IBM Systems and Technology Group

Why Hot Chips Are No Longer “COOL”

Ray Bryant
Director of PowerPC Products
IBM System and Technology Group

Product Designs – The Market Demands Innovation!

Market Expectations

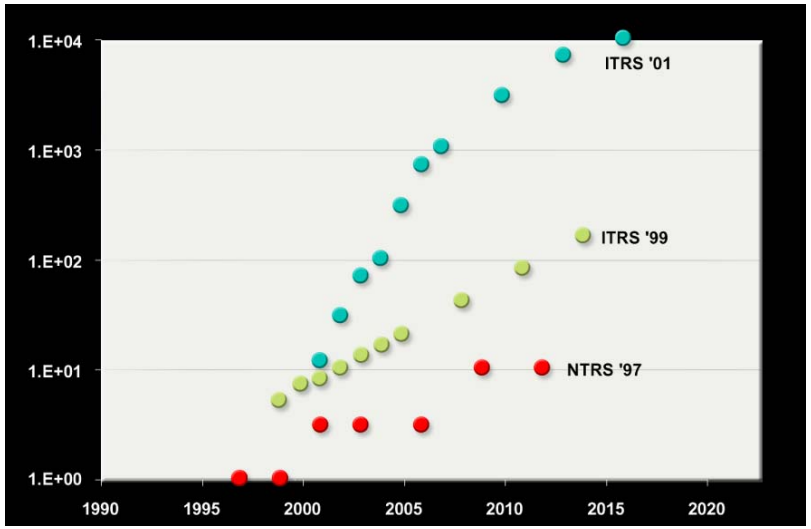
- **New products will be cheaper, lighter, use less power, and do WONDERFUL THINGS**
- **The Designer's Challenge**
  - Technology scaling used to provide 30% performance boost, 30% power reduction, and allowed integration with almost no "penalty"
  - Next generation designs were difficult, but could be mastered using enhancements to the same techniques that succeeded in the prior generations
  - Today, next generation designs are forcing designers to rethink basic architecture, methodology, and tool choices

Old Approach – More Gates Can Solve All Problems!

- **Almost unlimited growth in available transistors allowed architectural options for performance**
  - Multi-threading
  - Out of order branching
  - Predictive look-ahead / write-back techniques
  - Deep execution pipelines
- **Every time an operation is executed that does not create useful work, power is wasted**

Reality: Semiconductor Device Scaling Has Slowed

- **Threshold voltage limited by leakage, not by scaling theory**
- **Supply voltage reduction slowed**
- **"Conventional" gate oxide reaching a physical limit**
- **Gate leakage increasing and becoming a substantial portion of total leakage**



Frequency increases will slow with new technology generations due to power constraints

Market Drivers for Power Reduction

- **Consumer Electronics**
  - Battery operated devices
    - More audio, video, storage functions
  - Home electronics
    - Reduced cost, reduced noise, increased function
- **IT / Servers**
- **Networking / Telecom Access**

IT Reality

- **Servers / clusters**
  - Power density has become a major problem
  - Power distribution / cooling drive significant product cost
  - Rising impact on user satisfaction
    - Operating environment upgrades / costs
    - Degraded user experience
- **Alternate design points must be explored**

Power / Performance Challenge - IT

Virginia Tech's "Tech's X" Supercomputer

- **1100 Dual G5 Apple systems**
- **Ranked #3 Supercomputer in the world @ >10TeraFlops**

