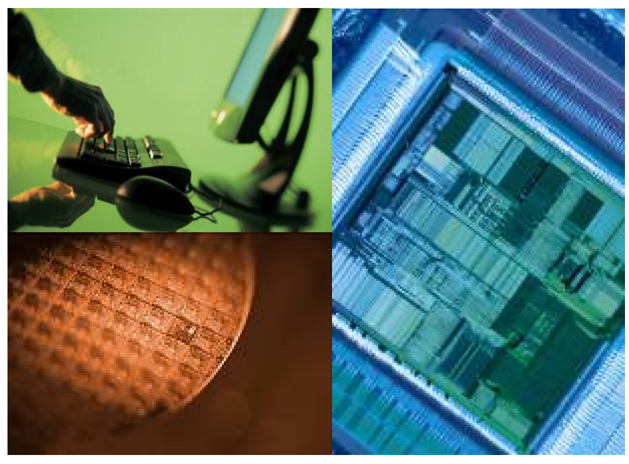
The Future Is Low Power

T. W. Williams, Ph.D.
Synopsys Fellow

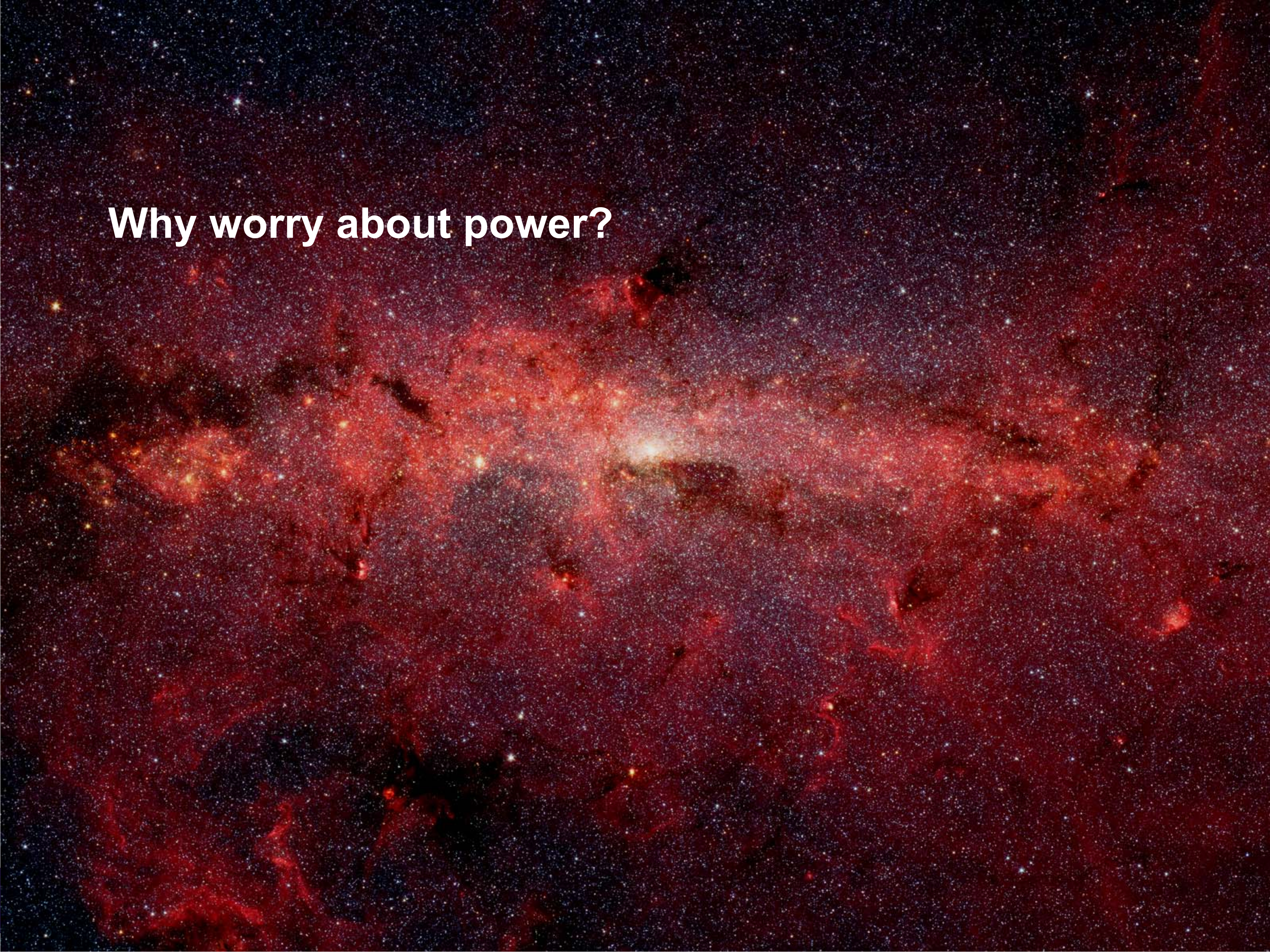


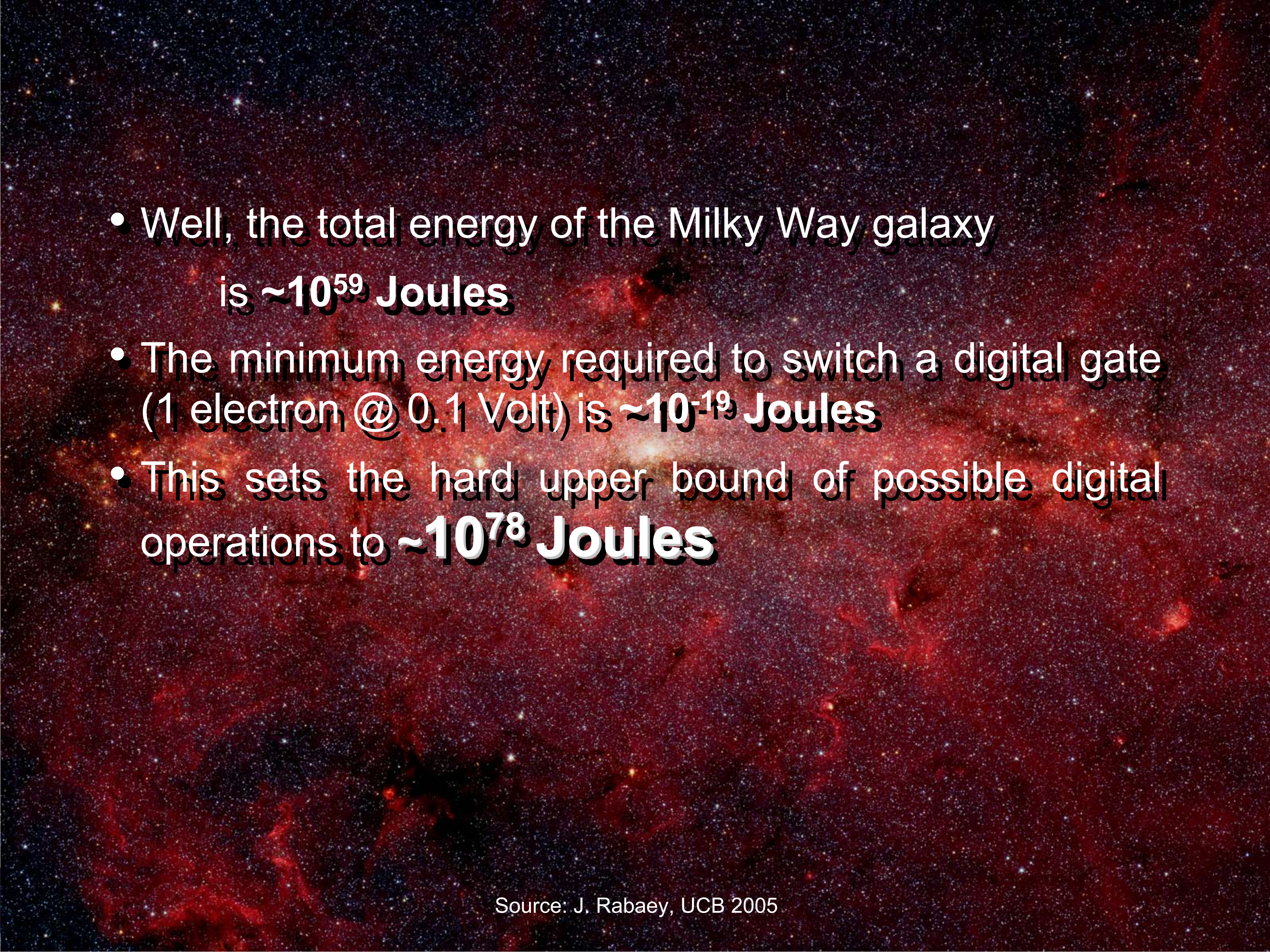
SYNOPSYS®
Predictable Success

the future **is** low power

PARTICIPANT
PRODUCTIONS
Copyright © 2006 by PARAMOUNT CLASSICS, a Division of PARAMOUNT PICTURES. All Rights Reserved.

Why worry about power?



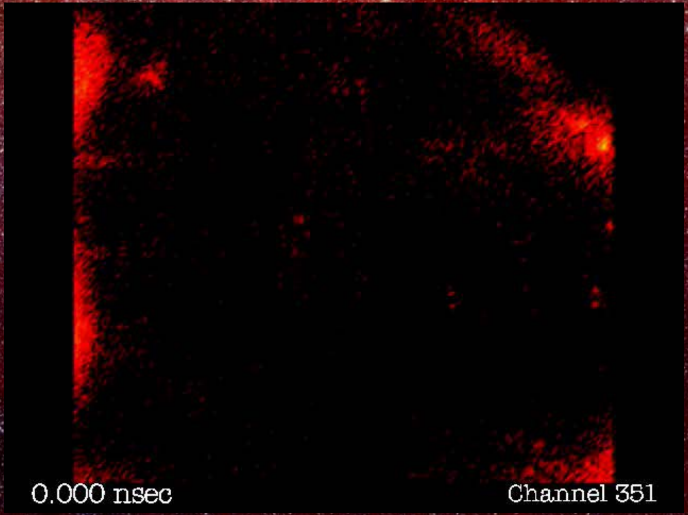


- Well, the total energy of the Milky Way galaxy

  is **~10^{59} Joules**
- The minimum energy required to switch a digital gate (1 electron @ 0.1 Volt) is **~10^{-19} Joules**
- This sets the hard upper bound of possible digital operations to **~10^{78} Joules**

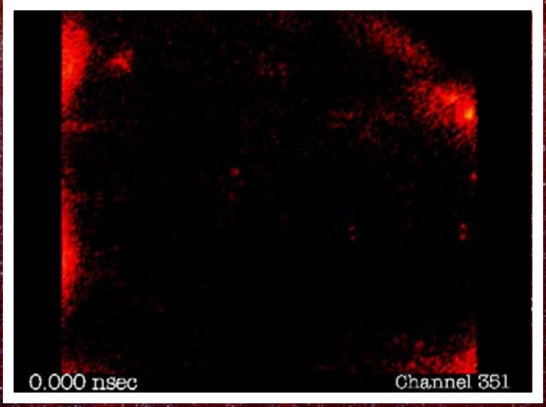
Source: J. Rabaey, UCB 2005

- Now, assuming 1 billion Core 2 Duo Extreme at work all over the world, this is ~10^{26} digital operations per year
- "Moore's Law is alive and well", isn't it? i.e. computing power doubles every 2 years
- **All galaxy's energy will be exhausted in ~343 years from now**



Source: J. Rabaey, UCB 2005

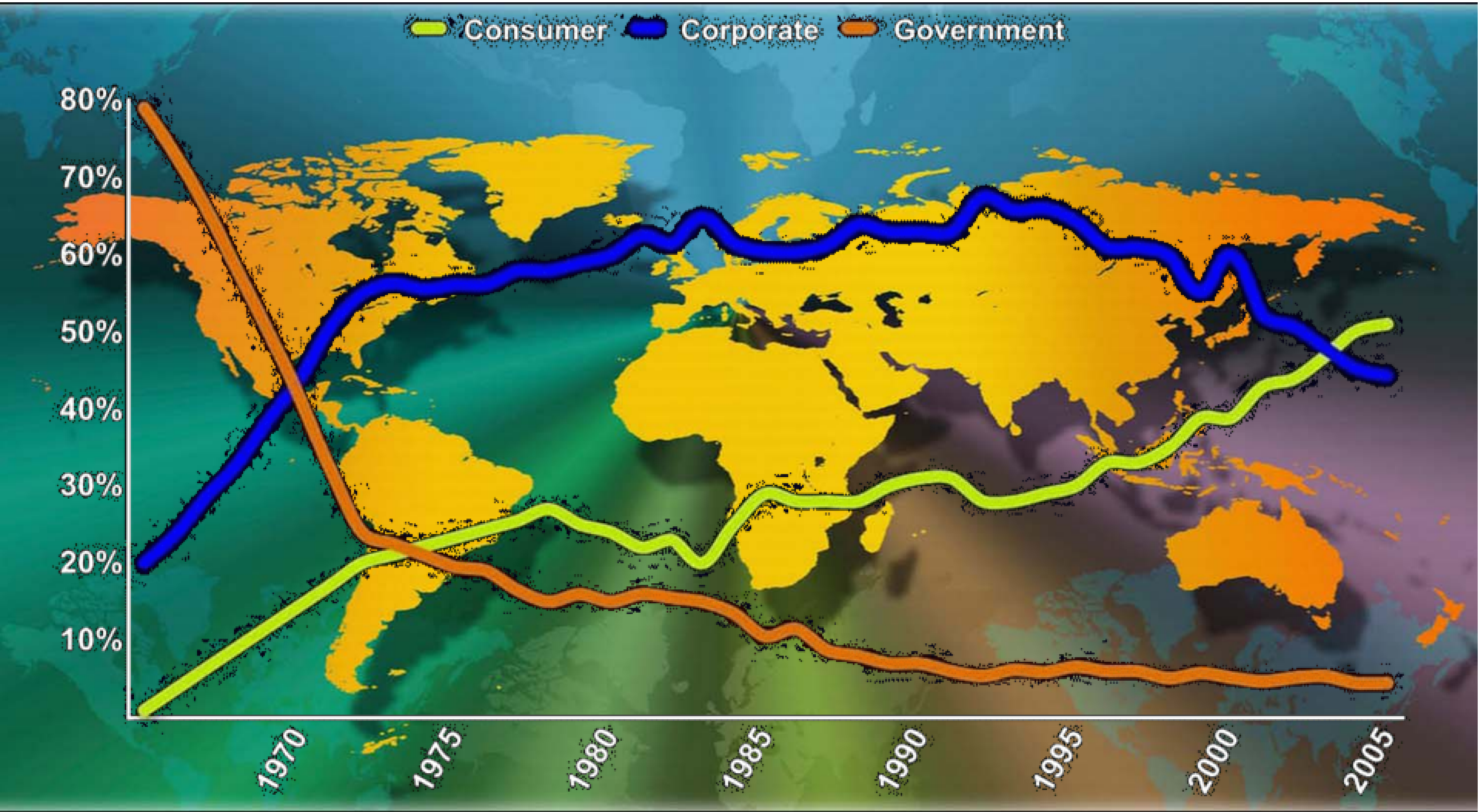
OK, Let’s worry about power!



Source: J. Rabaey, UCB 2005

The Era Of The Consumer

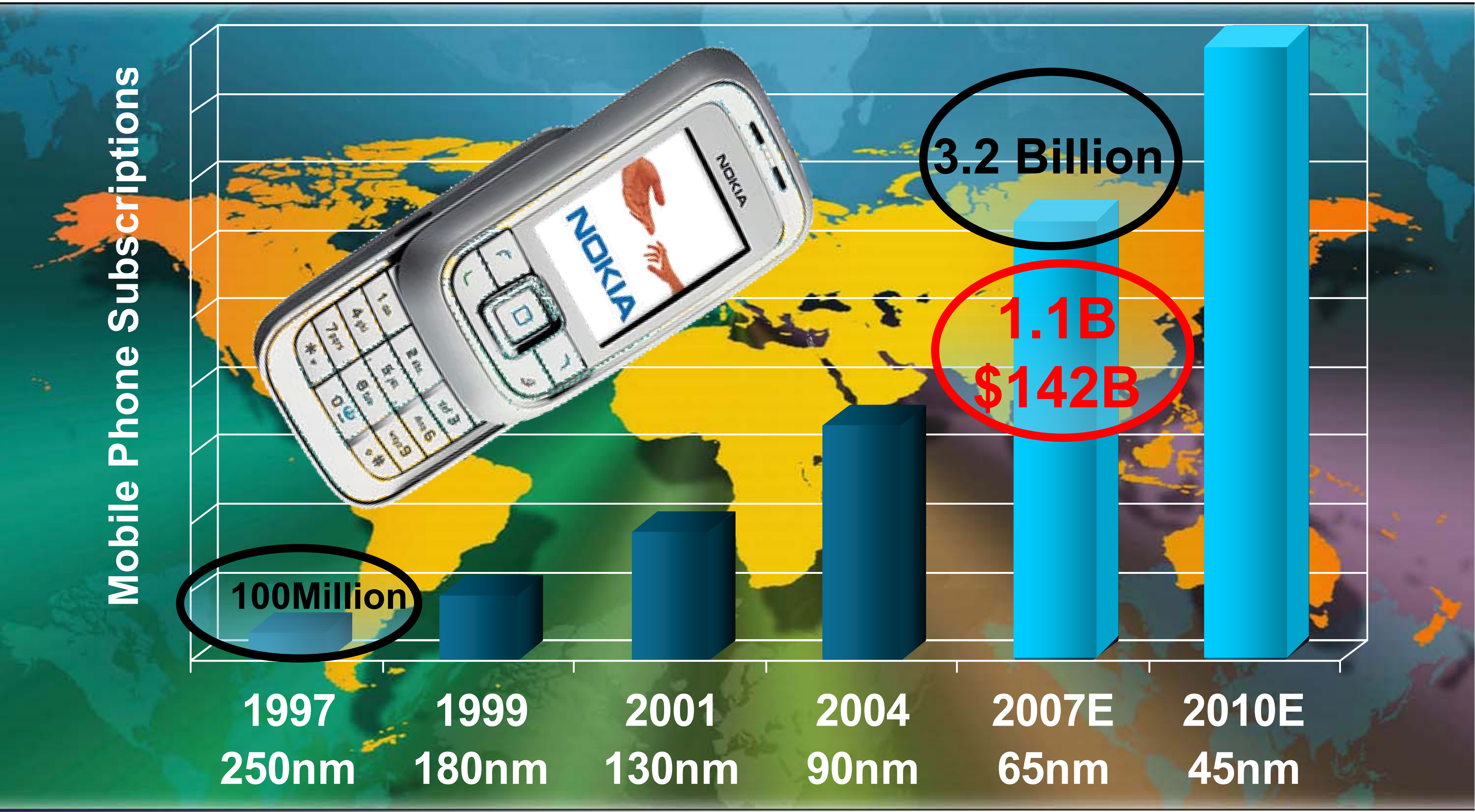
Households, Semiconductor Industry's #1 Customer



