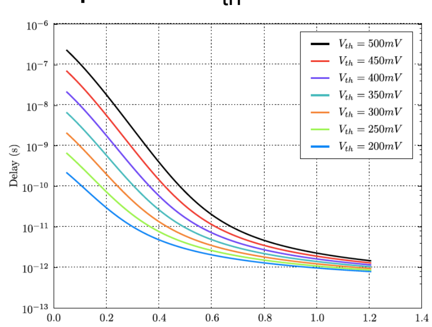
- $V_{th}$ → $I_{ds}$ → Delay ($R_{on}$ * $C_{load}$)



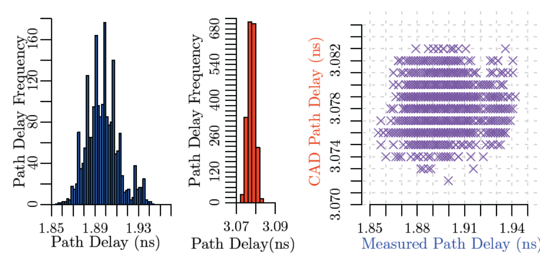



# Impact of $V_{th}$ Variation





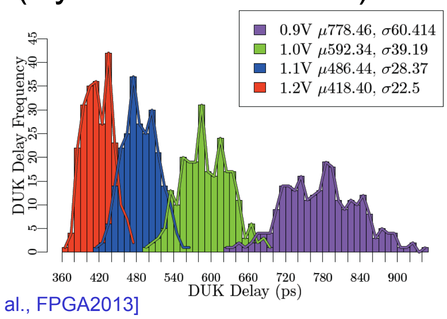

# Variation in Current FPGAs




(a) Measured Delay (b) CAD Delay (c) Correlation

[Gojman et al., FPGA2013]

## Reduce Vdd (Cyclone IV 60nm LP)




[Gojman et al., FPGA2013]



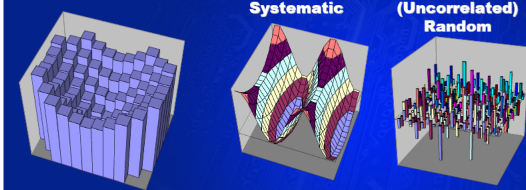
## Scale of Variations

Die-to-Die (D2D) Variations

Within-Die (WID) Variations

Systematic

(Uncorrelated) Random

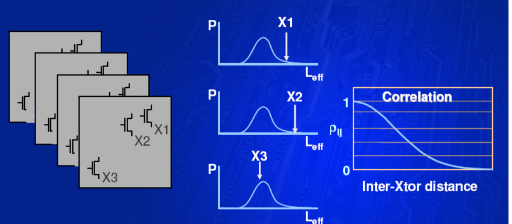


Wafer Scale | Die Scale | Feature Scale

Source: Noel Menezes, Intel ISPD2007



## Nature of correlated variation




- CDs of transistors that are close track
  - Tracking diminishes with distance

Source: Noel Menezes, Intel ISPD2007

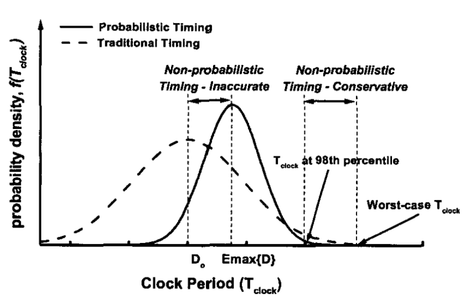


## Old Way

- Characterize gates by corner cases
  - Fast, nominal, slow
- Add up corners to estimate range
- Preclass:
  - Slow corner: 1.1
  - Nominal: 1.0
  - Fast corner: 0.9



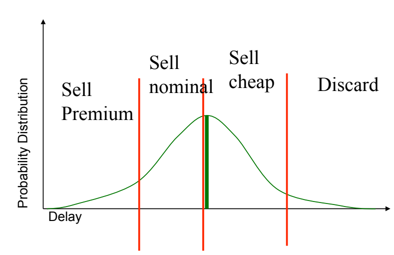
## Corners Misleading




[Orshansky+Keutzer DAC 2002]

