ESE535: Electronic Design Automation

Day 23: April 22, 2009

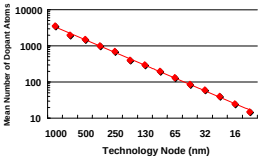
Statistical Static Timing Analysis

Today

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

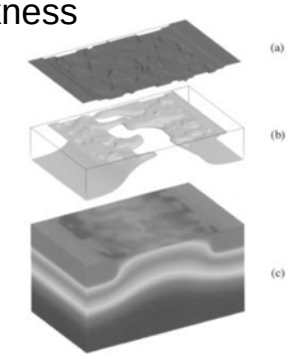
Central Problem

- 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

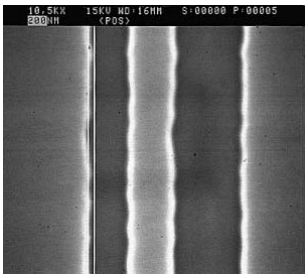


[Asenov et al. TRED 2002]

Fig. 1. (a) Typical profile of the random Si/SiO_2 interface in a 30×30 nm² MOSFET, followed by (b) an equiconcentration contour obtained from DG simulations, and (c) the potential distribution.

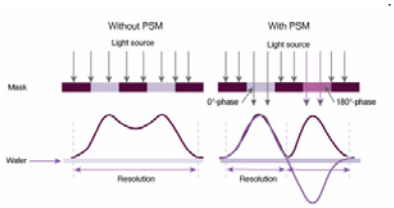
Line Edge Roughness

- 1.2μm and 2.4μm lines



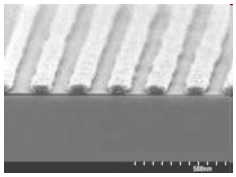
From: http://www.microtechweb.com/2d/lw_pict.htm

Phase Shift Masking



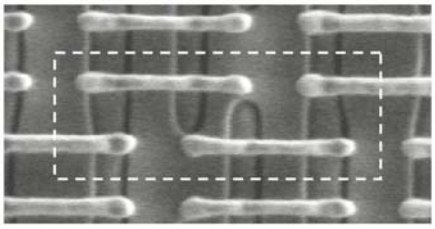
Source http://www.synopsys.com/Tools/Manufacturing/MaskSynthesis/PSMCreate/Pages/default.aspx

Line Edges (PSM)



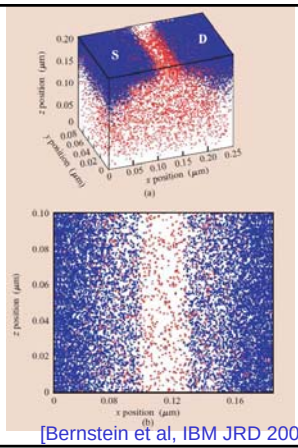
Source:
http://www.solid-state.com/display_article/122066/5/none/none/Feat/Developments-in-materials-for-157nm-photoresists

Intel 65nm SRAM (PSM)



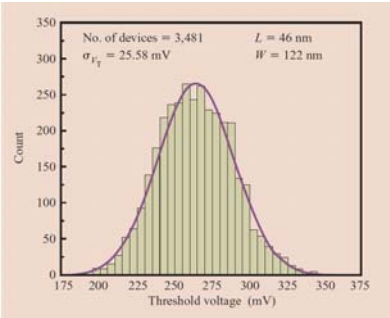
Source:
http://www.intel.com/technology/itj/2008/v12i2/5-design/figures/Figure_5_lggif

Statistical Dopant Placement



[Bernstein et al, IBM JRD 2006]

V_{th} Variability @ 65nm

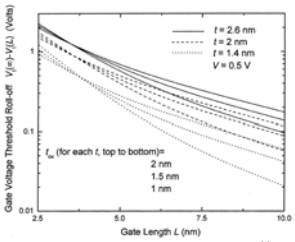


[Bernstein et al, IBM JRD 2006]

Parameter Variation

$$I_{ds} = \left(\frac{\mu C_{OX}}{2}\right)\left(\frac{W}{L}\right)\left(V_{gs} - V_{th}\right)^2$$

- Parameters will vary from device-to-device on the die
  - Include transistor threshold (V_{th})