Source: J.-H. Huang, SIA 2004

Nanometer Market

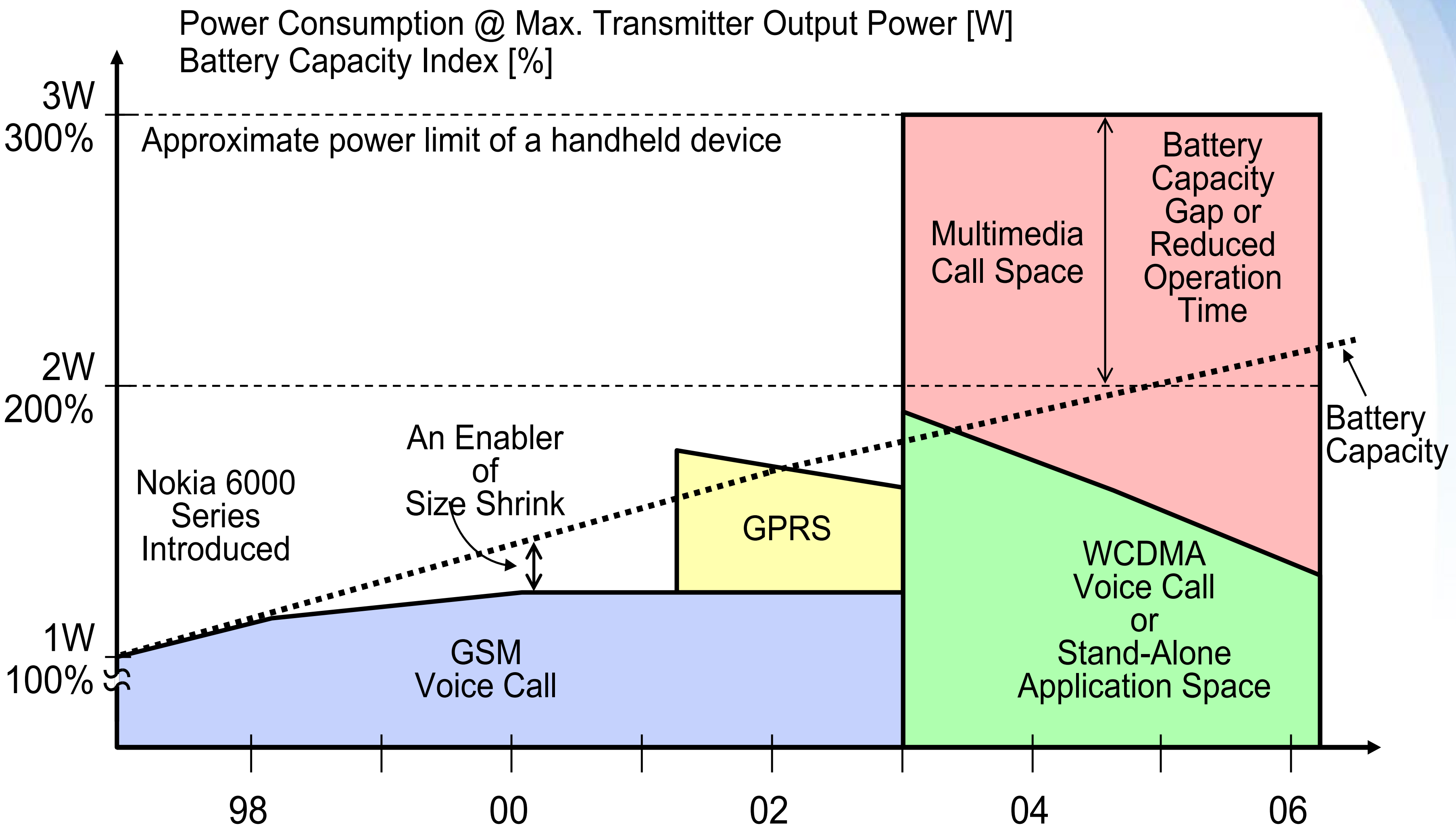
4.5 Billions Customers Already Within Reach



Source: IC Insights 2007

Power Consumption & Supply Trends

The Battery Capacity Gap



Source: M. Ryynänen, Nokia 2005

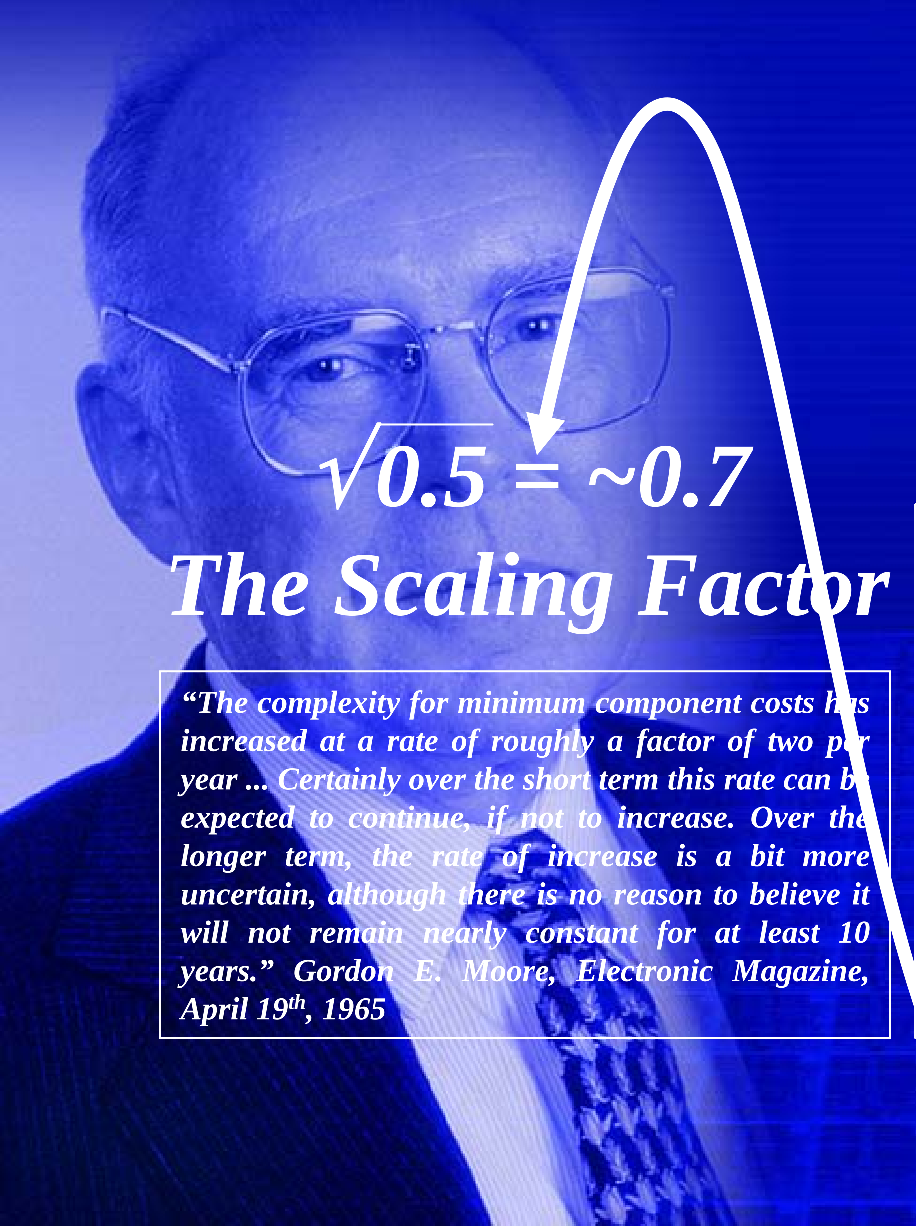
Where Do We Stand? Well…

“Houston, We Have A Problem”

Source: NASA 1970

Dr. Gordon E. Moore's Law

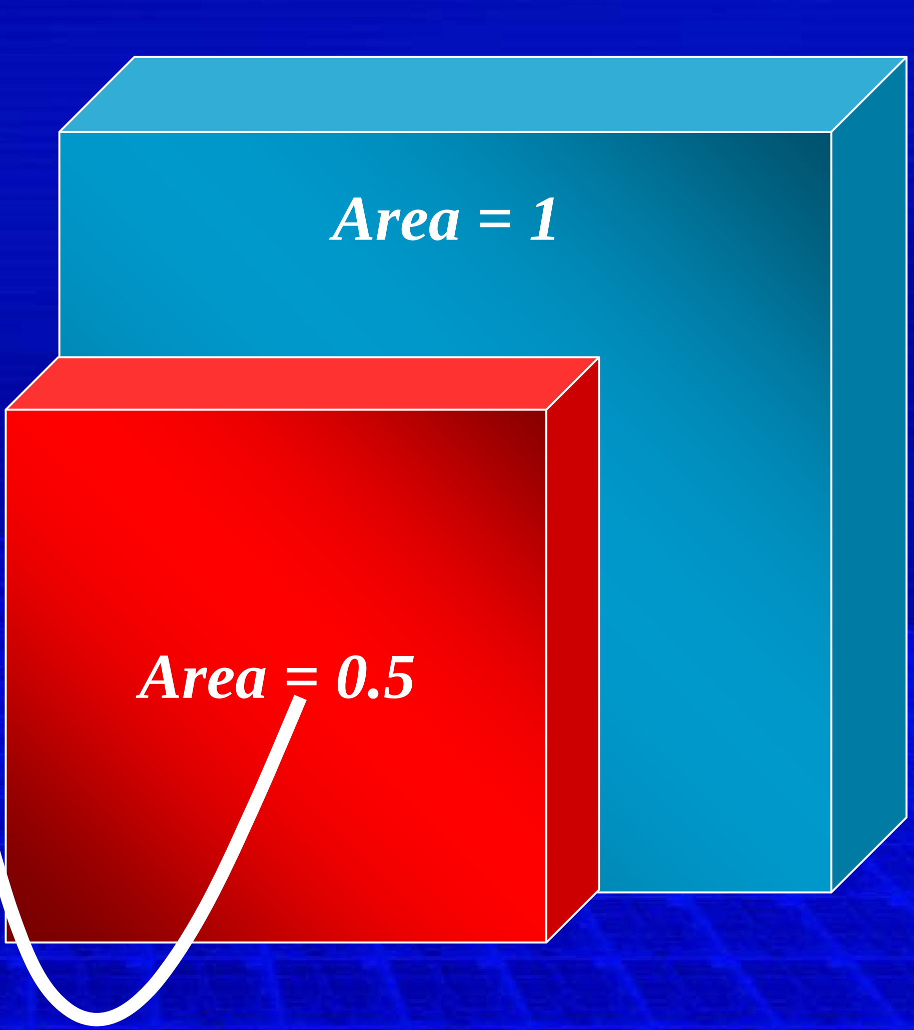
Integration's Capacity Doubles Every Two Years



$$\sqrt{0.5} = \sim 0.7$$

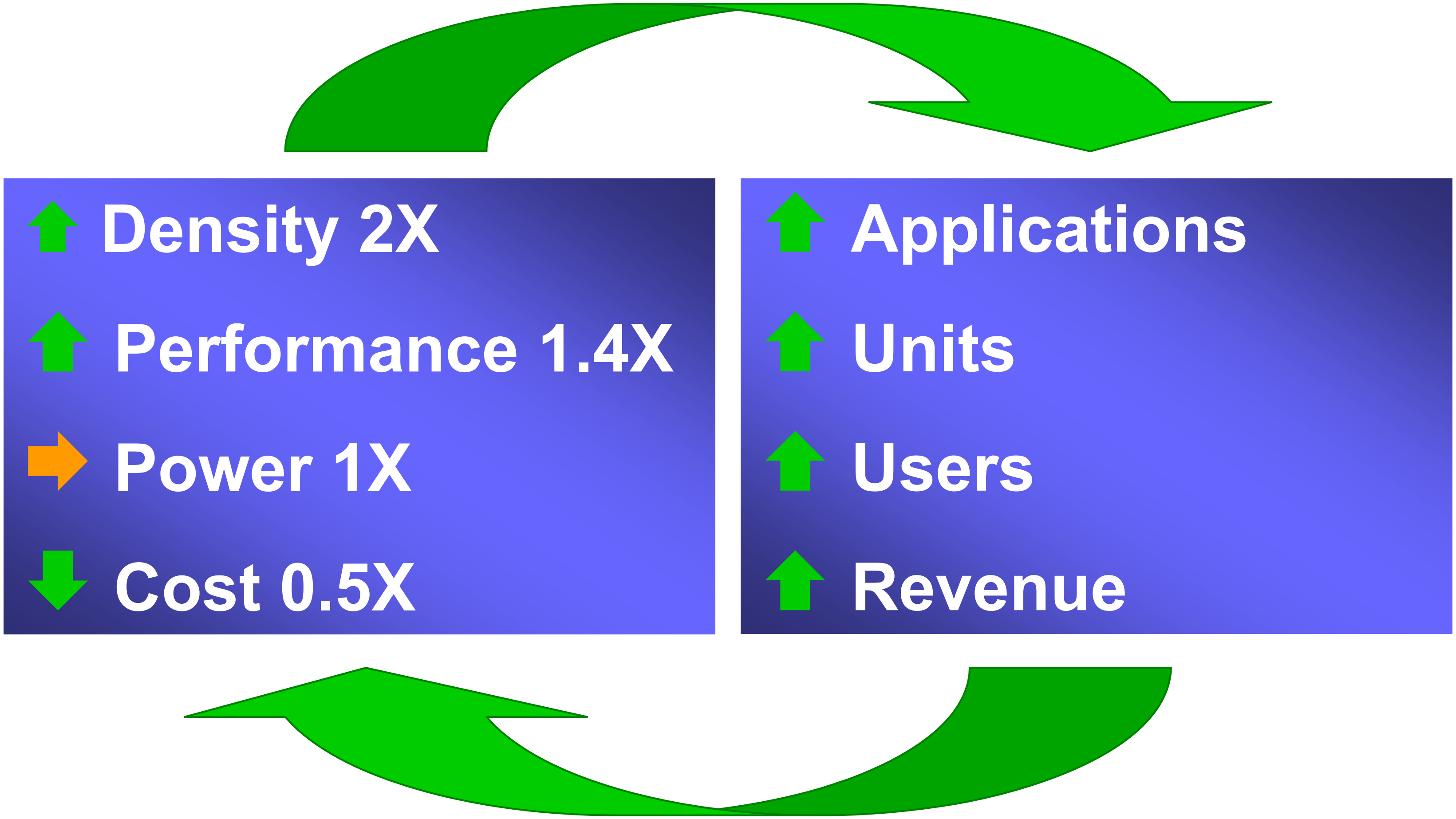
The Scaling Factor

"The complexity for minimum component costs has increased at a rate of roughly a factor of two per year ... Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years." Gordon E. Moore, Electronic Magazine, April 19th, 1965



Semiconductor Industry Cycle

Technology Advances, Market Grows



⬆ Density 2X

⬆ Performance 1.4X

➡ Power 1X

⬇ Cost 0.5X

⬆ Applications

⬆ Units

⬆ Users

⬆ Revenue

The Signs of Crisis Are Visible Already

Voltage Has Broken the Rules of Scaling

Source: ITRS 2005/2006	90nm	65nm	45nm
Device Length (nm) ⇩	1X	0.7X	0.5X
Delay (ps) ⇩	1X	0.7X	0.5X
Frequency (GHz) ⇧	1X	**1.2X**	**1.45X**
Integration Capacity (BTx) ⇧	1X	2X	4X
Capacitance (fF) ⇩	1X	0.7X	0.5X
Die Size (mm^2) ⇨	1X	**1X**	**1X**
Voltage (V) ⇘	1X	**0.85X**	**0.75X**
Dynamic Power (W) ⇘	1X	**> 0.7X**	**> 0.5X**
Manufacturing (microcents/Tx) ⇩	1X	0.35X	0.12X