Designed for Performance at Achievable Power

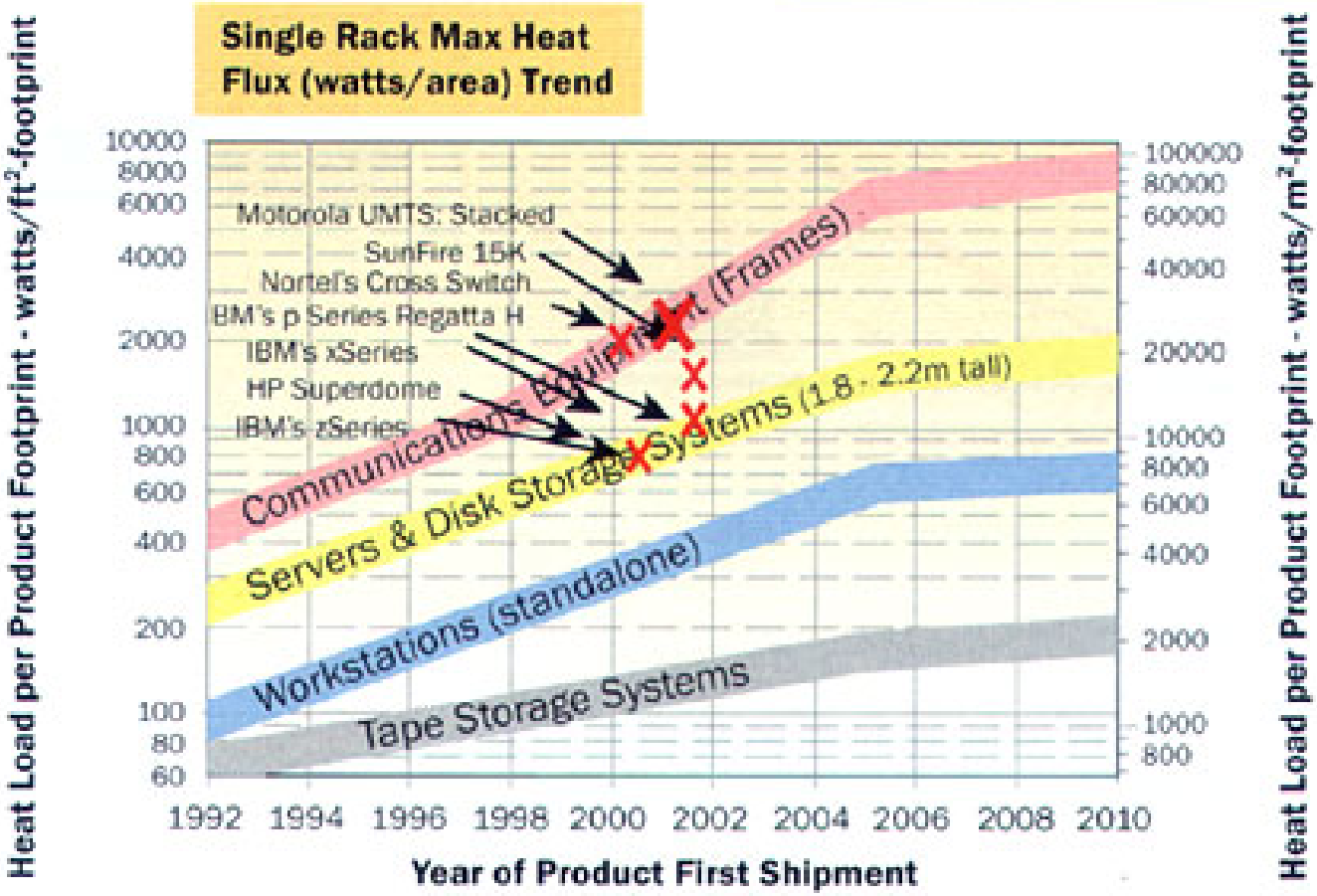
- 1.5 MW consumption
- 2+ million BTUs cooling capacity

Networking Reality

- **Fixed telecom installations were sized based on bandwidth needs 10-15 years ago**
  - Bulky core network switches/routers
  - Limited available space for base station / mobile infrastructure
- **Initial solution was rack mount / stackable system hardware**
- **Latest approach – blade form factors**
- **Current problem – available power service into the telecom closets**