Sverdlov et al. TRED v50n9p1926 2003

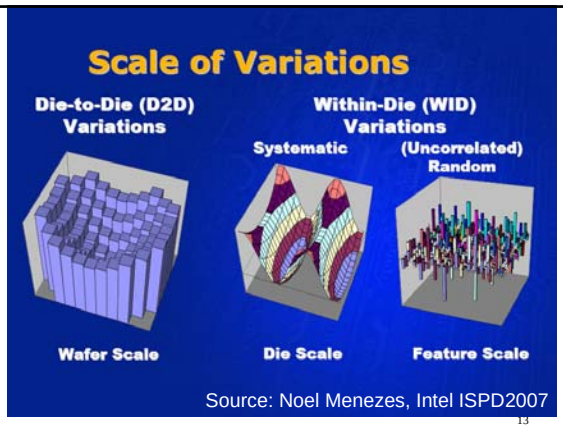
ITRS 2005 Variation

Table 15a Design-for-Manufacturability—Near-term Years

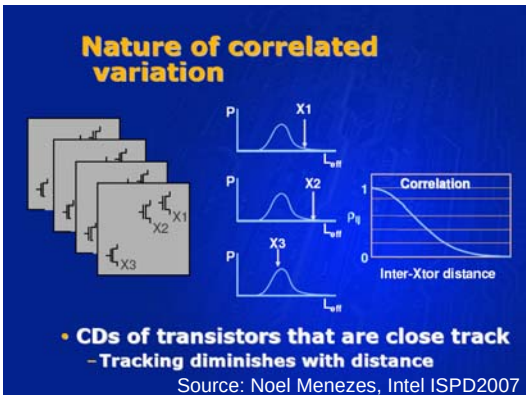
Year of Production	2005	2006	2007	2008	2009	2010	2011	2012	2013	Driver
DRAM ½ Pitch (nm) (contacted)	80	70	63	57	50	45	40	36	32	
Mask cost ($m) from publicly available data	1.5	2.2	3.0	4.5	6.0	9.0	12.0	18.0	24.0	SOC
% V_{dd} Variability	10%	10%	10%	10%	10%	10%	10%	10%	10%	SOC

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



Source: Noel Menezes, Intel ISPD2007

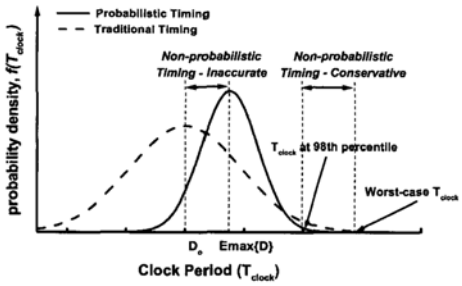


Source: Noel Menezes, Intel ISPD2007

Old Way

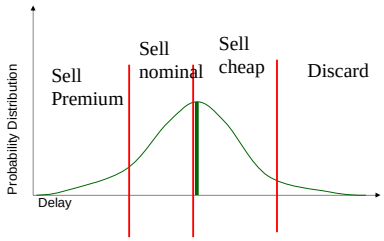
- Characterize gates by corner cases
  - Fast, nominal, slow
- Add up corners to estimate range

Corners Misleading



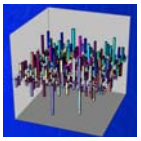
[Orshansky+Keutzer DAC 2002]

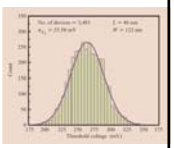
Parameteric Yield



Phenomena 1: Path Averaging

- $T_{path} = t_0+t_1+t_2+t_3+\ldots t_{(d-1)}$
- T_i – iid random variables
  - Mean τ
  - Variance σ
- 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 3σ
- SSTA reduces pessimism by 28%

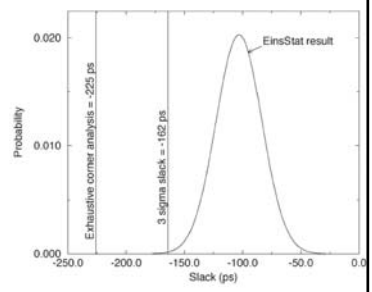


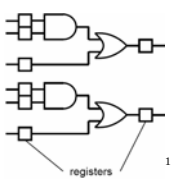
Fig. 11. EinsStat result on industrial ASIC design for early mode slacks.