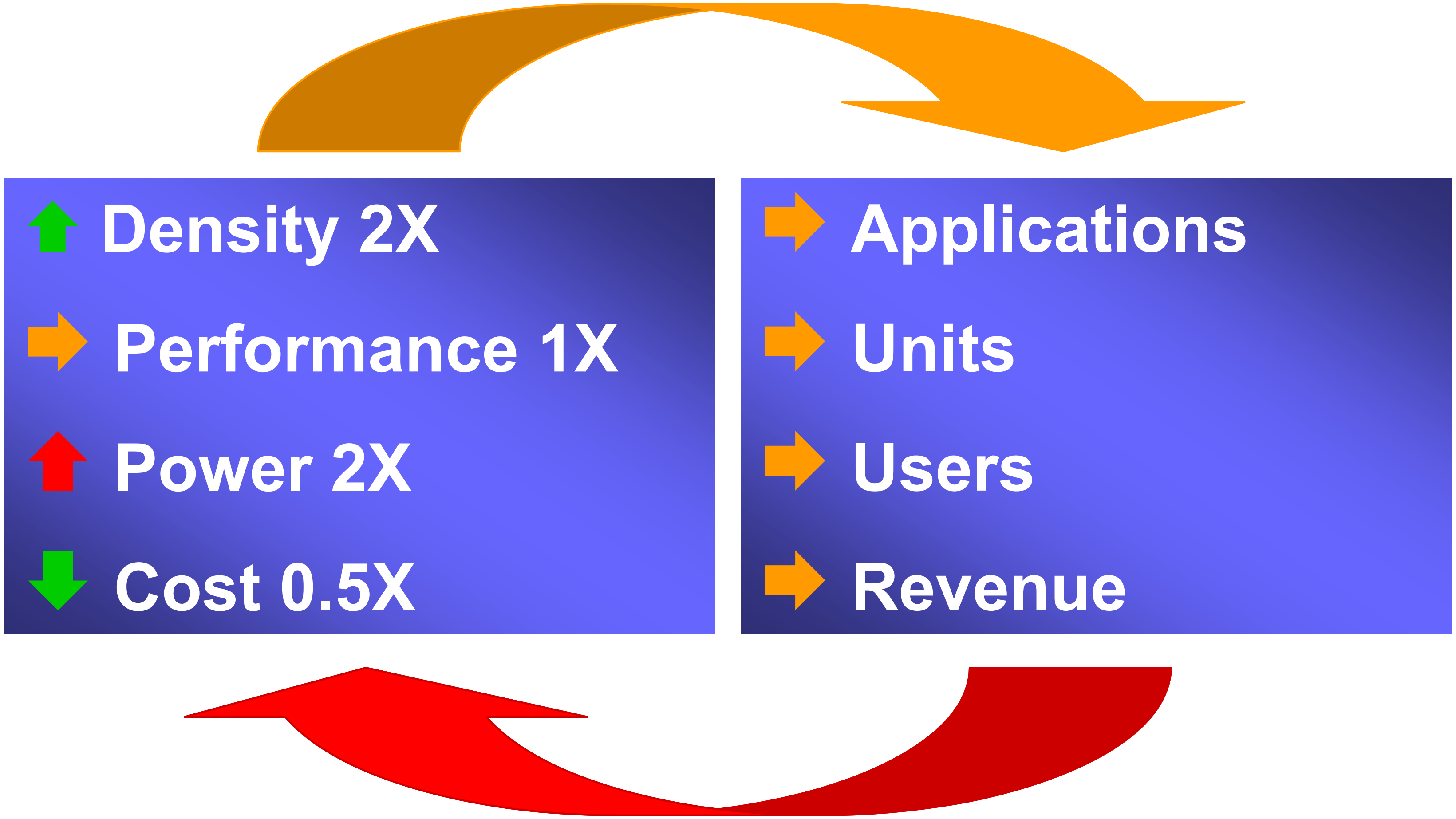
The [not so] Hidden Costs of Scaling

New Materials [and Devices] Are Badly Needed

Source: ITRS 2005	90nm	65nm	45nm
V_{TH} (V) ⇘	1X	**0.85X**	**0.75X**
I_{OFF} (nA/um) ⇧⇧	1X	**~3X**	**~9X**
Dynamic Power Density (W/cm^2) ⇧	1X	**1.43X**	**2X**
Leakage Power Density (W/cm^2) ⇧⇧	1X	**~2.5X**	**~6.5X**
Power Density (W/cm^2) ⇧	1X	**~2X**	**~4X**
Cu Resistance (Ω) ⇧	1X	**2X**	**4X**
Interconnect RC Delay (ps) ⇧	1X	**~2X**	**~5X**
Packaging (cents/pin) ⇘	1X	**0.86X**	**0.73X**
Test (nanocents/Tx) ⇨	1X	**1X**	**1X**

Semiconductor Industry Stalemate

Technology (CMOS) Shrinks, Doesn't Advance

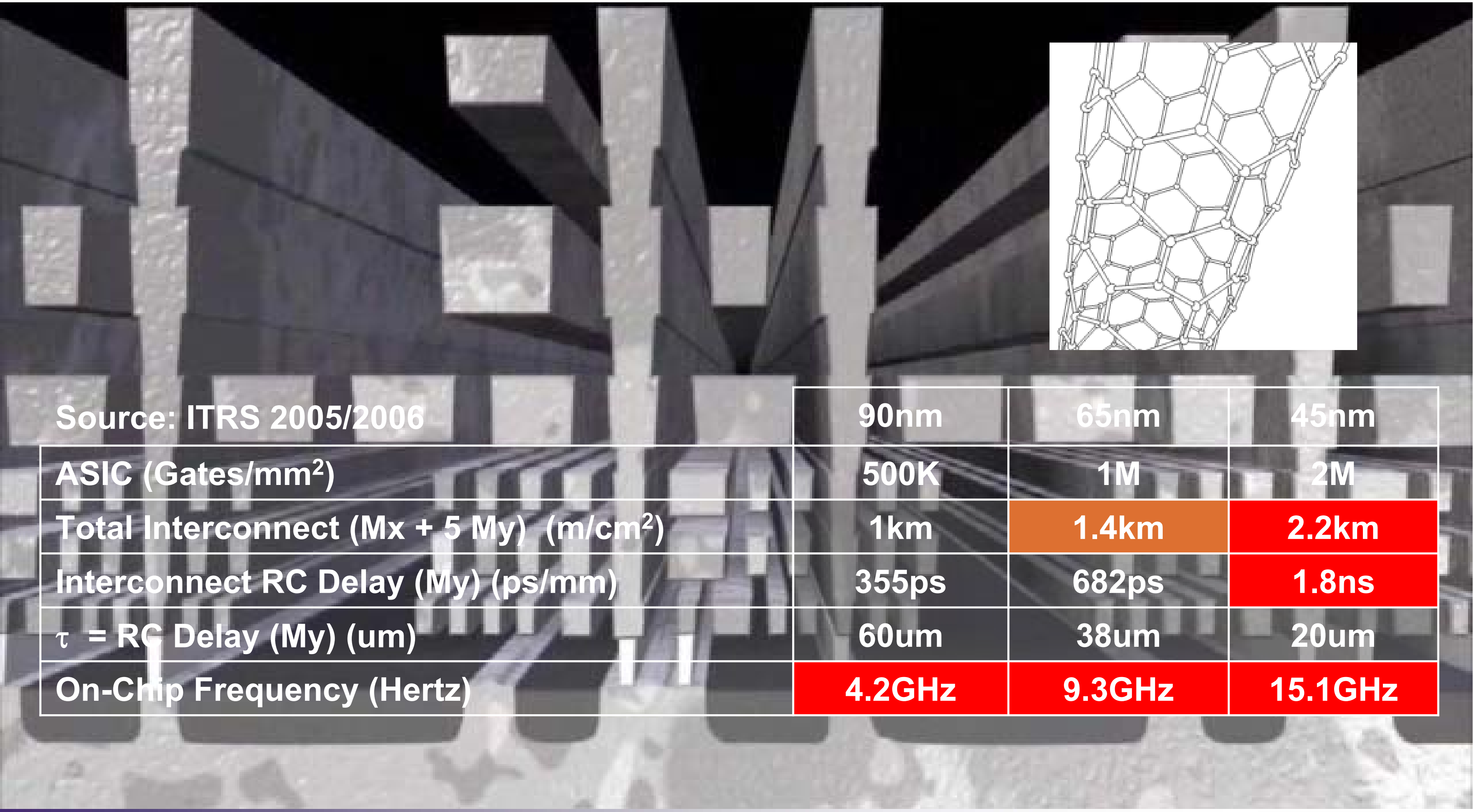


- ⬆ Density 2X
- ➡ Performance 1X
- ⬆ Power 2X
- ⬇ Cost 0.5X

- ➡ Applications
- ➡ Units
- ➡ Users
- ➡ Revenue

Quasi-Atomic-Level Interconnect

Main Contributor to both Timing and Dynamic Power

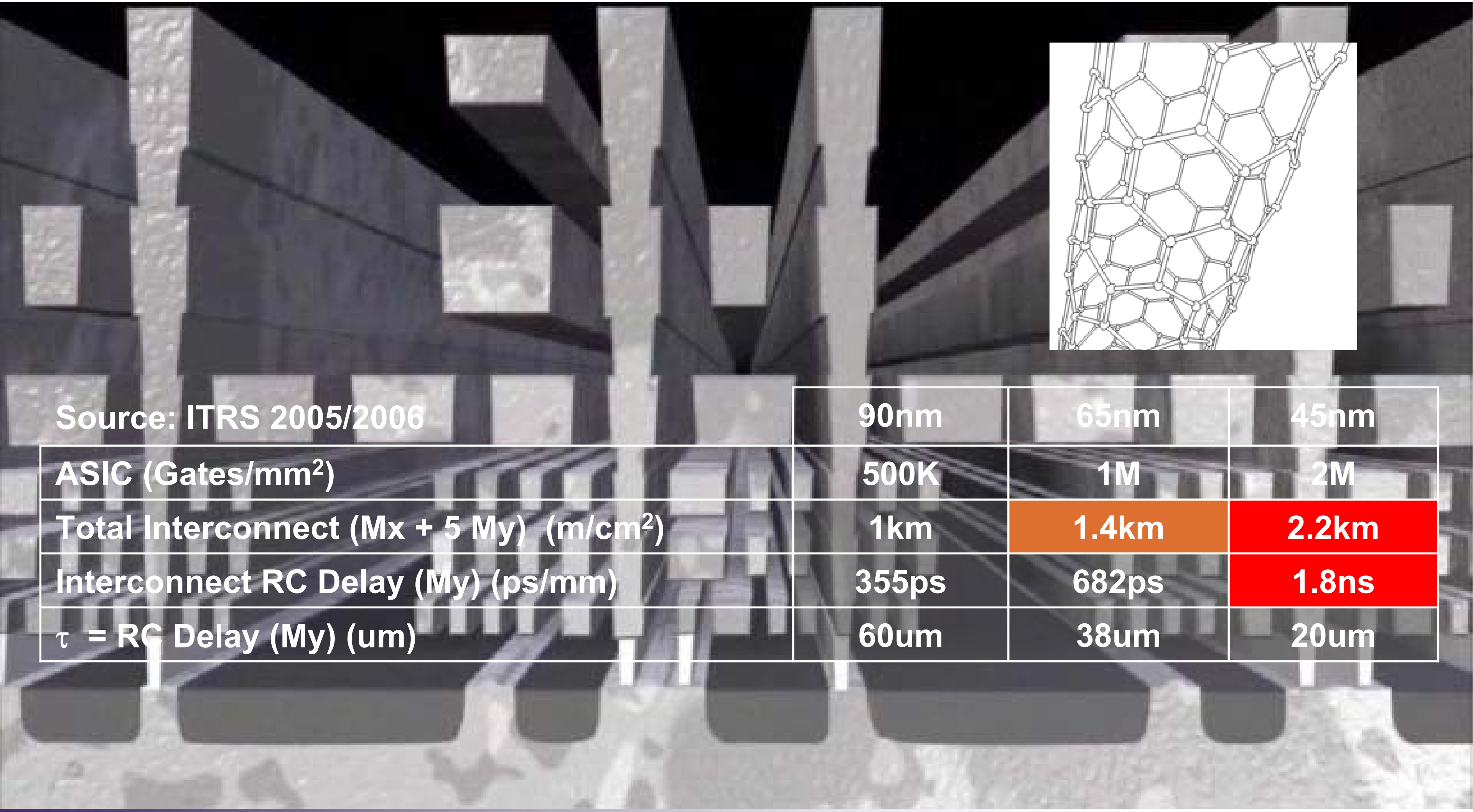


Source: ITRS 2005/2006	90nm	65nm	45nm
ASIC (Gates/mm^2)	500K	1M	2M
Total Interconnect (Mx + 5 My) (m/cm^2)	1km	1.4km	2.2km
Interconnect RC Delay (My) (ps/mm)	355ps	682ps	1.8ns
τ = RC Delay (My) (um)	60um	38um	20um
On-Chip Frequency (Hertz)	4.2GHz	9.3GHz	15.1GHz

Source: R. Chau, Intel 2003

Quasi-Atomic-Level Interconnect

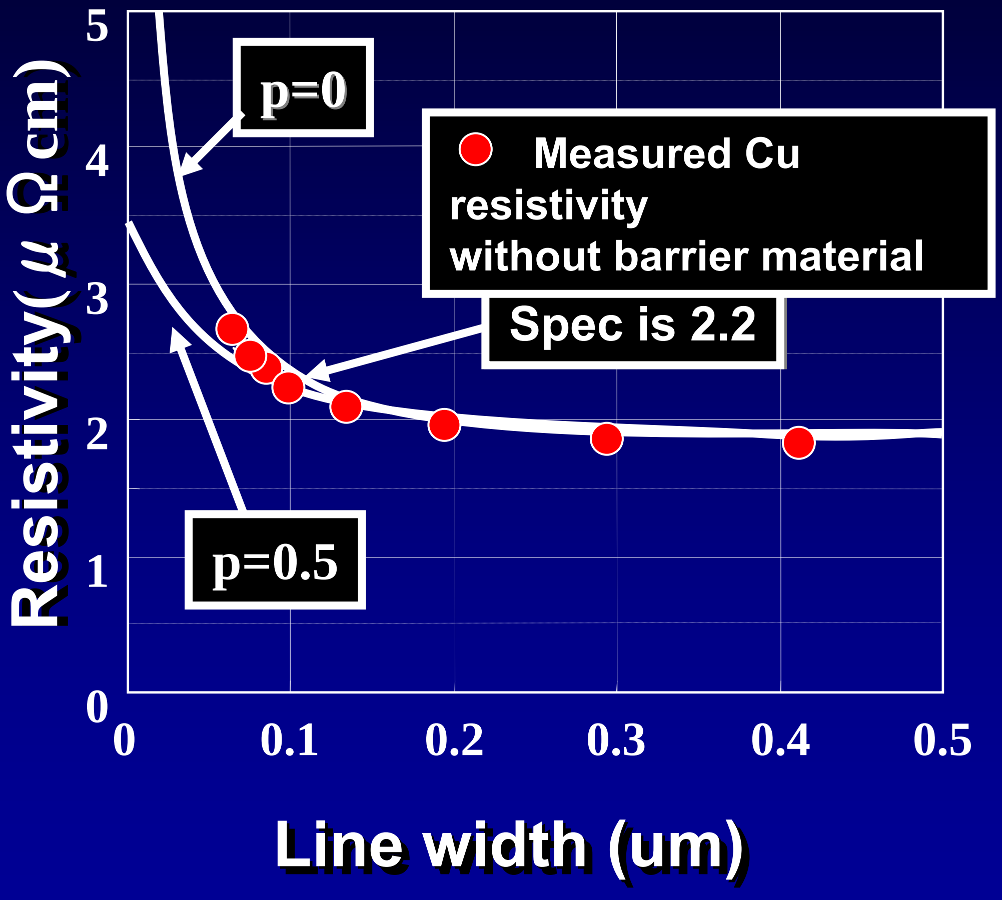
Main Contributor to both Timing and Dynamic Power

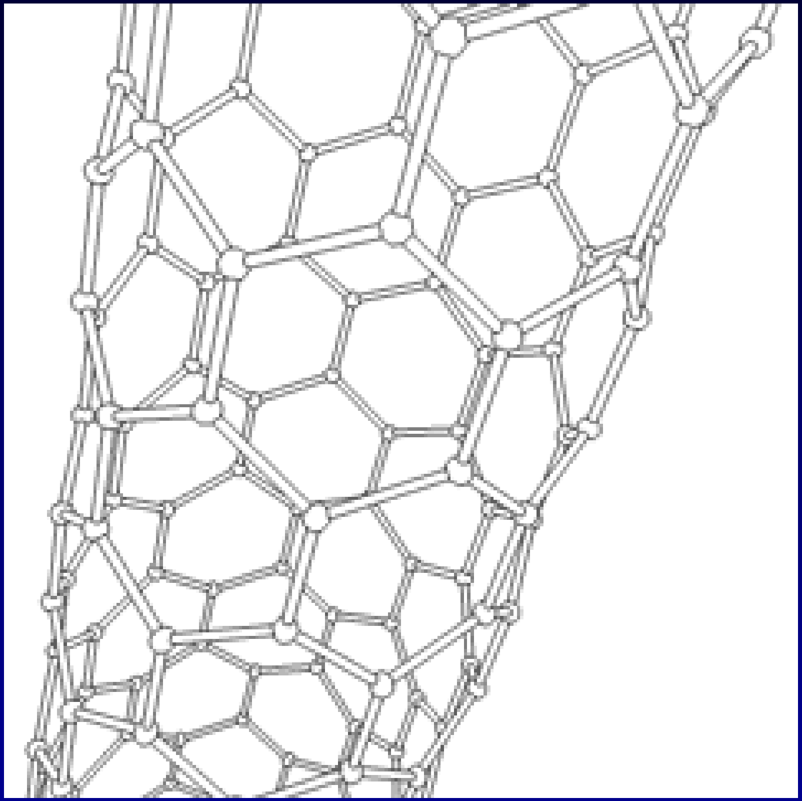


Source: ITRS 2005/2006	90nm	65nm	45nm
ASIC (Gates/mm^2)	500K	1M	2M
Total Interconnect (Mx + 5 My) (m/cm^2)	1km	1.4km	2.2km
Interconnect RC Delay (My) (ps/mm)	355ps	682ps	1.8ns
τ = RC Delay (My) (um)	60um	38um	20um