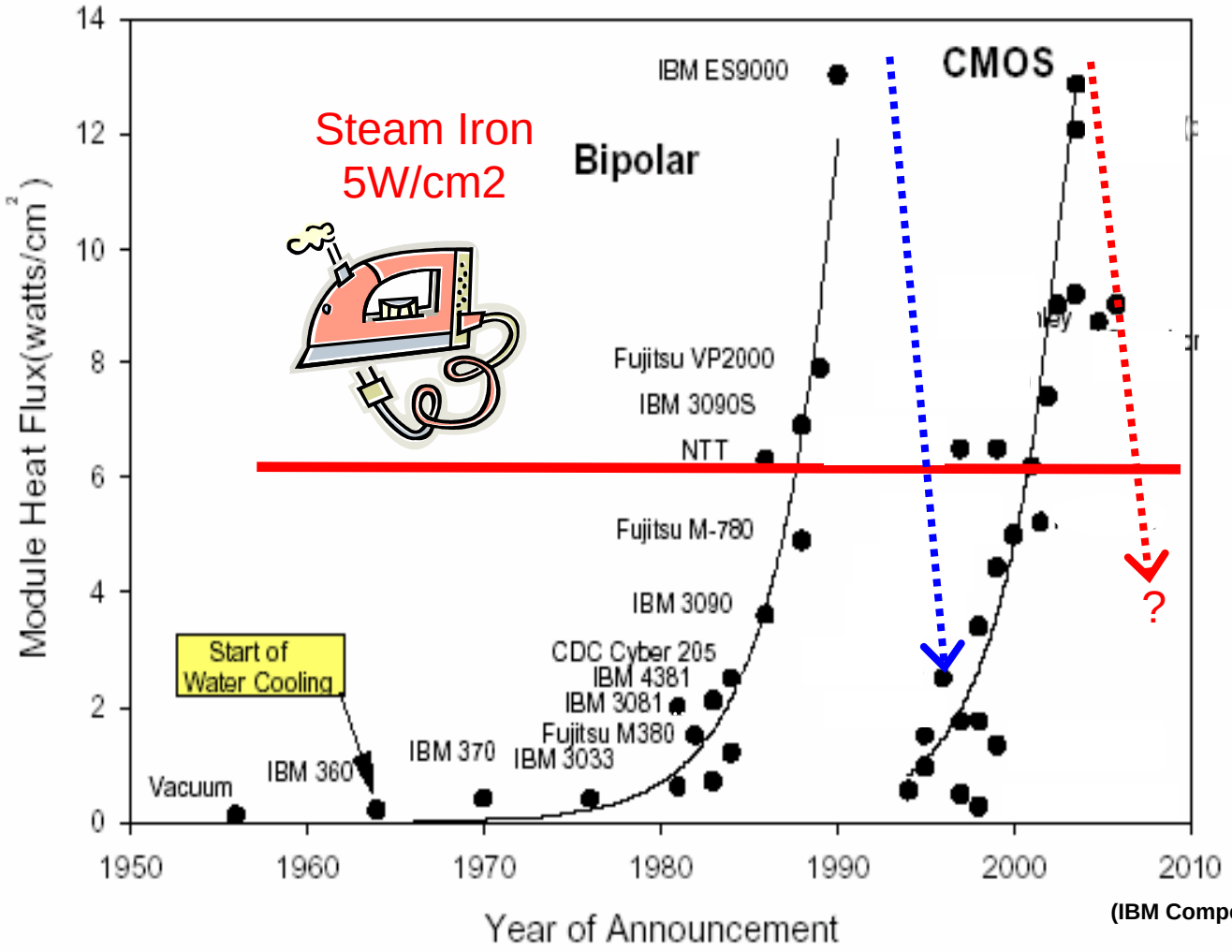
Heat Dissipation Trends

Data Center heat load doubling every 5 years
Telecom rooms quadrupling every 5 years



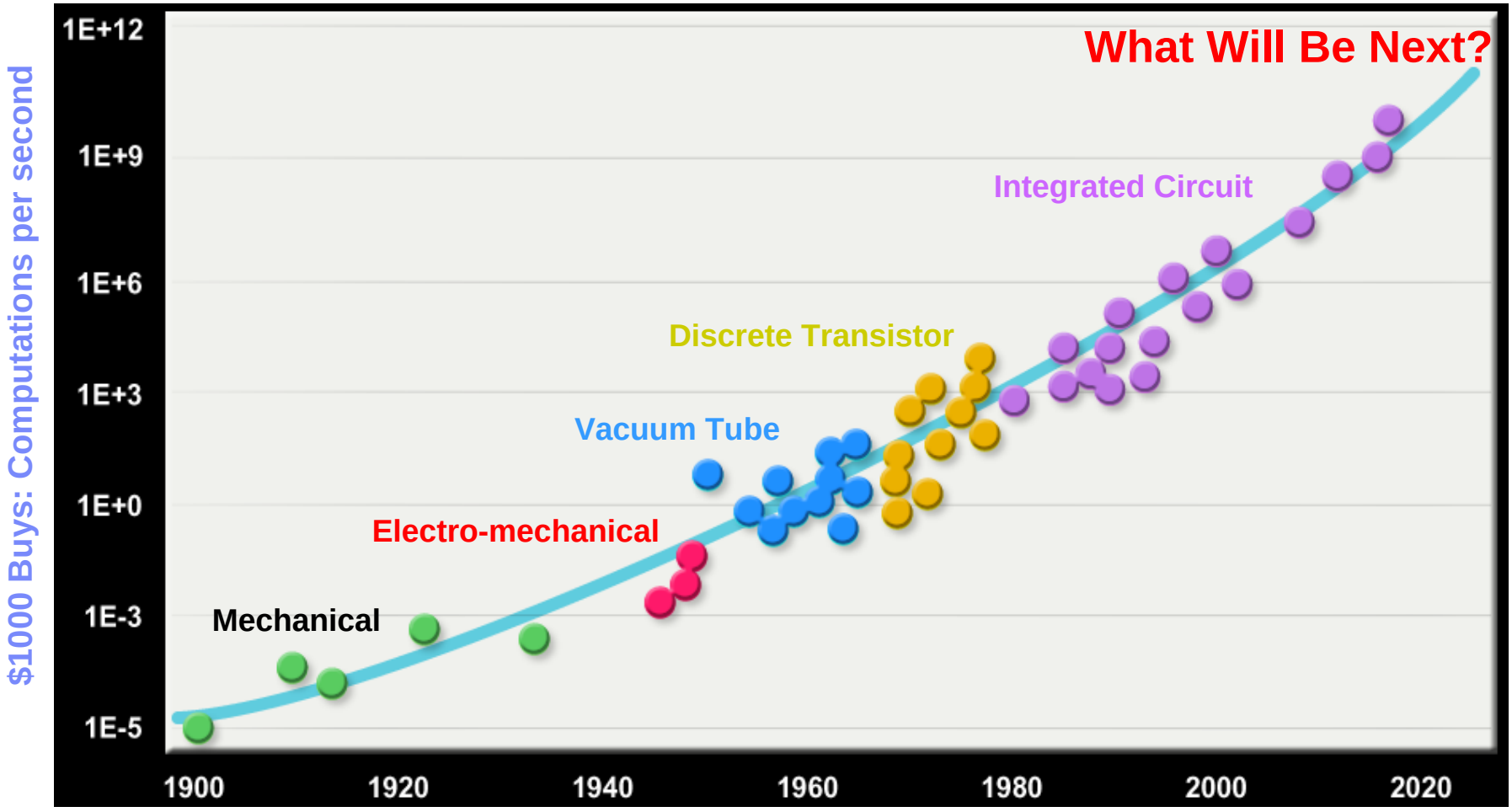
(The Uptime Institute)