[Slide composed by Nikil Mehta]

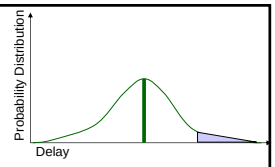
Source: IBM, TRCAD 2006

Phenomena 2: Parallel Paths

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

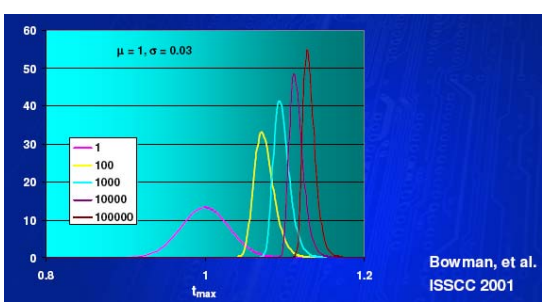


System Delay

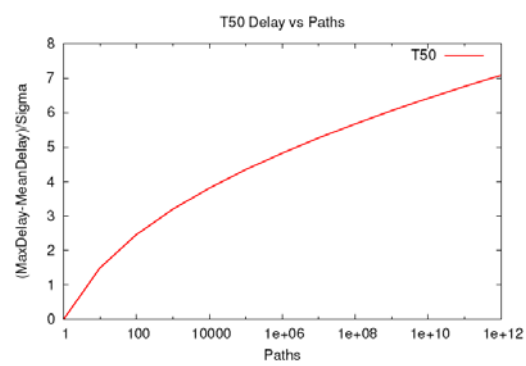


- $P(T_p<T_{50}) = (0.5)^{(1/n)}$
  - $N=10^8$ ➔ 0.999999993
    - $1-7\times10^{-9}$
  - $N=10^{10}$ ➔ 0.99999999993
    - $1-7\times10^{-11}$
- For 50% yield want
  - 6 to 7 σ
  - $T_{50}=T_{mean}+7\sigma_{path}$

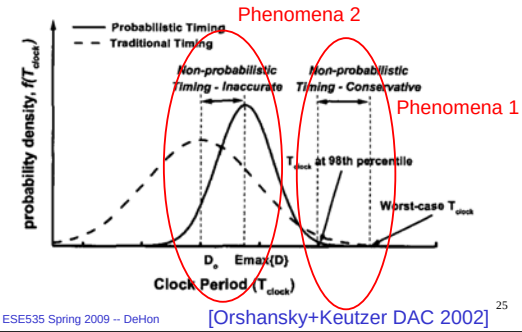
System Delay



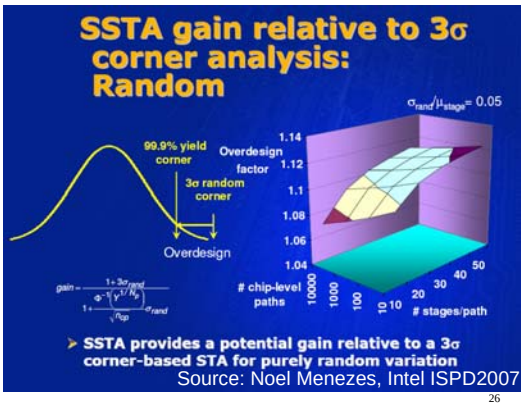
System Delay



Corners Misleading



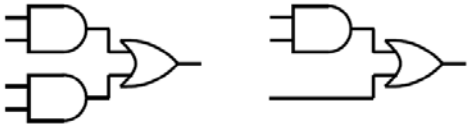
[Orshansky+Keutzer DAC 2002]



Source: Noel Menezes, Intel ISPD2007

But does worst-case mislead?

- STA with worst-case says these are equivalent:



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

Delay Calculation

Day 15

AND rules

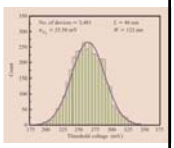
$i_1 \rightarrow$ $i_2 \downarrow$	0	1	2
0	0 $MIN(l_1,l_2)+d$ $MIN(u_1,u_2)+d$	0 l_2+d u_2+d	0 $MIN(l_1,l_2)+d$ u_2+d
1	0 l_1+d u_1+d	1 $MAX(l_1,l_2)+d$ $MAX(u_1,u_2)+d$	2 l_1+d $MAX(u_1,u_2)+d$
2	0 $MIN(l_1,l_2)+d$ u_1+d	2 l_2+d $MAX(u_1,u_2)+d$	2 $MIN(l_1,l_2)+d$ $MAX(u_1,u_2)+d$

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)

Dealing with PDFs

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



Sum of Gaussians

- Two Guassians
  - A,σ_A and B, σ_B
  - SUM = (A+B), $\sqrt{\sigma_A^2+\sigma_B^2}$
  - If identical
    - SUM = 2A, $\sigma_A\sqrt{2}$