Source: R. Chau, Intel 2003

Cu Resistivity Increase

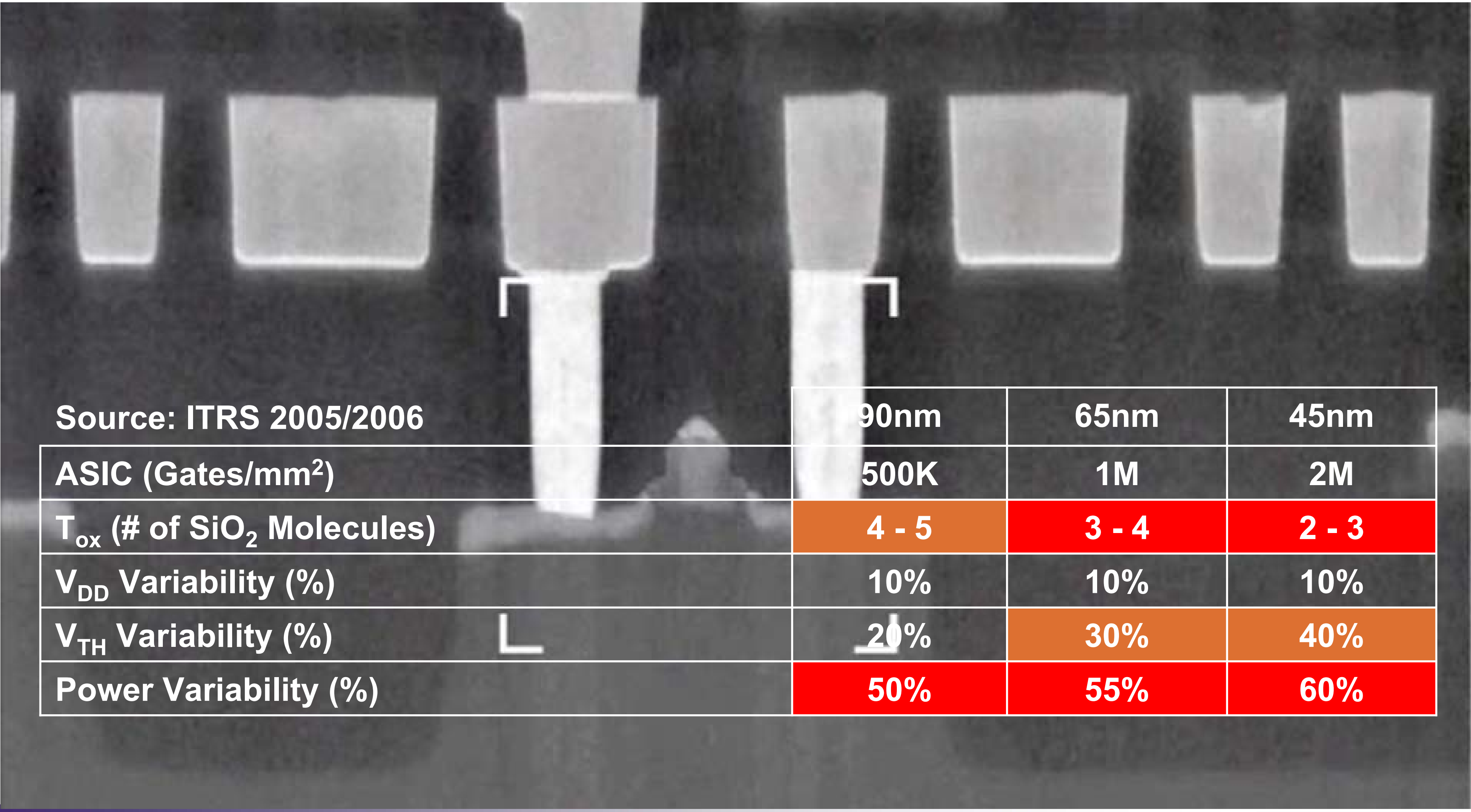




Free Mean Path 400-600 A for cupper, Al is > 1000 A

Atomic-Level Lithography

Main Contributor to both Leakage Power & Variability

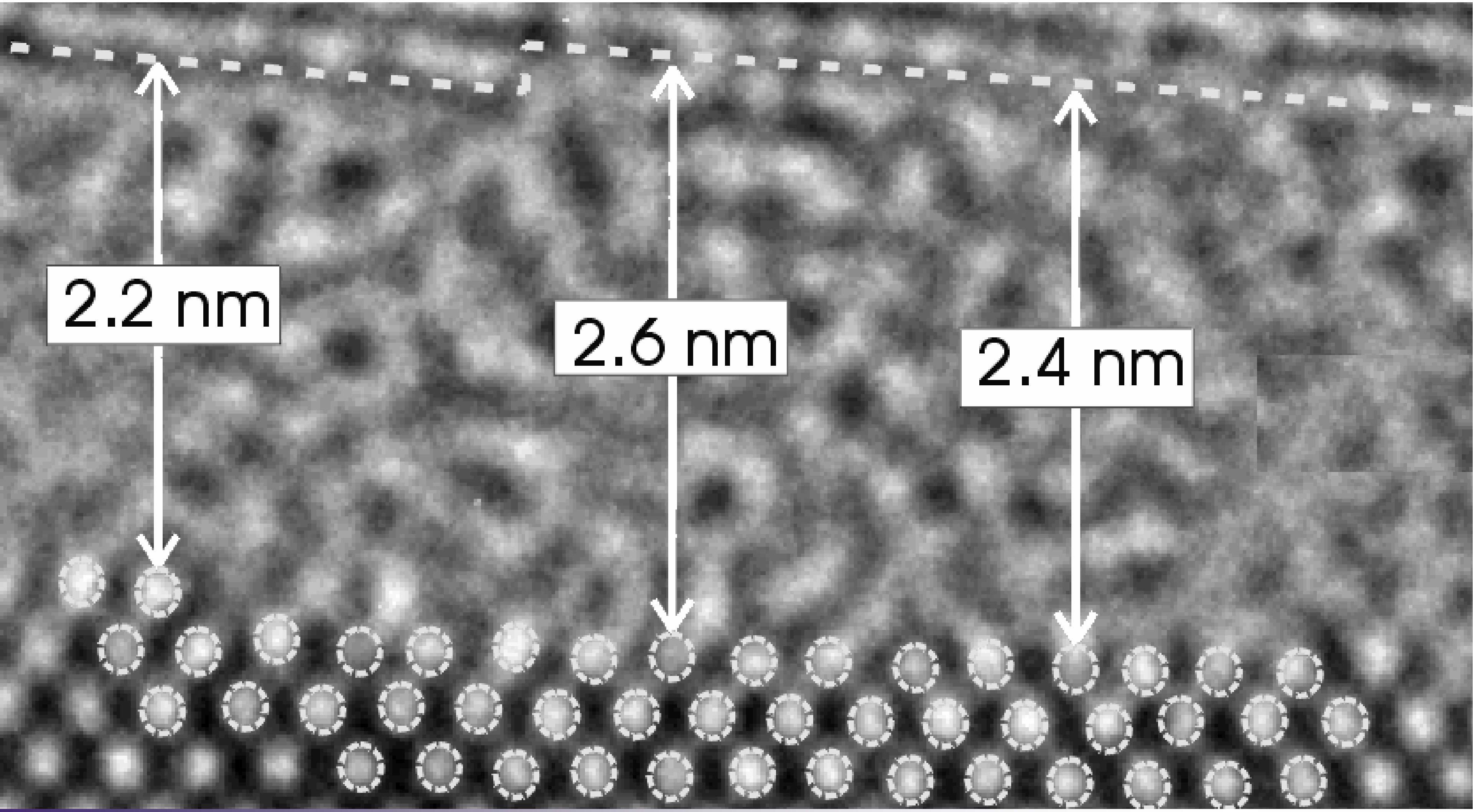


Source: ITRS 2005/2006	90nm	65nm	45nm
ASIC (Gates/mm^2)	500K	1M	2M
T_{ox} (# of SiO_2 Molecules)	4 - 5	3 - 4	2 - 3
V_{DD} Variability (%)	10%	10%	10%
V_{TH} Variability (%)	20%	30%	40%
Power Variability (%)	50%	55%	60%

Source: R. Chau, Intel 2003

Molecules/Atoms Make The Difference

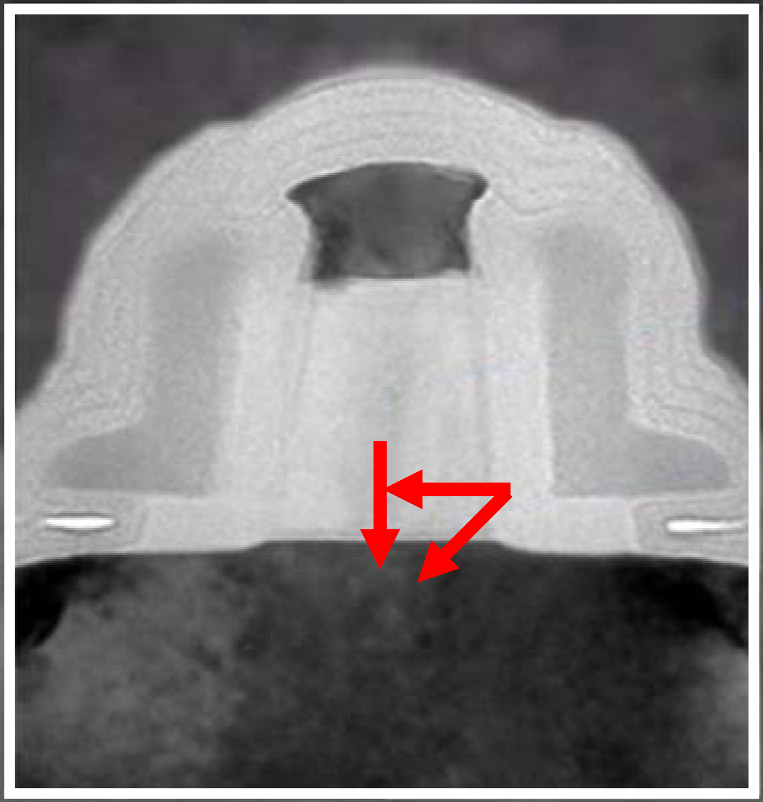
50% Power Variability Already



© Chandu Visweswariah, IBM 2004, ITRS 2005/2006

SiO_2 Dielectric/Polysilicon Gate

High Performance vs. Low Power Dilemma



Subthreshold Leakage

V_{TH} ⇩ **I_{OFF} ⇧**

Junction Leakage

I_{JUNC}

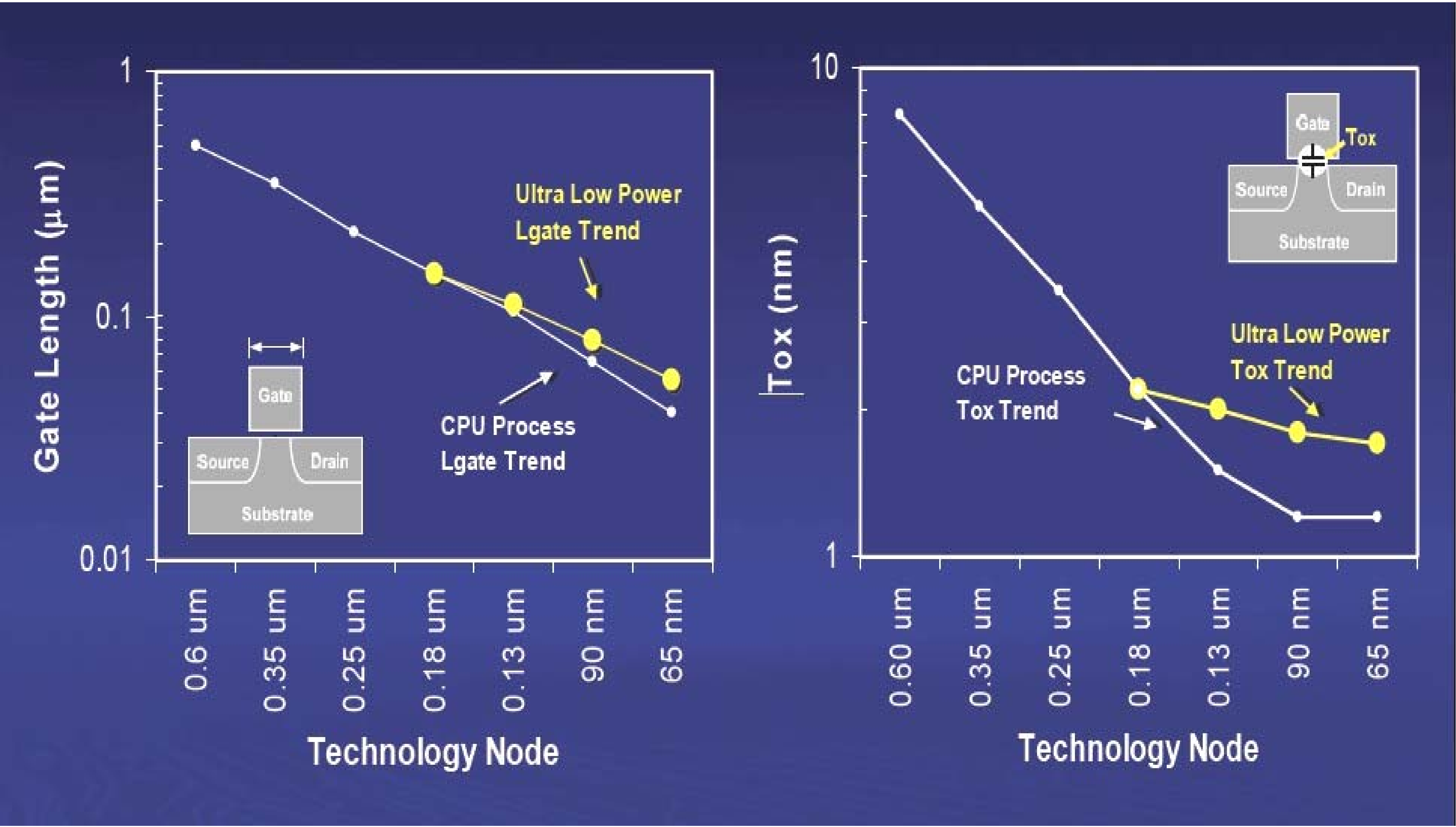
Gate Leakage

T_{OX} ⇩ **I_{GATE} ⇧**

Source: R. Chau, Intel 2003

65 Nanometers, Ultra-Low Power…

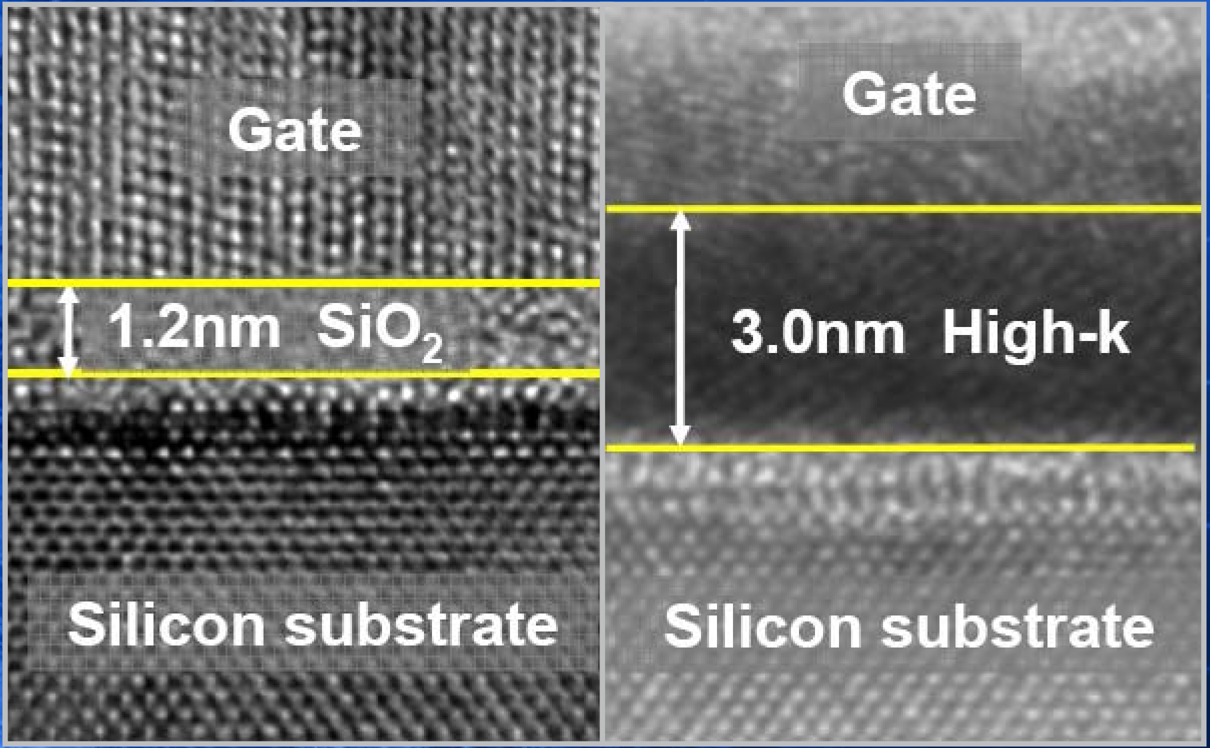
It's Actually 130nm (T_{OX}) 90nm (L_{GATE}) High Performance



Source: P. Bai, Intel, IEDM 2004; C.H. Jan, Intel, IEDM 2005

High-k Dielectric & Metal Gate

@ 45 Nanometers



Source: P. Gargini, Intel 2004

Is High-k Dielectric *The* Solution?