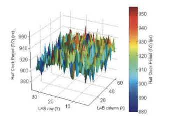


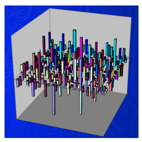
## Parameteric Yield

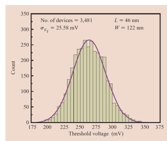


## Phenomena 1: Path Averaging

- $T_{path} = t_0+t_1+t_2+t_3+\ldots t_{(d-1)}$
- $T_i$ – iid random variables
  - Mean $\tau$
  - Variance $\sigma$
- Tpath
  - Mean $d \times \tau$
  - Variance = $\sqrt{d} \times \sigma$









## Sequential Paths

- $T_{path} = t_0+t_1+t_2+t_3+\ldots t_{(d-1)}$
- $T_{path}$
  - Mean $d \times \tau$
  - Variance = $\sqrt{d} \times \sigma$
- 3 sigma delay on path: $d \times \tau + 3\sqrt{d} \times \sigma$
  - Worst case per component would be: $d \times (\tau + 3\sigma)$
  - Overestimate d vs. $\sqrt{d}$



## SSTA vs. Corner Models

- STA with corners predicts 225ps
- SSTA predicts 162ps at $3\sigma$
- SSTA reduces pessimism by 28%

[Slide composed by Nikil Mehta]

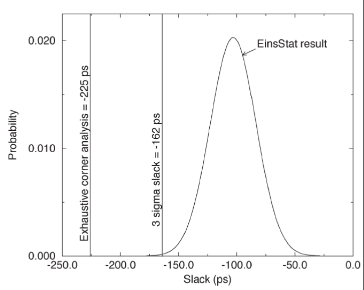


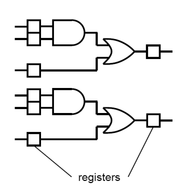

Fig. 11. EinsStat result on industrial ASIC design for early mode slacks.

Source: IBM, TRCAD 2006



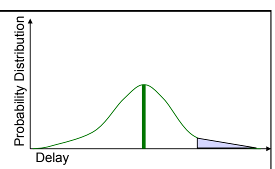
## Phenomena 2: Parallel Paths

- Cycle time limited by slowest path
- $T_{cycle} = \max(T_{p0}, T_{p1}, T_{p2}, \ldots T_{p(n-1)})$
- $P(T_{cycle}<T_0) = P(T_{p0}<T_0) \times P(T_{p1}<T_0)\ldots$
- $= [P(T_p<T_0)]^n$
- $0.5 = [P(T_p<T_{50})]^n$
- $P(T_p<T_{50}) = (0.5)^{(1/n)}$





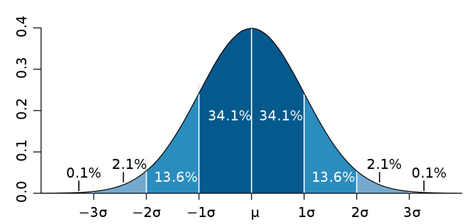

## System Delay




- $P(T_p<T_{50}) = (0.5)^{(1/n)}$
  - $N=10^8$ ➔ 0.999999993
    - $1-7\times10^{-9}$
  - $N=10^{10}$ ➔ 0.99999999993
    - $1-7\times10^{-11}$

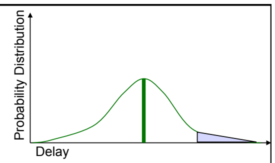


## Gaussian Distribution




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

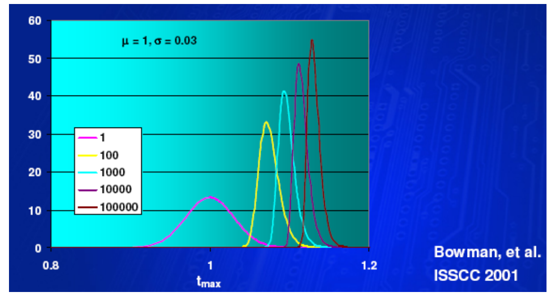
# System Delay




- $P(T_p<T_{50}) = (0.5)^{(1/n)}$
  - $N=10^8 \rightarrow 0.999999993$
    - $1-7\times10^{-9}$
  - $N=10^{10} \rightarrow 0.99999999993$
    - $1-7\times10^{-11}$
- For 50% yield want
  - 6 to 7 $\sigma$
  - $T_{50}=T_{mean}+7\sigma_{path}$