Tightness Probability (toward max)

$$\phi(x) \equiv \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$ → Gaussian PDF with zero mean and unit variance

$$\Phi(y) \equiv \int_{-\infty}^{y} \phi(x)\,dx$$ → Prob(X<y)

$$\theta \equiv \left(\sigma_A^2+\sigma_B^2-2\rho\sigma_A\sigma_B\right)^{\frac{1}{2}}$$ → Standard deviation of SUM(A,B) (if A & B uncorrelated)

$$T_A = \int_{-\infty}^{\infty} \frac{1}{\sigma_A}\phi\left(\frac{x-a_0}{\sigma_A}\right)\Phi\left(\frac{\left(\frac{x-b_0}{\sigma_B}\right)-\rho\left(\frac{x-a_0}{\sigma_A}\right)}{\sqrt{1-\rho^2}}\right)dx$$

$$= \Phi\left(\frac{a_0-b_0}{\theta}\right).$$ → Prob(X<(A_{nom} – B_{nom})/μ_{A+B}))

[Source: Nikil Mehta]

MAX of Two Gaussians

$$E[\max(A,B)] = a_0T_A + b_0(1-T_A) + \theta\phi\left[\frac{a_0-b_0}{\theta}\right]$$

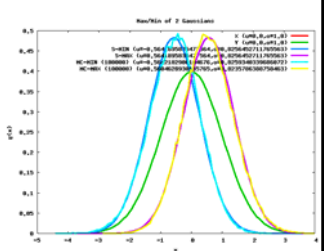
$$\begin{aligned}\mathrm{var}[\max(A,B)] = &\left(\sigma_A^2+a_0^2\right)T_A + \left(\sigma_B^2+b_0^2\right)(1-T_A)\\ &+(a_0+b_0)\theta\phi\left(\frac{a_0-b_0}{\theta}\right)\\ &-\{E[\max(A,B)]\}^2. \qquad (10)\end{aligned}$$

- Expected value
  - Weighted sum of means
  - Additional term which adds fraction of σ of SUM(A,B)
- Variance
  - Weighted sum of variance
  - Some other terms?

[Source: Nikil Mehta]

MAX of Two Identical Gaussians

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

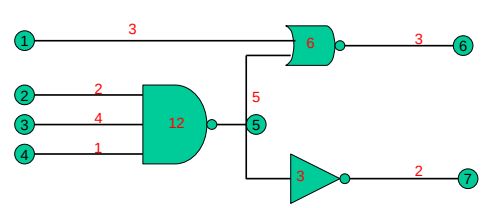


[Source: Nikil Mehta]

[Source: Nikil Mehta]

STA Example

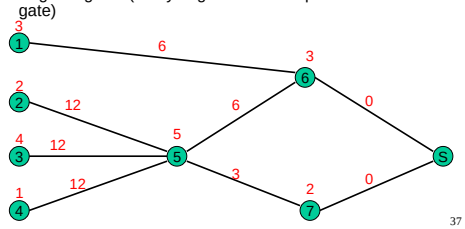
- Example circuit
  - Each component has a delay
  - Nets are numbered



[Source: Nikil Mehta]

STA Example

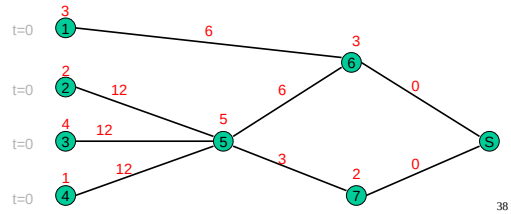
- Transform into a timing graph
  - Nodes = nets
  - Edges = gates (many edges can correspond to the same gate)



[Source: Nikil Mehta]

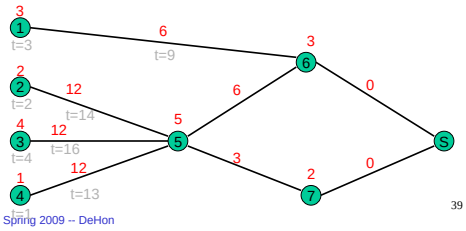
STA Example

- Goal is to compute arrival time on output



[Source: Nikil Mehta]

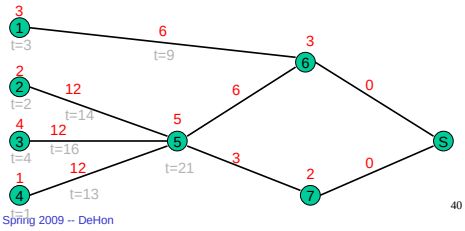
STA Example



[Source: Nikil Mehta]