Benefits of Intel’s 45nm Technology vs. 65nm One

> 20% improvement in transistor switching speed

OR

5X I_{OFF} reduction

10X I_{GATE} reduction

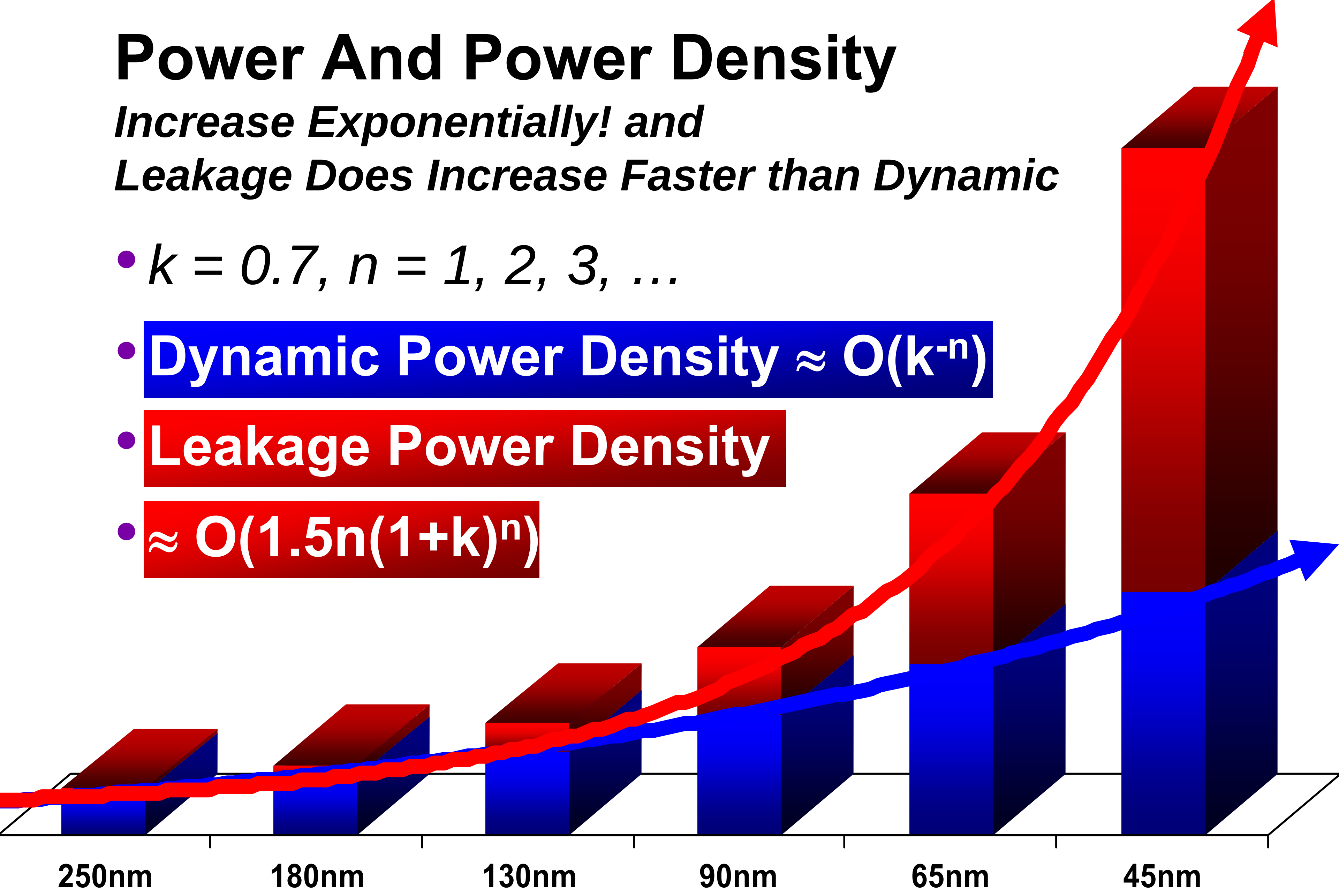
> 30% dynamic power reduction

Source: W. Holt, Intel 2007

Power And Power Density

Increase Exponentially! and Leakage Does Increase Faster than Dynamic

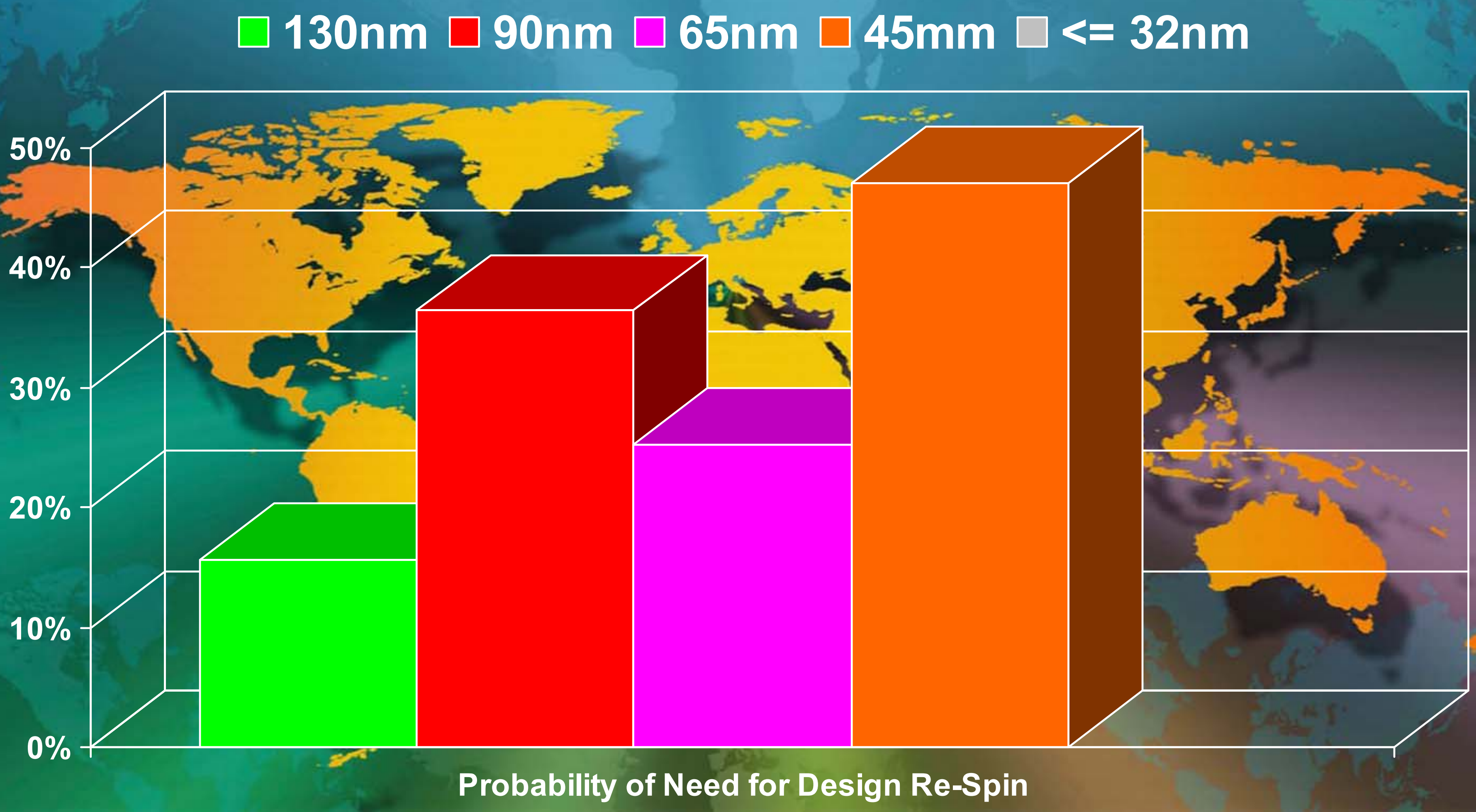
- *k = 0.7, n = 1, 2, 3, …*
- **Dynamic Power Density $\approx O(k^{-n})$**
- **Leakage Power Density**
- **$\approx O(1.5n(1+k)^n)$**



Source: ITRS 2005/2006

Probability Of Need For Design Re-Spin

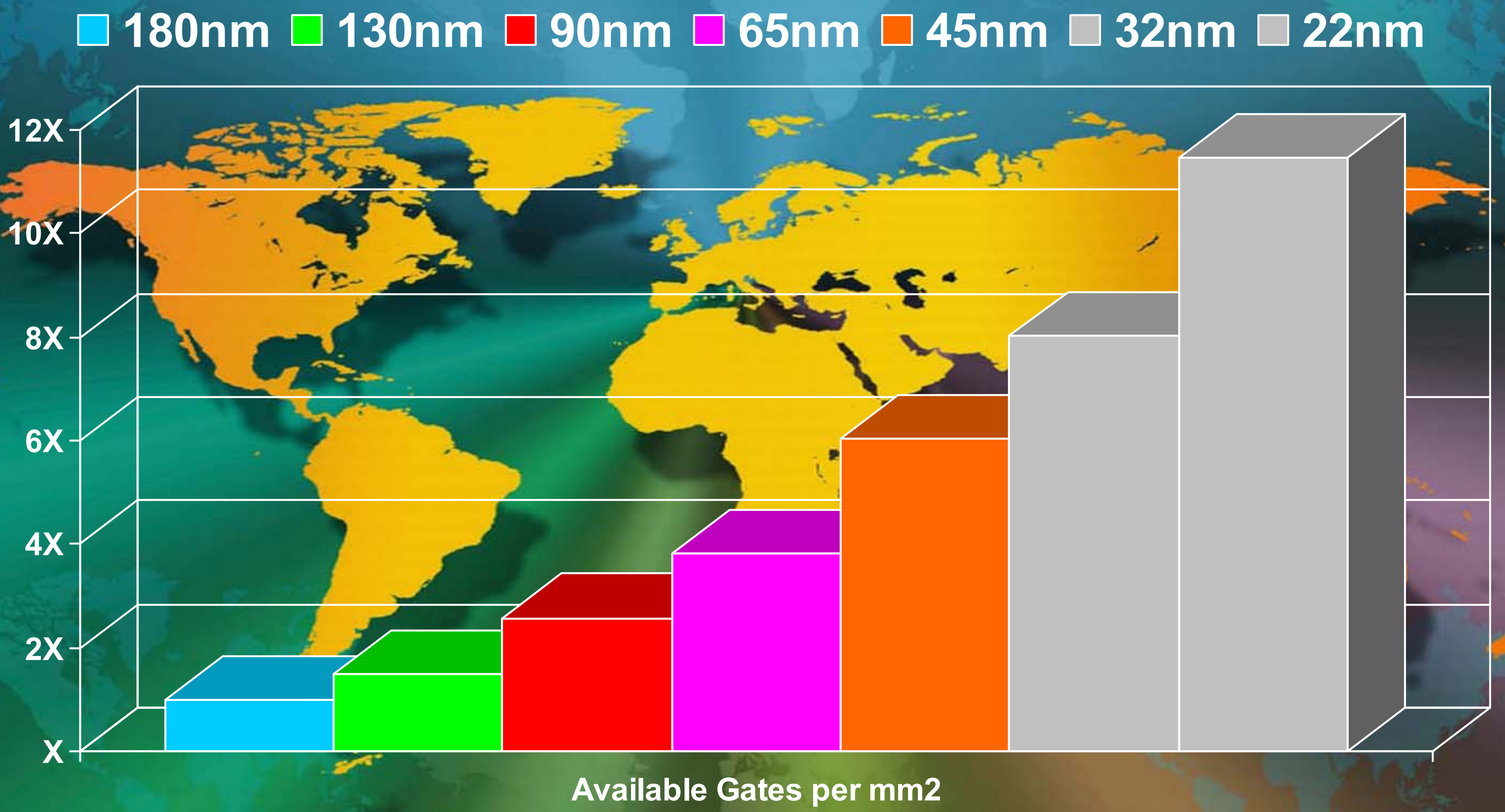
@ 45nm 50% of Design Re-Spins Due to Leakage



Source: IBS 2007

From Exponential To Asymptotic

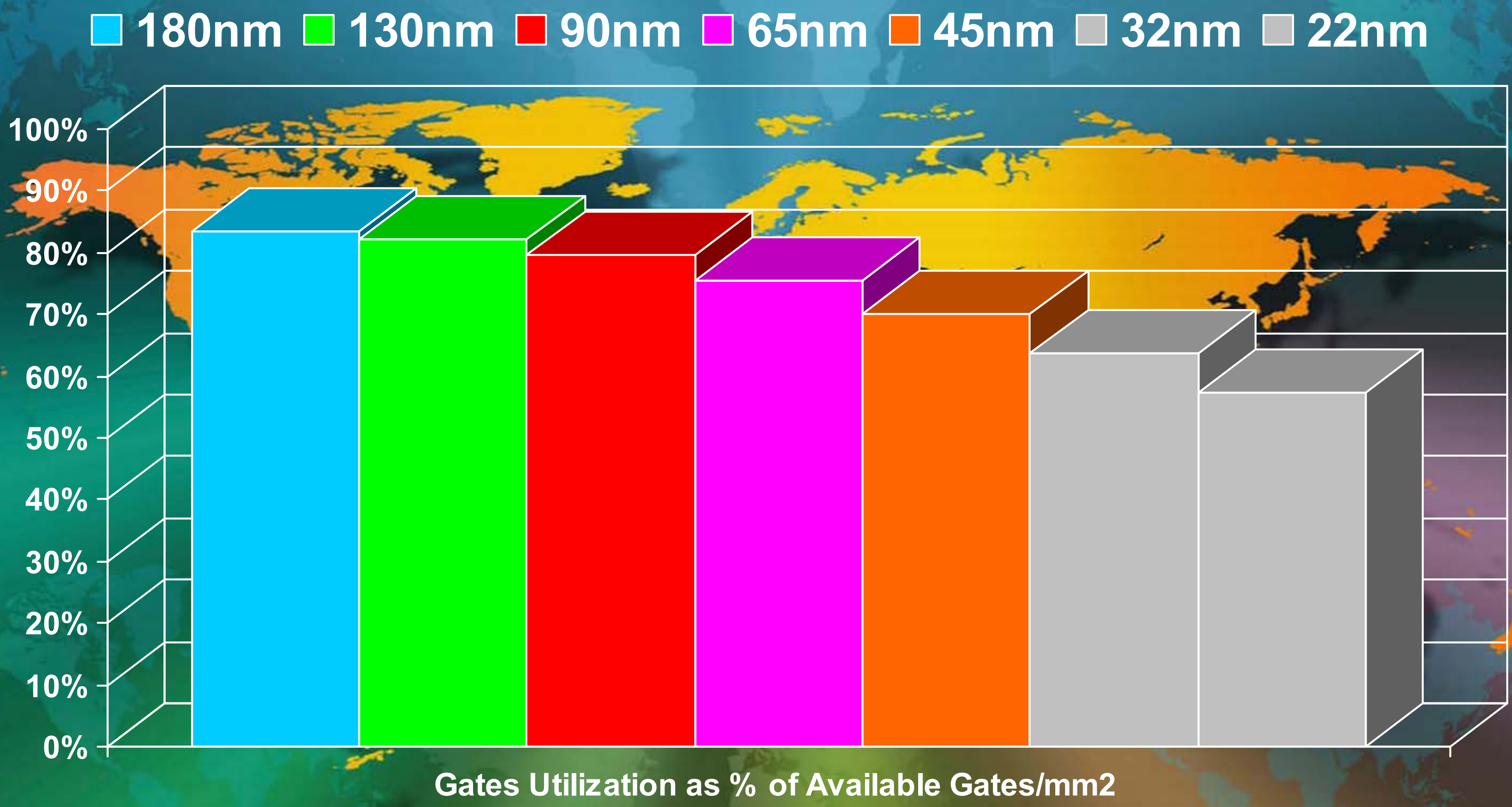
Actual Number of Gates/mm^2 Increase, But Slower!



Source: IBS 2007

It’s All About [In]Efficiency

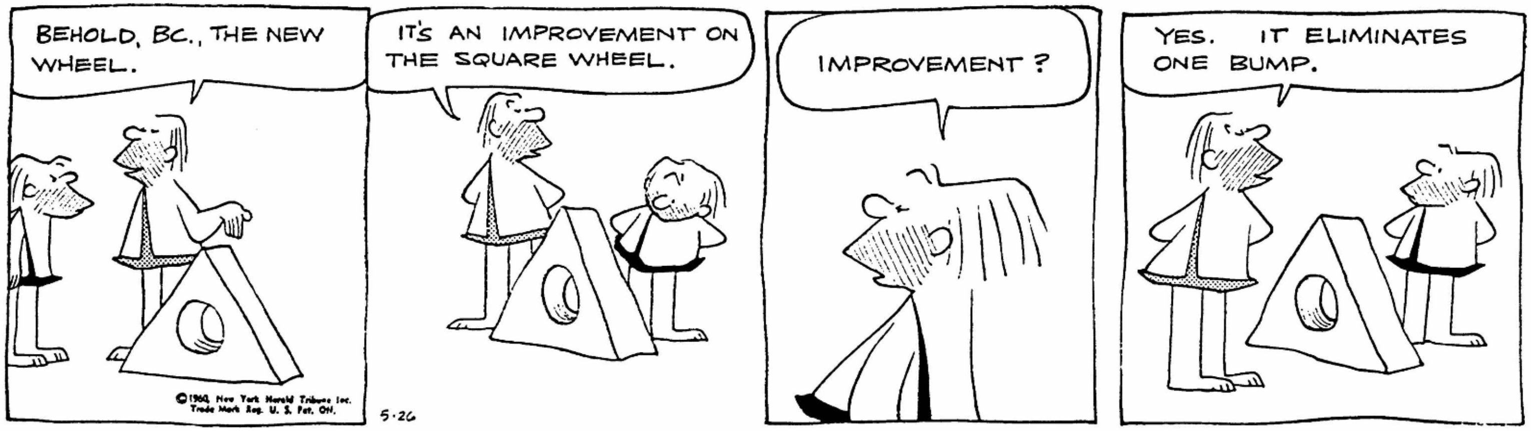
Gates Utilization as % of Available Gates/mm^2 Declines



Source: IBS 2007

The Nanometer Application

The Most Obvious Solutions Aren’t Necessarily the Right Ones


BEHOLD, BC., THE NEW WHEEL.
IT'S AN IMPROVEMENT ON THE SQUARE WHEEL.
IMPROVEMENT ?
YES. IT ELIMINATES ONE BUMP.
©1960 New York Herald Tribune Inc. Trade Mark Reg. U. S. Pat. Off.
5·26

Design For Low Power

All the Right Ingredients Are Available

	Constant Throughput/Latency		Variable Throughput/Latency
	Design Time	Non-Active Modules	Run Time
Dynamic & Short Circuit	Logic Re-Structuring, Logic Sizing Reduced V_{DD} Multi-V_{DD} **2.5X**	Clock Gating **2X**	Dynamic or Adaptive Frequency & Voltage Scaling **2.5X**
Leakage	Stack Effect + Multi-V_{TH} **2X-10X**	Sleep Transistors Multi-V_{DD} Variable V_{TH} **10X-1000X**	Variable V_{TH} **2X-10X**