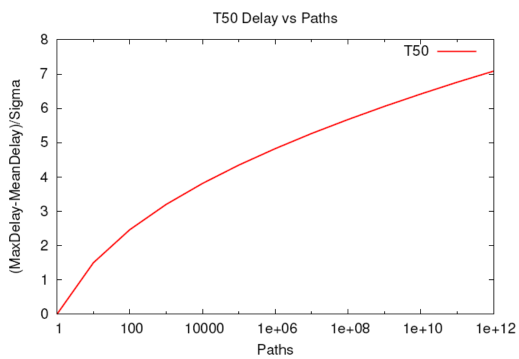


# System Delay



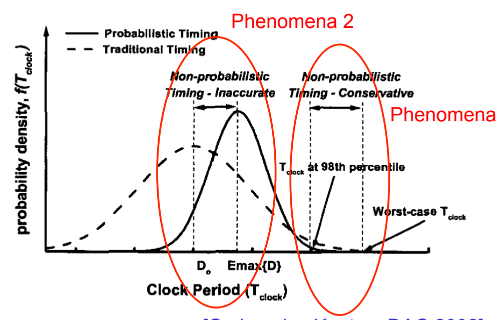



# System Delay



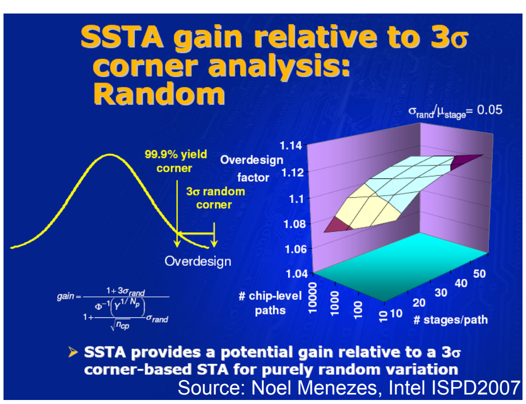



# Corners Misleading




[Orshansky+Keutzer DAC 2002]

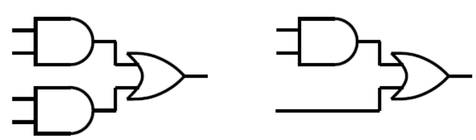






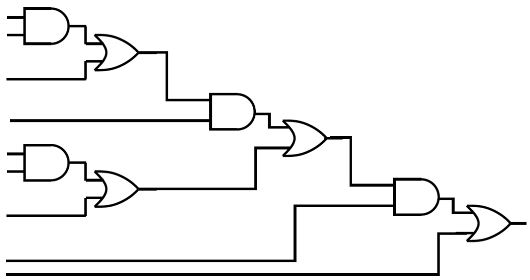

# But does worst-case mislead?

- STA with worst-case says these are equivalent:



# But does worst-case mislead?

- STA Worst case delay for this?

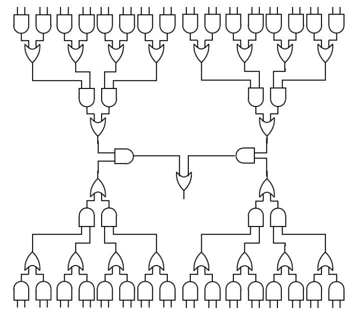


Penn ESE535 Spring 2013 -- DeHon

37

# But does worst-case mislead?

- STA Worst case delay for this?