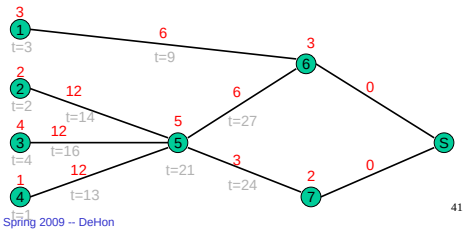
STA Example

- For nodes with multiple inputs
  - Arrival time = MAX(input arrival times)



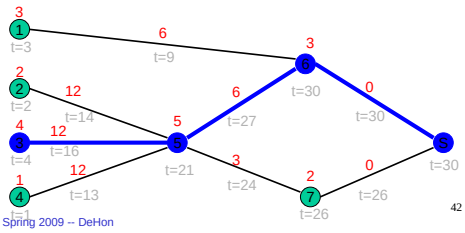
[Source: Nikil Mehta]

STA Example



[Source: Nikil Mehta]

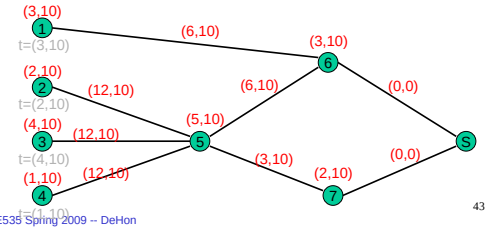
STA Example



[Source: Nikil Mehta]

SSTA Example

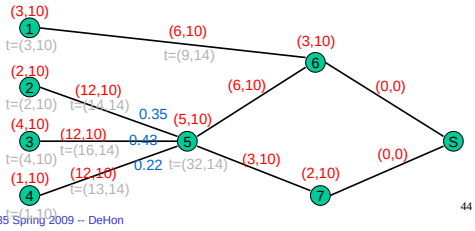
- Represent delay and arrival time statistically (μ,σ)
- Picking large variance (10) for all delays



[Source: Nikil Mehta]

SSTA Example

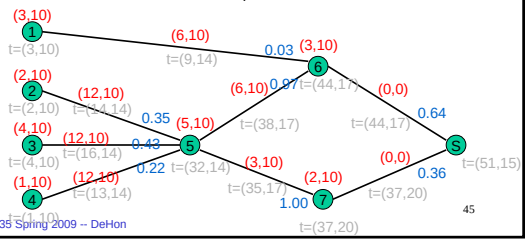
- Perform statistical SUM's
- Once we get to node 5, calculate tightness probabilities of input edges
  - Will allow us to perform statistical MAX
  - MAX is computed pairwise



[Source: Nikil Mehta]

SSTA Example

- Finish forward pass
  - Now, have statistical delay pdf of circuit
  - Normal distribution with μ=51 and σ=15

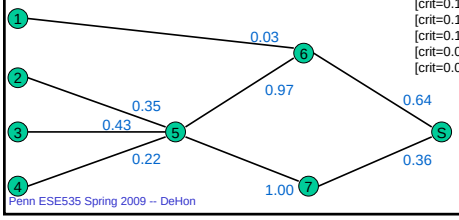


[Source: Nikil Mehta]

SSTA Example

- Also have statistical criticality of all paths
  - Criticality ´ Product of tightness probabilities along path
  - SSTA outputs list of paths in order of criticality
- On backward pass can calculate
  - Statistical slack
  - Statistical node/edge criticality

Printing critical paths (7/7) …
[crit=0.27](node3->sink)
[crit=0.21](node2->sink)
[crit=0.15](node3->sink)
[crit=0.13](node4->sink)
[crit=0.12](node2->sink)
[crit=0.07](node4->sink)
[crit=0.02](node1->sink)



Probability of Path Being Critical

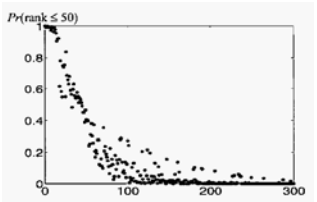
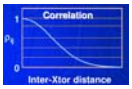
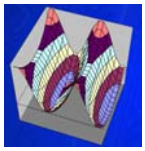


Figure 1 Probability that a path shows up in top 50 paths (Data from Monte Carlo simulation of a 90nm microprocessor block)

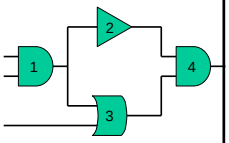
[Source: Intel DAC 2005]

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)

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 (3σ)						
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%)	21.3 (-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

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

Admin

- Reading for Monday online
- Online Course Evaluations
  - http://www.upenn.edu/eval

Big Ideas:

- Coping with uncertainty
- Statistical Reasoning and Calculation