Source: J. Rabaey, UCB 2005

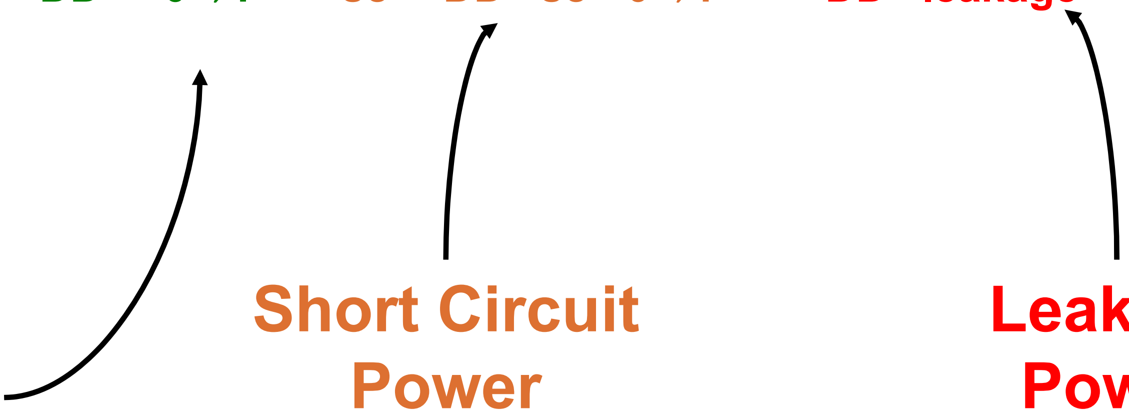
CMOS Energy & Power Equations

Dynamic, Short Circuit & Leakage

$$\text{Energy} = C_L V_{DD}^2 P_{0\rightarrow 1} + t_{sc} V_{DD} I_{sc} P_{0\rightarrow 1} + V_{DD} I_{leakage}$$

$$f_{0\rightarrow 1} = P_{0\rightarrow 1}\ \text{frequency}$$

$$\text{Power} = C_L V_{DD}^2 f_{0\rightarrow 1} + t_{sc} V_{DD} I_{sc} f_{0\rightarrow 1} + V_{DD} I_{leakage}$$



Dynamic Power

50% @ 65nm
decreasing <u>relatively</u>

Short Circuit Power

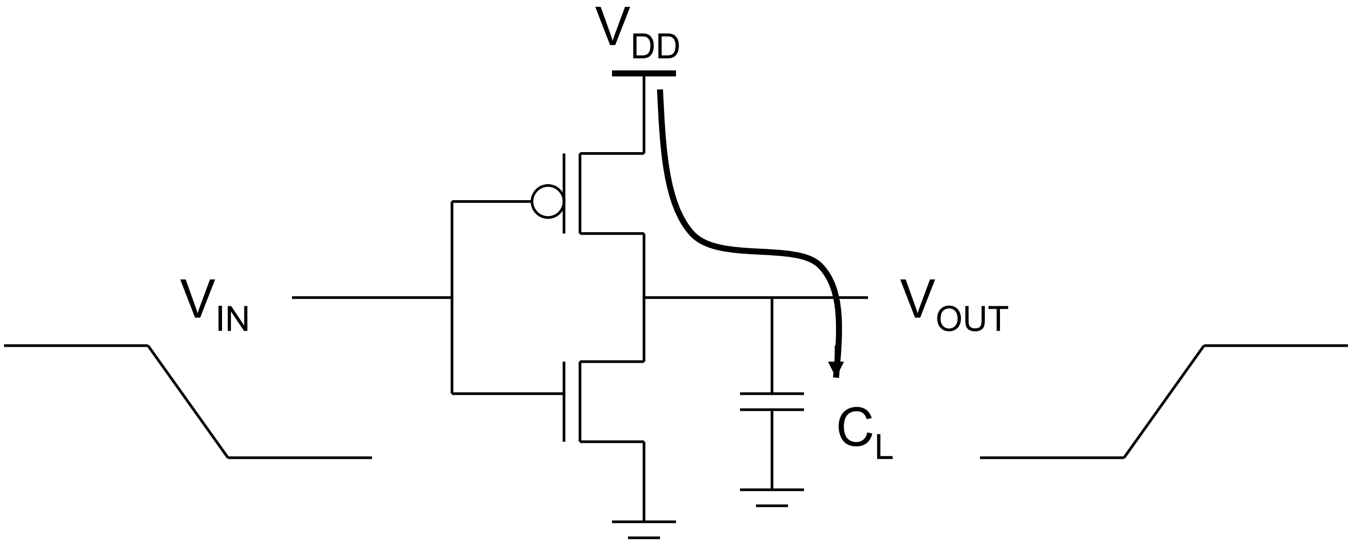
1-2% @ 65nm,
decreasing <u>absolutely</u>

Leakage Power

50% @ 65nm,
increasing

Dynamic Power Consumption

Depends on Output Load, V_{DD} & Switching Activity



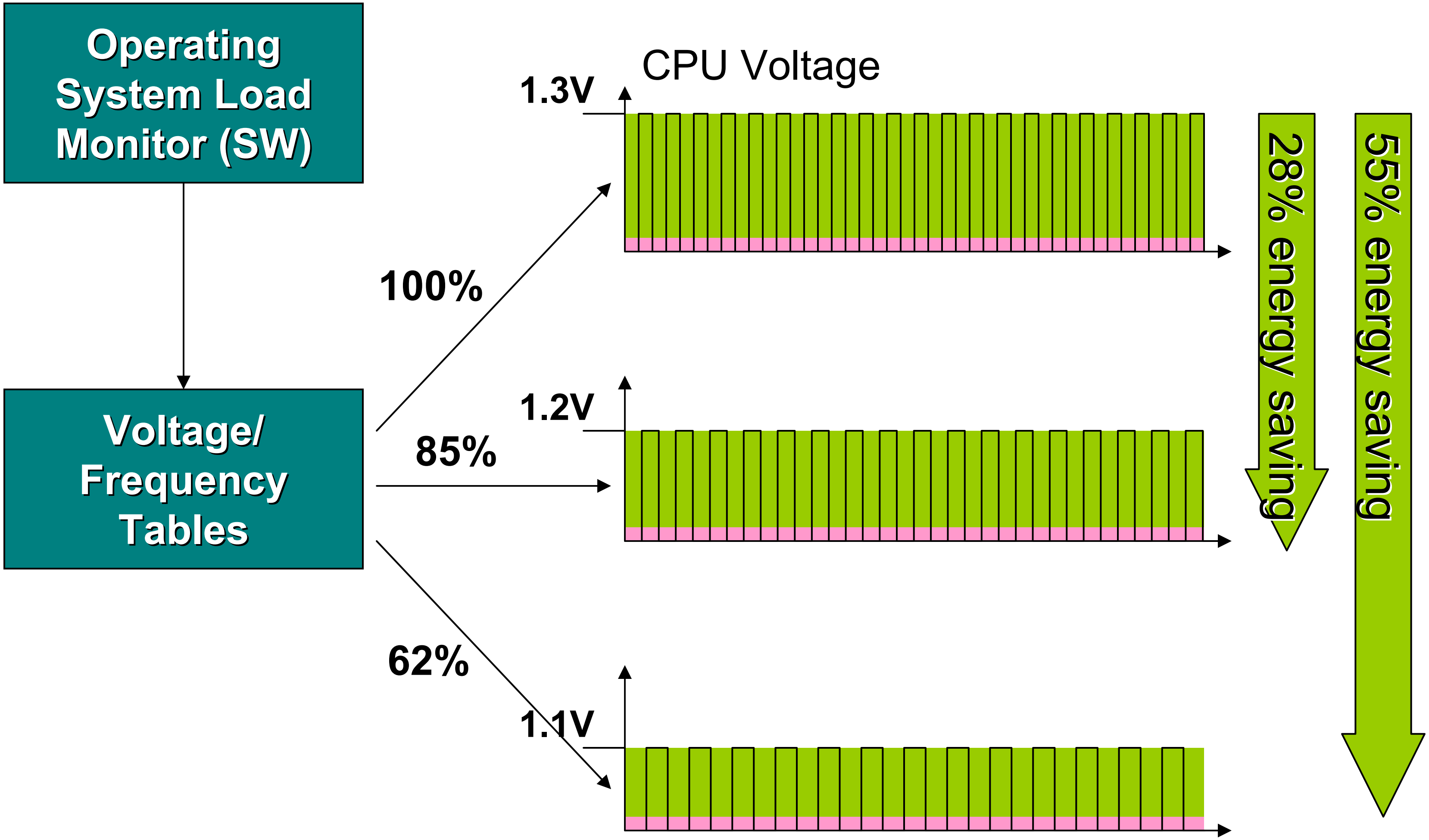
Energy = $C_L * V_{DD}^2 * P_{0\rightarrow 1}$ $f_{0\rightarrow 1} = P_{0\rightarrow 1}$ * frequency

Power = $C_L * V_{DD}^2 * f_{0\rightarrow 1}$

Transition Probability & Frequency Dominate

Dynamic Power Reduction

Voltage & Frequency Scaling



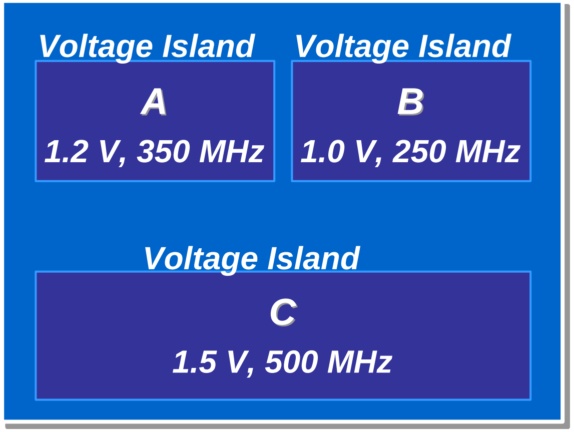
Operating System Load Monitor (SW)

Voltage/ Frequency Tables

Dynamic Power Reduction

Multi-Supply, Multi-Voltage, Dynamic & Adaptive Voltage Scaling

(1)

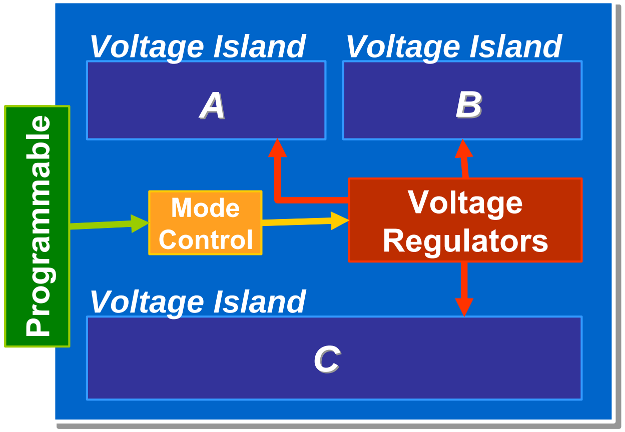


Multiple Supply

Multi-Voltage (MV) Islands

- Voltage areas with fixed, single voltages

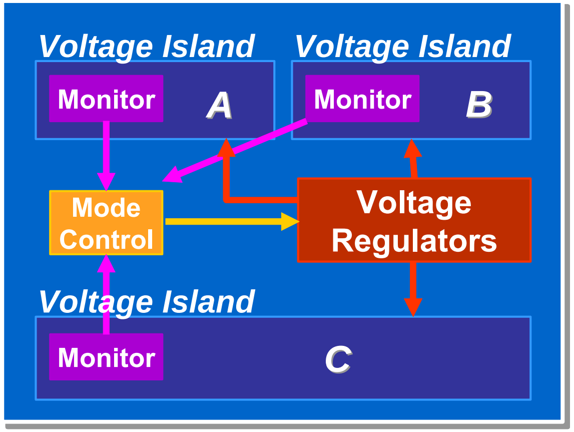
(2)



Dynamic Voltage Scaling (DVS)

- Voltage areas with multiple, but fixed voltages
- Software controlled

(3)

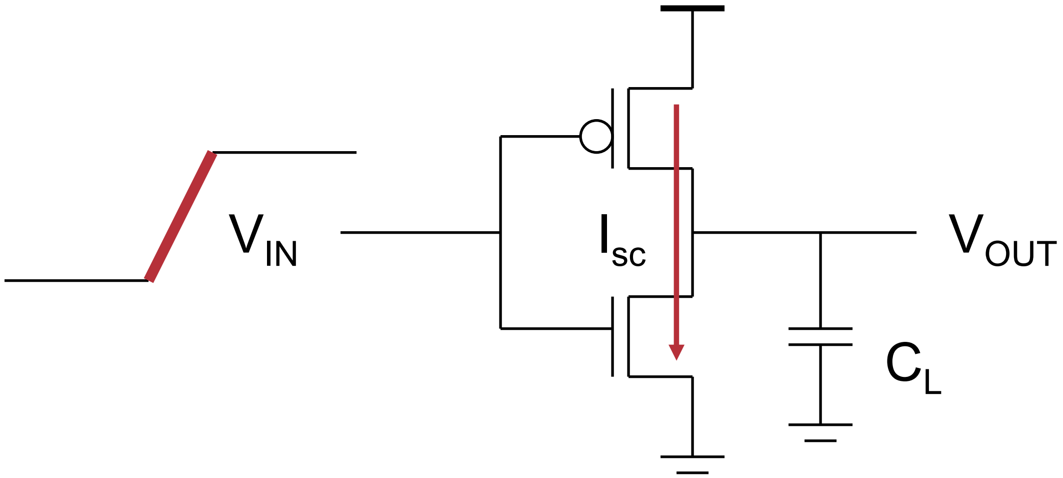


Adaptive Voltage Scaling (AVS)

- Voltage areas with variable V_{DD}
- Software controlled

Short-Circuit Power Consumption

Depends on Input Slope, Output Load & V_{DD}

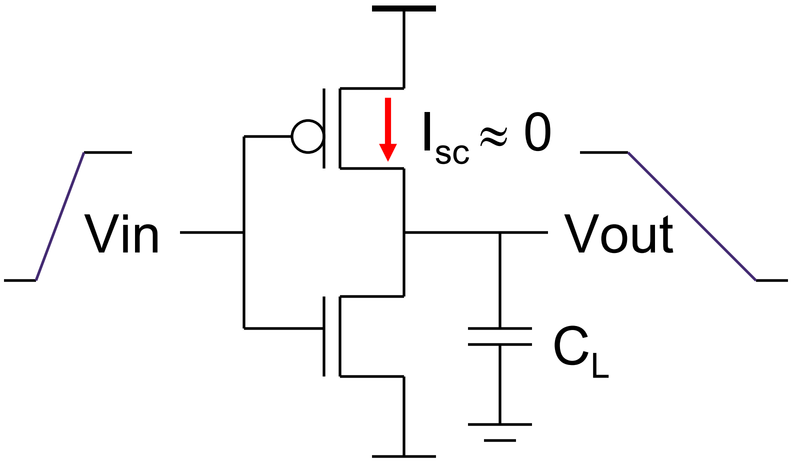


Energy = $t_{sc} * V_{DD} * I_{sc} * P_{0\rightarrow1}$ $f_{0\rightarrow1} = P_{0\rightarrow1}$ * frequency

Power = $t_{sc} * V_{DD} * I_{sc} * f_{0\rightarrow1}$

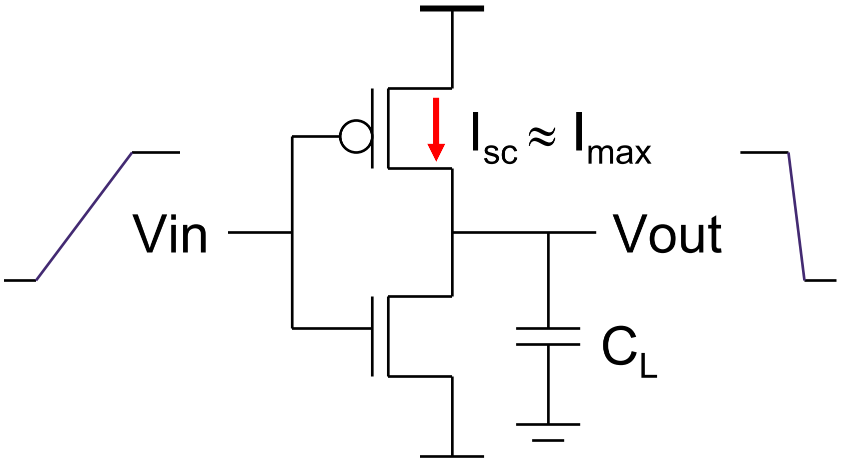
Today, Interconnects Capacitance Dominates

Impact Of C_L



Large capacitive load

Output fall time significantly larger than input rise time.



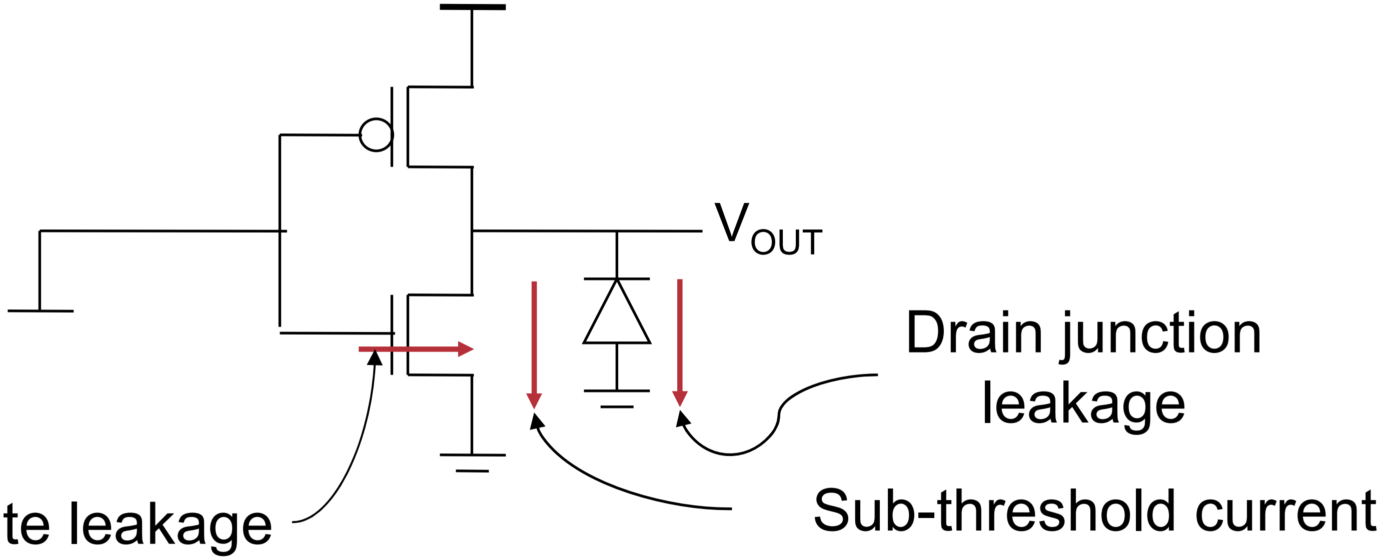
Small capacitive load

Output fall time substantially smaller than input rise time.