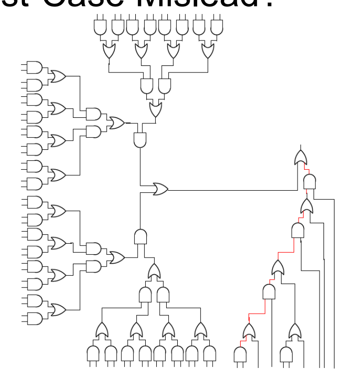


Penn ESE535 Spring 2013 -- DeHon

38

# Does Worst-Case Mislead?

- Delay of off-critical path may matter
- Best case delay of critical path?
- Wost-case delay of non-critical path?



Penn ESE535 Spring 2013 -- DeHon

39

# What do we need to do?

- Ideal:
  - Compute PDF for delay at each gate
  - Compute delay of a gate as a PDF from:
    - PDF of inputs
    - PDF of gate delay

Penn ESE535 Spring 2013 -- DeHon

40

# Delay Calculation

Day 22

AND rules

| $i_1 \rightarrow$ $i_2 \downarrow$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 0<br>$MIN(l_1, l_2) + d$<br>$MIN(u_1, u_2) + d$ | 0<br>$l_2 + d$<br>$u_2 + d$ | 0<br>$MIN(l_1, l_2) + d$<br>$u_2 + d$ |
| 1 | 0<br>$l_1 + d$<br>$u_1 + d$ | 1<br>$MAX(l_1, l_2) + d$<br>$MAX(u_1, u_2) + d$ | 2<br>$l_1 + d$<br>$MAX(u_1, u_2) + d$ |
| 2 | 0<br>$MIN(l_1, l_2) + d$<br>$u_1 + d$ | 2<br>$l_2 + d$<br>$MAX(u_1, u_2) + d$ | 2<br>$MIN(l_1, l_2) + d$<br>$MAX(u_1, u_2) + d$ |

Penn ESE535 Spring 2013 -- DeHon

41

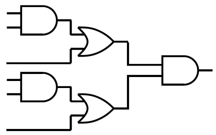
# What do we need to do?

- Ideal:
  - compute PDF for delay at each gate
  - Compute delay of a gate as a PDF from:
    - PDF of inputs
    - PDF of gate delay
  - Need to compute for distributions
    - SUM
    - MAX (maybe MIN)

Penn ESE535 Spring 2013 -- DeHon

42

## For Example



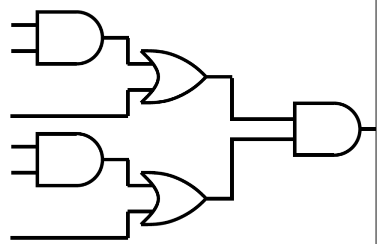
- Consider entry:
  - $MAX(u_1,u_2)+d$

| $i_1 \rightarrow$ / $i_2 \downarrow$ | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 0 $MIN(l_1,l_2)+d$ $MIN(u_1,u_2)+d$ | 0 $l_2+d$ $u_2+d$ | 0 $MIN(l_1,l_2)+d$ $u_2+d$ |
| 1 | 0 $l_1+d$ $u_1+d$ | 1 $MAX(l_1,l_2)+d$ $MAX(u_1,u_2)+d$ | 2 $l_1+d$ $MAX(u_1,u_2)+d$ |
| 2 | 0 $MIN(l_1,l_2)+d$ $u_1+d$ | 2 $l_2+d$ $MAX(u_1,u_2)+d$ | 2 $MIN(l_1,l_2)+d$ $MAX(u_1,u_2)+d$ |



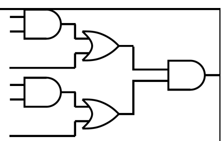
## MAX

- Compute MAX of two input distributions.





## SUM



- Add that distribution to gate distribution.