Negligible When $V_{DD} < V_{THn} + |V_{THp}|$, Slope Engineering

Leakage Power Consumption

Depends on Sub-Threshold Current & V_{DD}

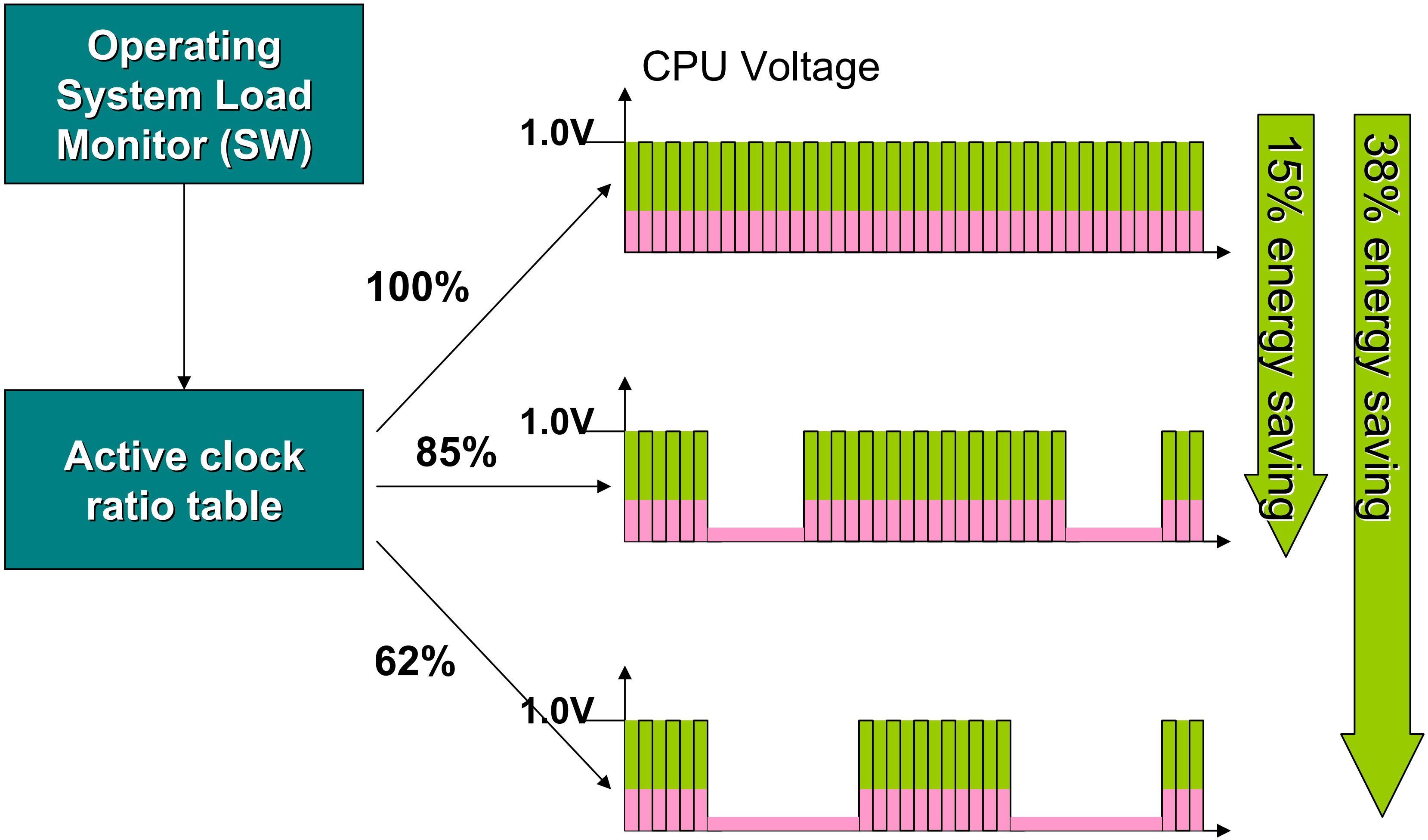


Energy = Power = $V_{DD} * I_{leakage}$

Today Sub-Threshold Current, But Gate Leakage Coming

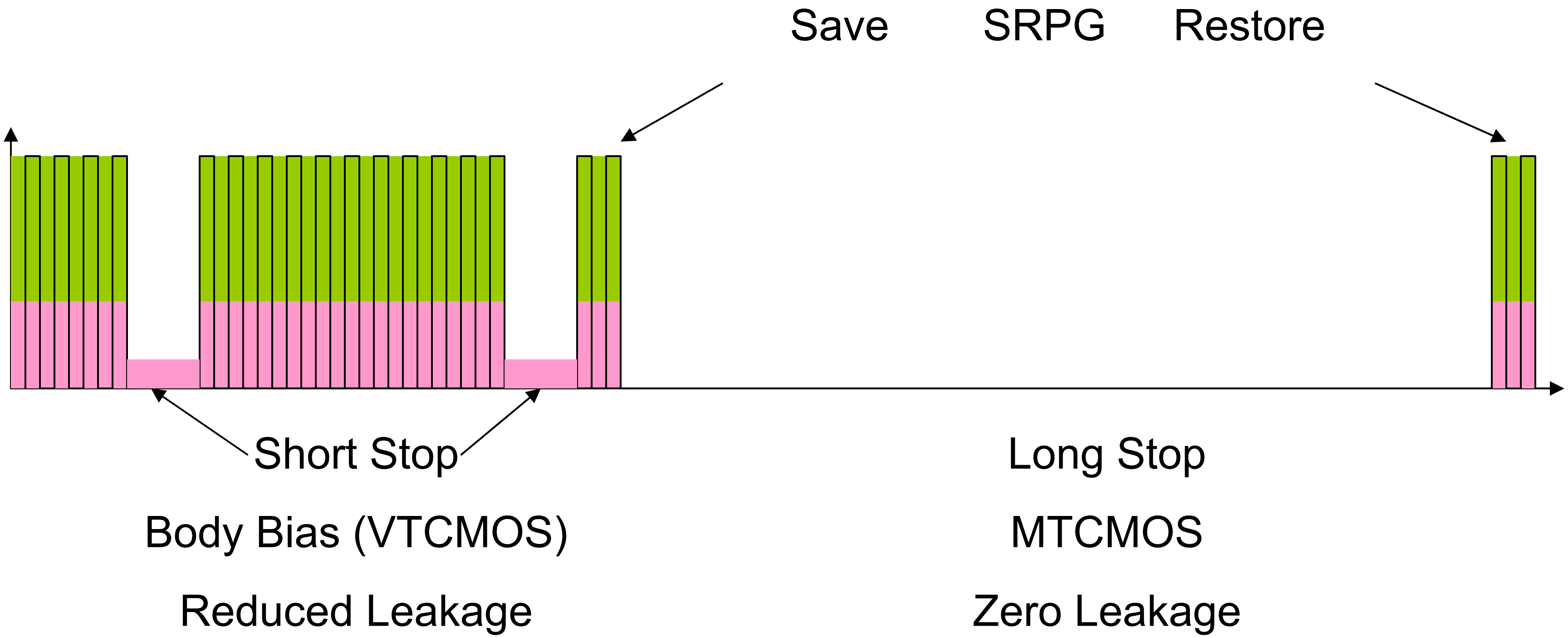
Leakage Power Reduction

Body Bias (VTCMOS) and (SRPG)/MTCMOS

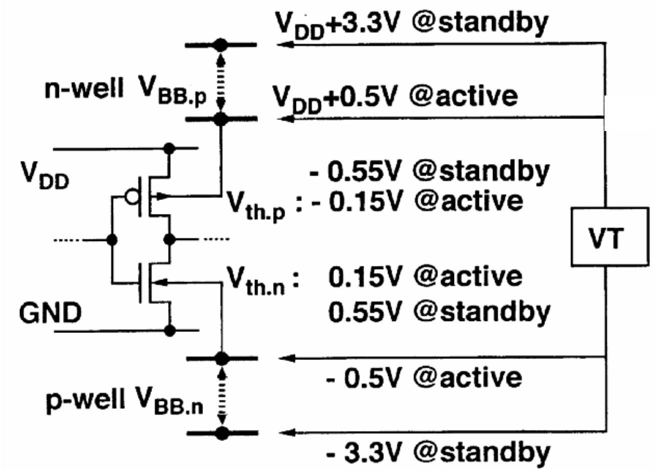
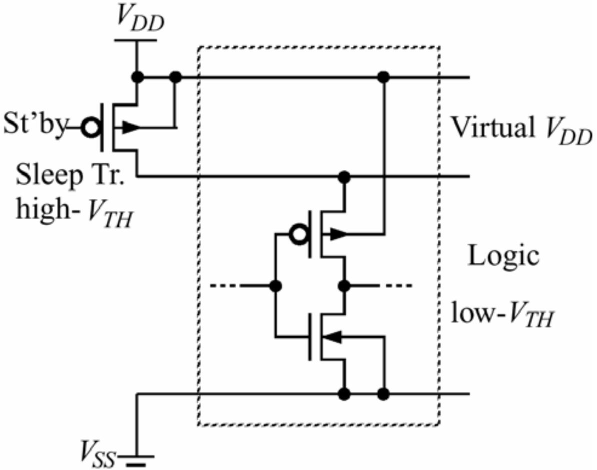


Leakage Power Reduction

Body Bias (VTCMOS) vs. (SRPG)/MTCMOS

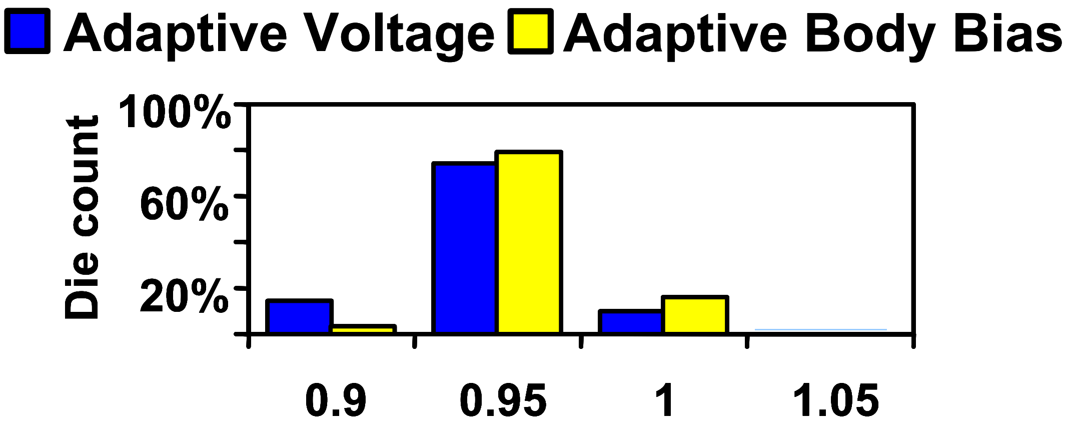


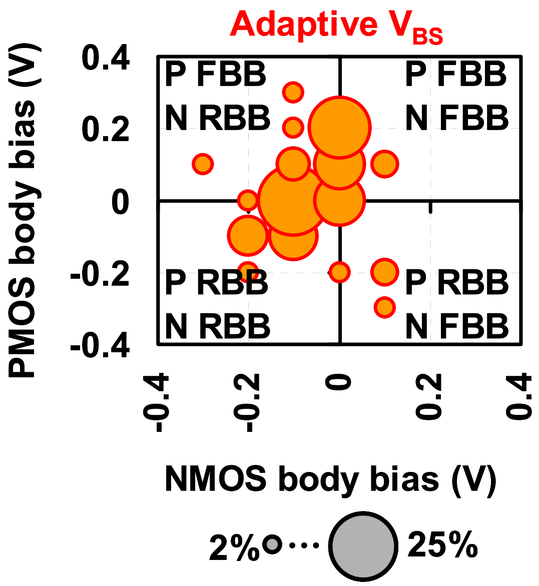
Body Bias (VTCMOS) Vs. MTCMOS

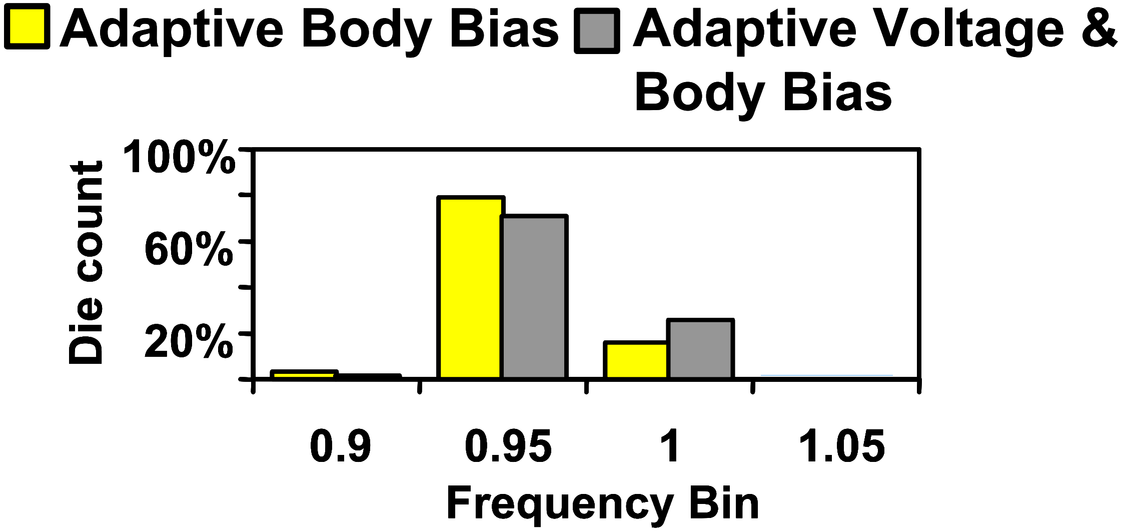
Body Bias (VTCMOS) **V_{TH} Control With Substrate Bias**	**MTCMOS** **On/Off Control of Internal V_{DD}/V_{SS}**
 V_{DD}+3.3V @standby n-well $V_{BB.p}$ V_{DD}+0.5V @active V_{DD} - 0.55V @standby $V_{th.p}$: - 0.15V @active VT $V_{th.n}$: 0.15V @active 0.55V @standby GND - 0.5V @active p-well $V_{BB.n}$ - 3.3V @standby	 V_{DD} St'by Sleep Tr. high- V_{TH} Virtual V_{DD} Logic low-V_{TH} V_{SS}
Needs Circuit Development	**Conceptually Simpler**
Compensate ΔV_{TH} Fluctuation	**Compensate ΔV_{TH} Fluctuation**
IDDQ Test	**IDDQ Test**
No Serial MOSFET	**Large Serial MOSFET**
Conventional EDA Tools	**Conventional EDA Tools**
Re-Use of Existing Design	**Retention Registers**
Triple Well Desirable	**Conventional Well structure**

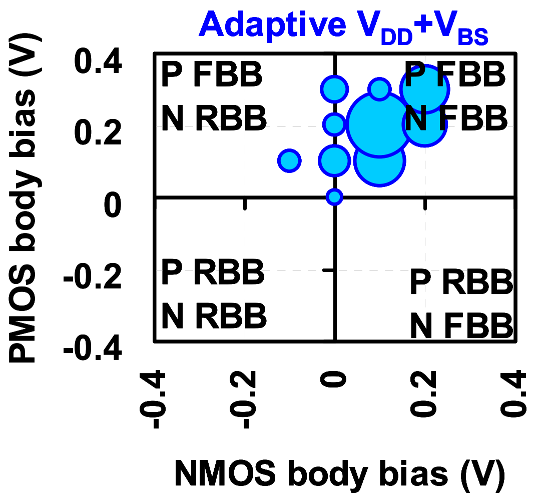
Source: J. Rabaey, UCB 2005

Adaptive Voltage Scaling Vs. Adaptive Body Bias





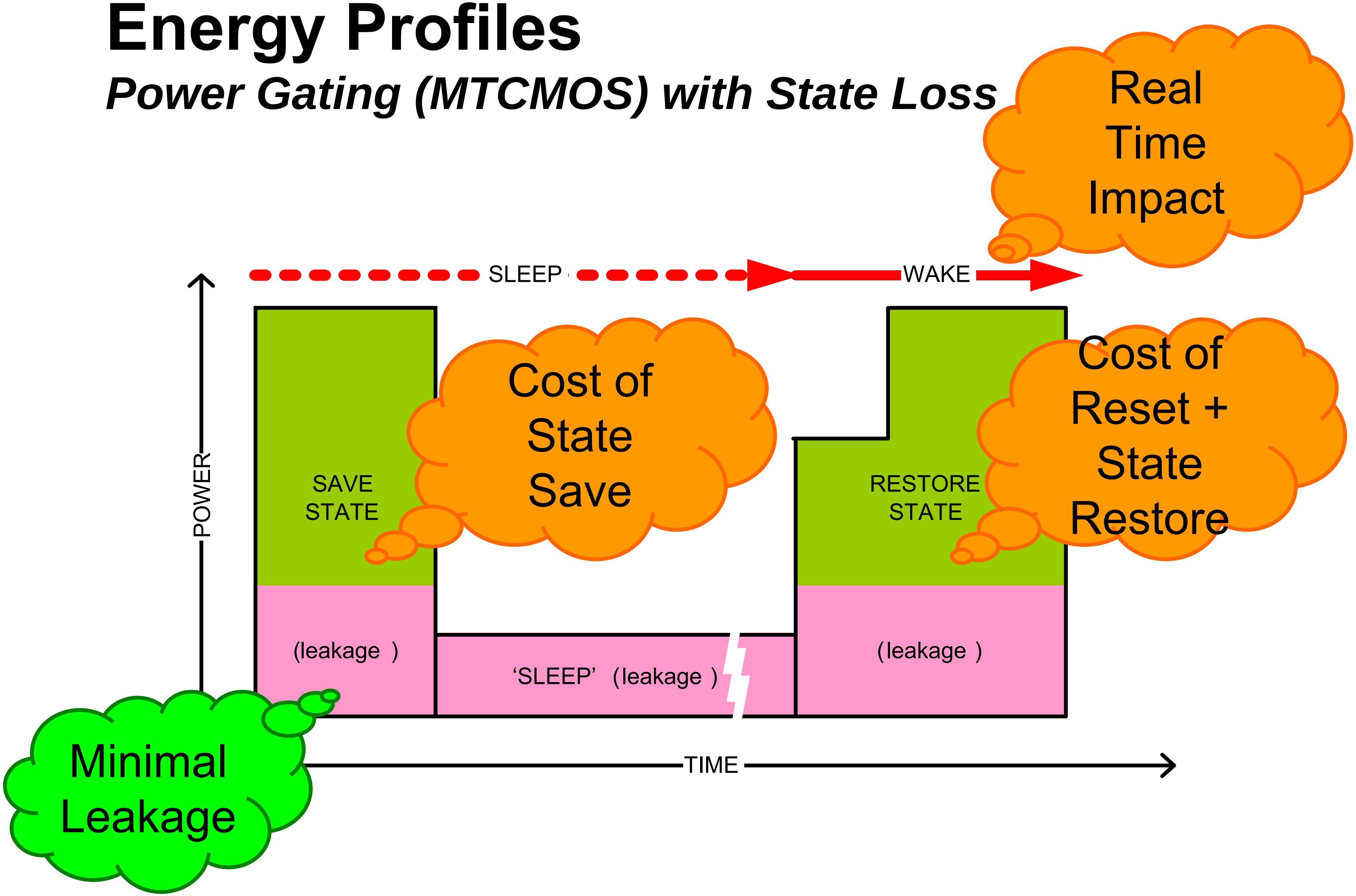




Source: S. Borkar, Intel 2004

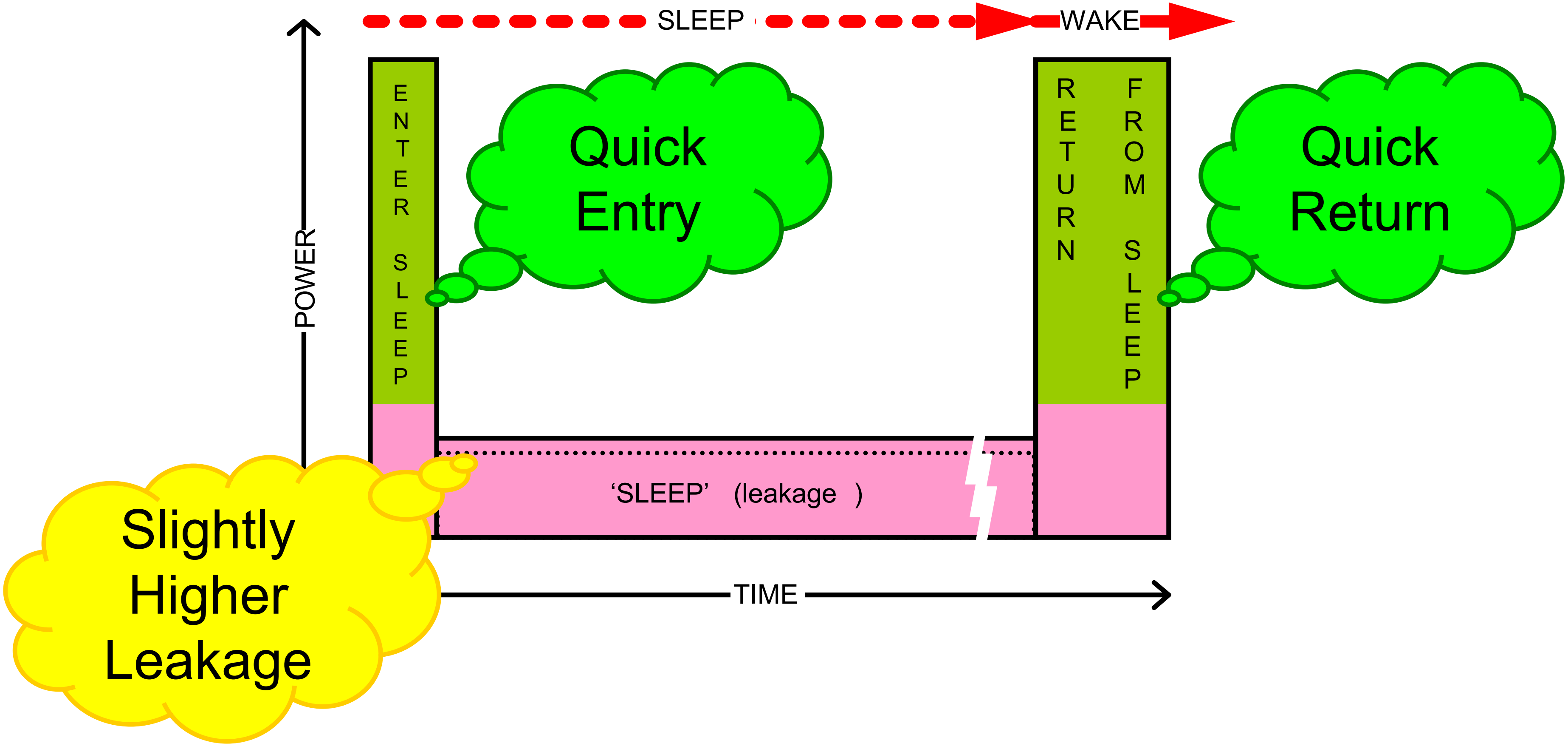
Energy Profiles

Power Gating (MTCMOS) with State Loss



Energy Profiles

Power Gating (MTCMOS) with State Retention



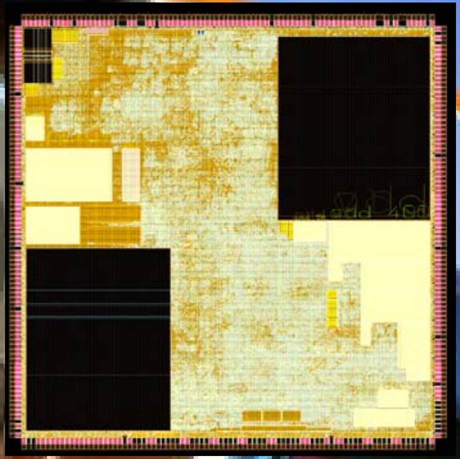
Design For Low Power Helps!

Generations Of Low Power, First Time Silicon Success

2003

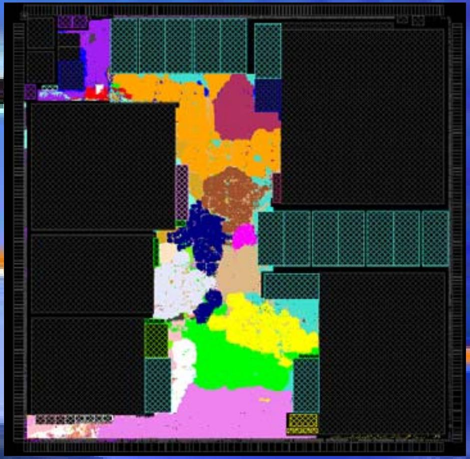
- 130 nanometers
- Clock Gating
- Multi-V_{TH}

2005



- 130 nanometers
- Clock Gating
- Multi-V_{TH} & Multi-Supply

2006



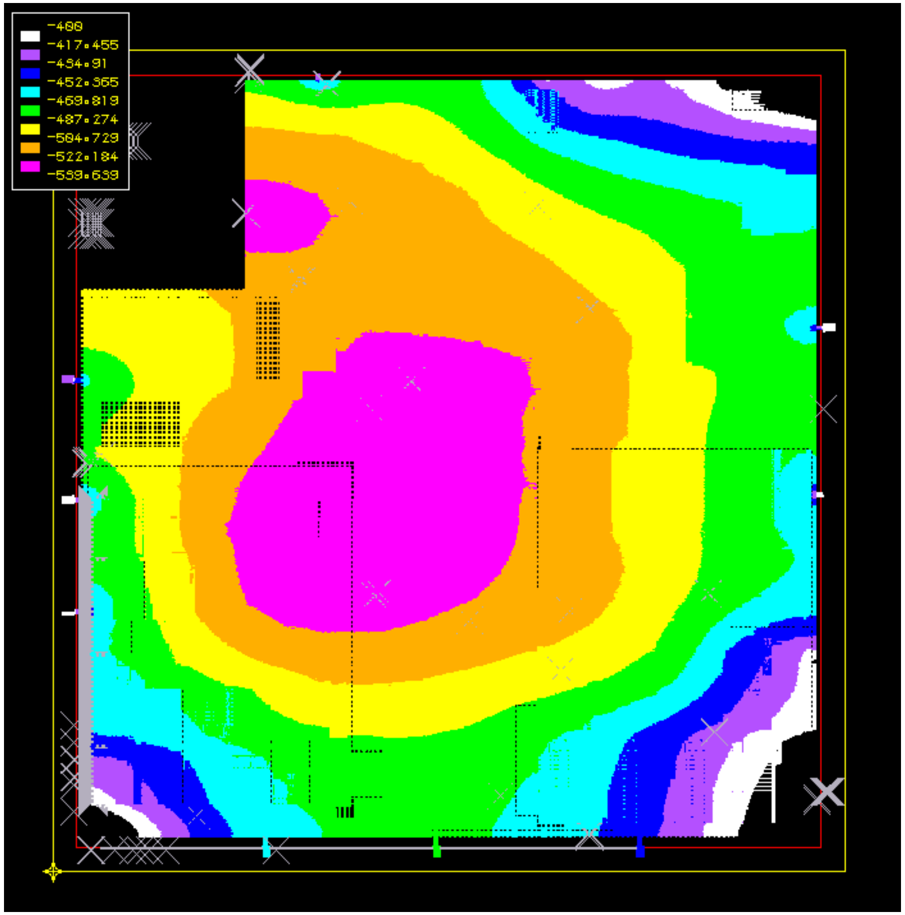
- 90 nanometers
- Clock Gating
- Multi-V_{TH} & Multi-Supply

2007

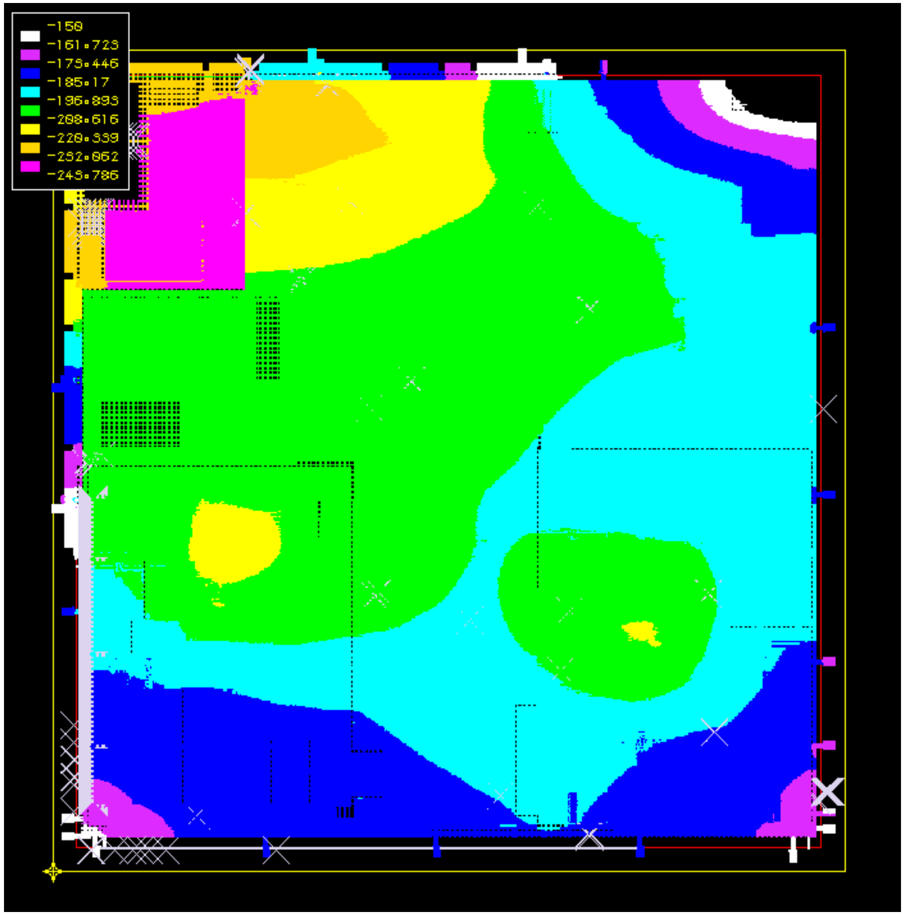
- 65 nanometers
- Clock Gating
- Multi-V_{TH} & Nested Multi-Supply

Design For Low Power Helps!

Voltage Drop in Implementation & Sign-Off



Without packaging parasitics - Range (-400mV, -540mV)

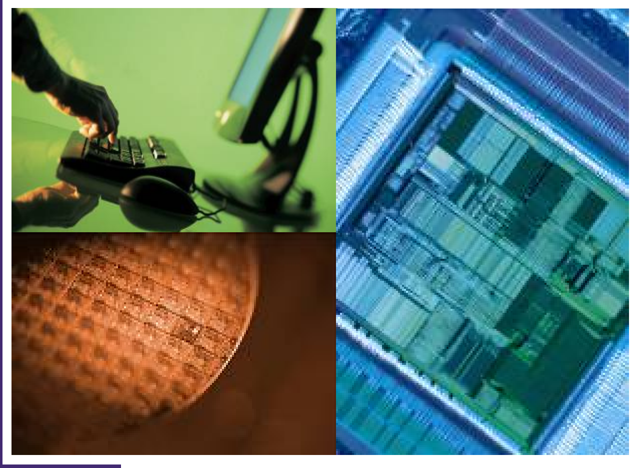


With packaging parasitics + C_{EXT} (330nf) - Range (-150mV, -243mV)

Source: European IDM, Multimedia Application, 130nm

The Future Is Low Power

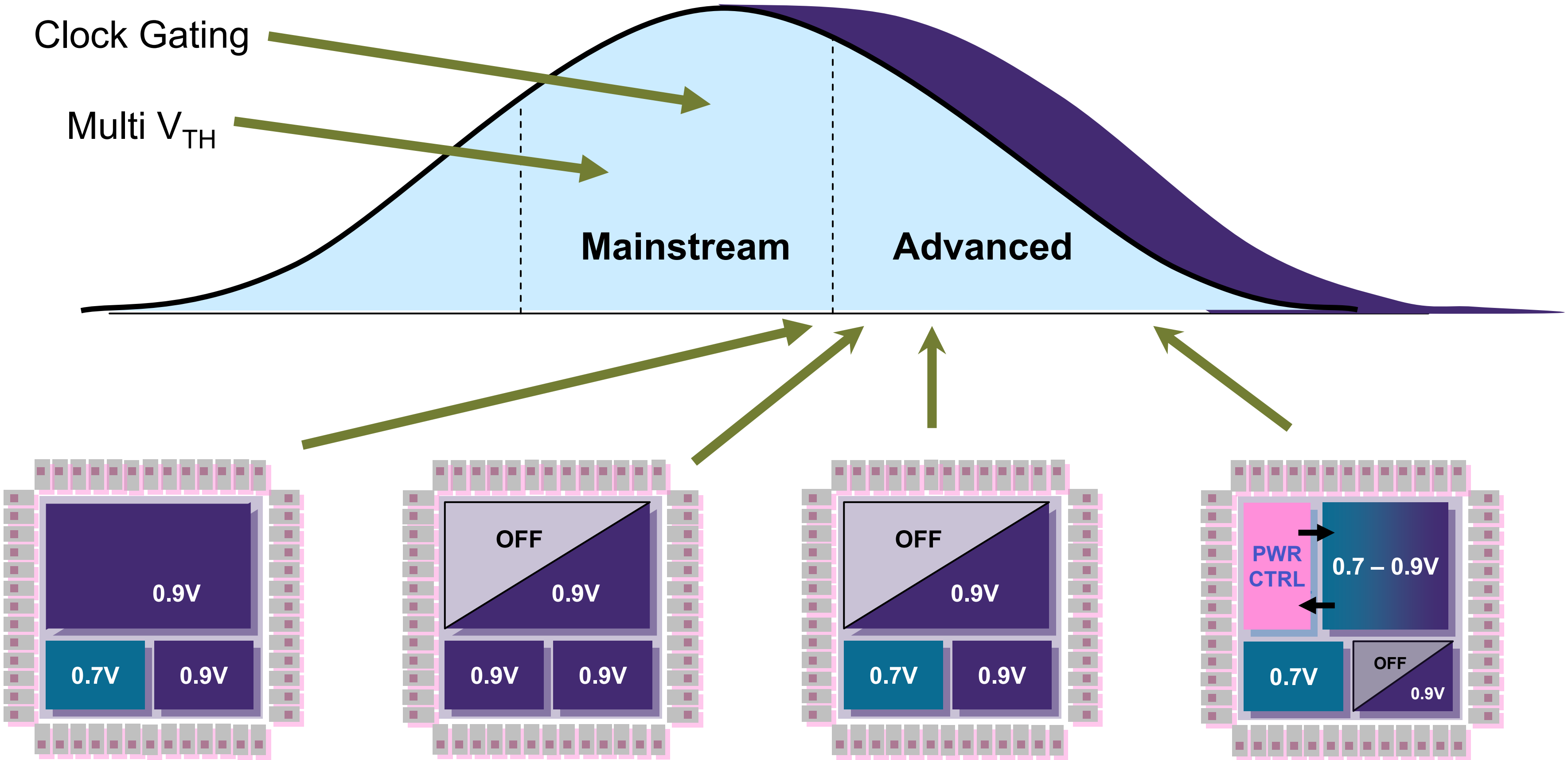
T. W. Williams, Ph.D.
Synopsys Fellow



Predictable Success

Predictable Success

Power Management Techniques



Multi-Voltage

Power Gating (MTCMOS)

Multi-Voltage & Power Gating

Dynamic or Adaptive Voltage Scaling and Power Gating