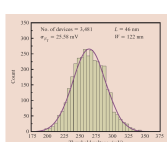
## Continuous

- Can roughly carry through PDF calculations with Guassian distributions rather than discrete PDFs



## Dealing with PDFs

- Simple model assume all PDFs are Gaussian
  - Model with mean, $\sigma$
  - Imperfect
    - Not all phenomena are Gaussian
    - Sum of Gaussians is Gaussian
    - Max of Gaussians is **not** a Gaussian

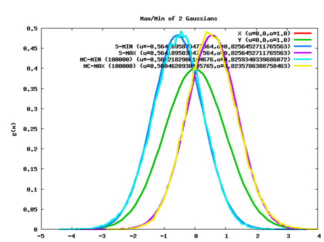




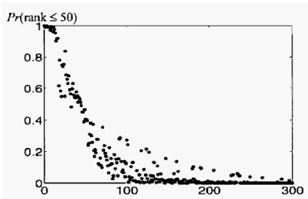
## MAX of Two Identical Gaussians

- Max of Gaussians is not a Gaussian
- Can try to approximate as Guassian
- Given two identical Gaussians A and B with $\mu$ and $\sigma$
- Plug into equations
- $E[MAX(A,B)] = \mu + \sigma/(\pi)^{1/2}$
- $VAR[MAX(A,B)] = \sigma^2 - \sigma/\pi$

[Source: Nikil Mehta]



## Probability of Path Being Critical

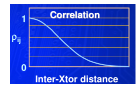


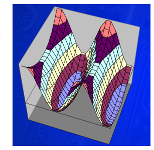

Path Rank (from deterministic timing analysis)
**Figure 1 Probability that a path shows up in top 50 paths (Data from Monte Carlo simulation of a 90nm microprocessor block )**

[Source: Intel DAC 2005]

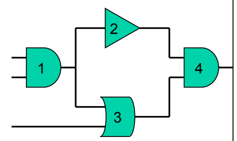


## More Technicalities



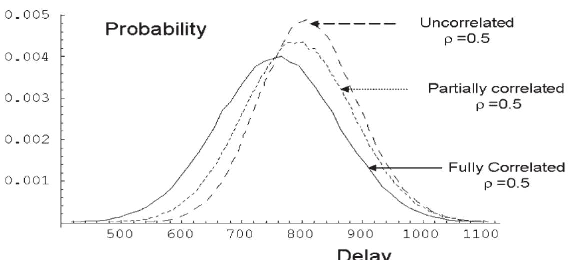
- Correlation
  - Physical on die
  - In path (reconvergent fanout)
    - Makes result conservative
      - Gives upper bound
      - Can compute lower




Graphics from: Noel Menezes (top) and Nikil Mehta (bottom)



## Max of Gaussians with Correlation



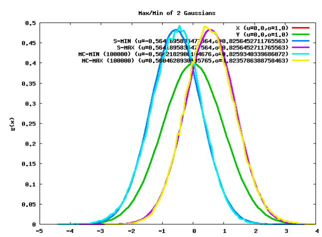

- Max of identical Gaussians

 [Blaauw et al. TRCAD v27n4p589] 

## MAX of Two Identical Gaussians

- Given two identical Gaussians A and B with $\mu$ and $\sigma$
- Plug into equations
- $E[MAX(A,B)] = \mu + \sigma/(\pi)^{1/2}$
- $VAR[MAX(A,B)] = \sigma^2 - \sigma/\pi$

[Source: Nikil Mehta]



Extreme of correlated: is just the input Gaussian



## SSTA vs. Monte Carlo Verification Time

TABLE II
MONTE CARLO VERSUS EinsStat COMPARISON

| Test case | Gates | EinsStat CPU | Monte Carlo | | |
|---|---|---|---|---|---|
| | | | Samples | Sequential CPU dd:hh:mm:ss | Parallel CPU dd:hh:mm:ss |
| 1 | 18 | 1 sec. | 100000 | 5:57 | N/A |
| 2 | 3042 | 2 sec. | 100000 | 2:01:15:10 | 2:46:55 |
| 3 | 11937 | 7 sec. | 10000 | 0:20:33:40 | 51:05 |
| 4 | 70216 | 59 sec. | 10000 | N/A | 4:36:12 |

Source: IBM, TRCAD 2006



## Using SSTA in FPGA CAD

[Slide composed by Nikil Mehta]

- Le Hei
  - FPGA2007
  - SSTA Synthesis, Place, Route
- Kia
  - FPGA2007
  - Route with SSTA

| process variation settings ($3\sigma$) | | | | | | |
|---|---|---|---|---|---|---|
| global | 5.0% | 10.0% | 15.0% | 20.0% | 25.0% | 30.0% |
| spatial | 5.0% | 10.0% | 15.0% | 20.0% | 25.0% | 30.0% |
| local | 3.0% | 6.0% | 9.0% | 12.0% | 15.0% | 18.0% |
| deterministic flow | | | | | | |
| Tmean (ns) | 21.7 | 22.9 | 24.4 | 26.2 | 28.2 | 30.2 |
| Tsigma (ns) | 1.8 | 3.4 | 5.0 | 6.4 | 7.8 | 9.2 |
| stochastic flow | | | | | | |
| Tmean (ns) | 20.3 (-6.5%) | 21.5 (-6.2%) | 23.0 (-5.8%) | 24.8 (-5.5%) | 26.7 (-5.3%) | 28.6 (-5.1%) |
| Tsigma (ns) | 1.6 (-6.6%) | 3.1 (-7.5%) | 4.6 (-8.1%) | 5.9 (-8.2%) | 7.2 (-8.1%) | 8.5 (-8.0%) |

**Table 6: Comparison of mean delay and standard deviation between deterministic and stochastic flows under various process variation assumptions (based on the geometric mean of 20 MCNC designs).**

| Circuit | Delay Impr.(%) |
|---|---|
| ex5p | 6.58 |
| alu4 | 1.50 |
| misex3 | 5.76 |
| apex2 | 3.24 |
| apex4 | 2.57 |
| pdc | 4.74 |
| seq | 4.37 |
| des | 3.73 |
| spla | 4.82 |
| ex1010 | 1.83 |
| frisc | 2.84 |
| elliptic | 0.17 |
| bigkey | 0.35 |
| s298 | 7.10 |
| tseng | 5.93 |
| diffeq | 4.16 |
| dsip | 7.37 |
| s38417 | 7.56 |
| s38584.1 | 5.43 |
| clma | -1.17 |
| **Mean** | **3.95** |

# Impact of SSTA in High-Level Synthesis

| Design (#ops) | #ALU, #MUL | $p_{clk}$ (ns) | Latency(cycles) | | Reduction | Run time(s) |
|---|---|---|---|---|---|---|
| | | | LS[15] | HLS-tv(Y) | | |
| DIFF (18) | 3, 3 | 3.5 | 32 | 28 (94.5%) | 12.5% | 928 |
| | | 4.0 | 29 | 24 (90.7%) | 17.2% | 930 |
| | | 4.5 | 26 | 22 (93.2%) | 15.4% | 637 |
| LATT (22) | 3, 2 | 3.5 | 47 | 36 (94.3%) | 23.4% | 2122 |
| | | 4.0 | 42 | 32 (94.3%) | 23.8% | 3325 |
| | | 4.5 | 37 | 30 (90.2%) | 18.9% | 1207 |
| AR (28) | 2, 3 | 3.5 | 57 | 45 (93.9%) | 21.1% | 1241 |
| | | 4.0 | 51 | 40 (93.9%) | 21.6% | 1534 |
| | | 4.5 | 45 | 36 (90.8%) | 20.0% | 680 |
| EWF (34) | 2, 3 | 3.5 | 46 | 37 (93.6%) | 19.6% | 157 |
| | | 4.0 | 42 | 34 (93.6%) | 19.0% | 367 |
| | | 4.5 | 38 | 33 (91.5%) | 13.2% | 113 |
| avg. | | | | **- (92.9%)** | **18.8%** | |

- Scheduling and provisioning
  - ALU/MUL $\sigma$=5% $t_{nominal}$

[Jung&Kim ICCAD2007]



# Summary

- Nanoscale fabrication is a statistical process
- Delays are PDFs
- Assuming each device is worst-case delay is too pessimistic
  - Wrong prediction about timing
  - Leads optimization in wrong direction
- Reformulate timing analysis as statistical calculation
- Estimate the PDF of circuit delays
- Use this to drive optimizations



# Big Ideas:

- Coping with uncertainty
- Statistical Reasoning and Calculation



# Admin

- Reading for Monday on blackboard
- Milestone Mondays