CMOS, Once a Breakthrough, Now Needs Innovation



(IBM Competitive Analysis)

IT Leadership Through Technological Innovation



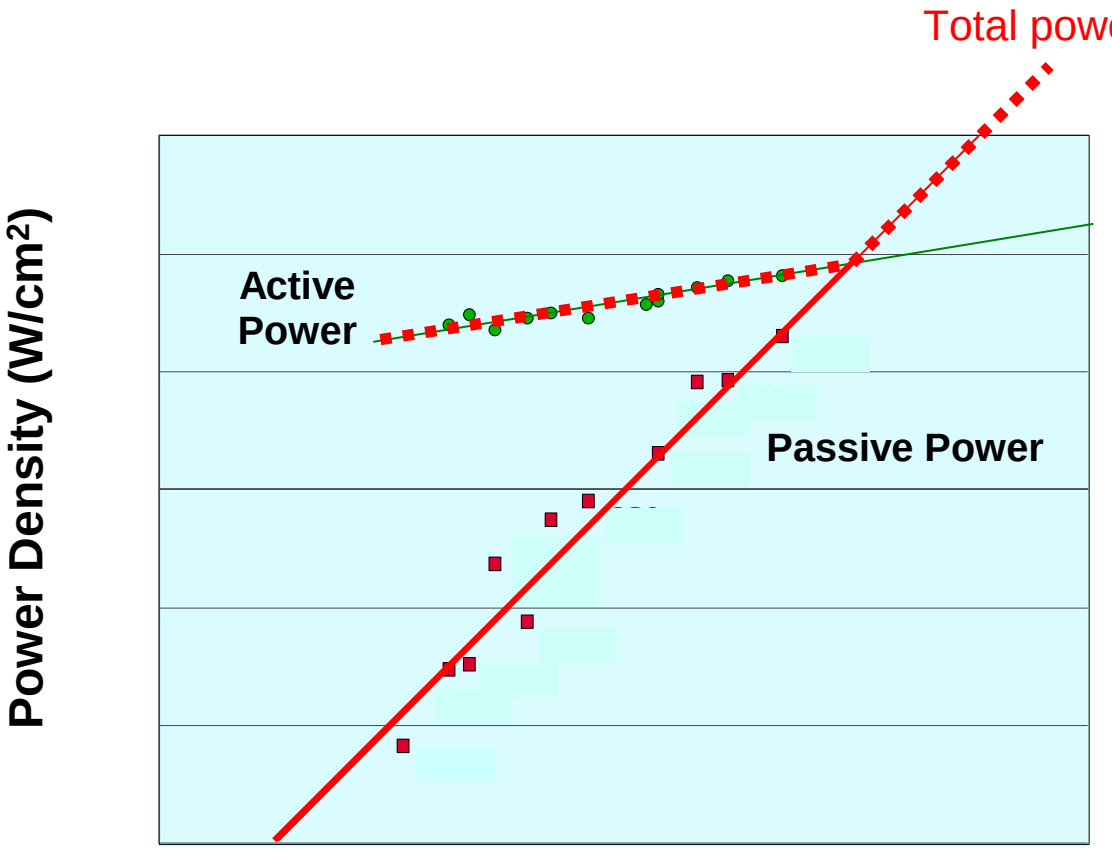
Source: Kurzweil 1999 – Moravec 1998

Traditional Solutions

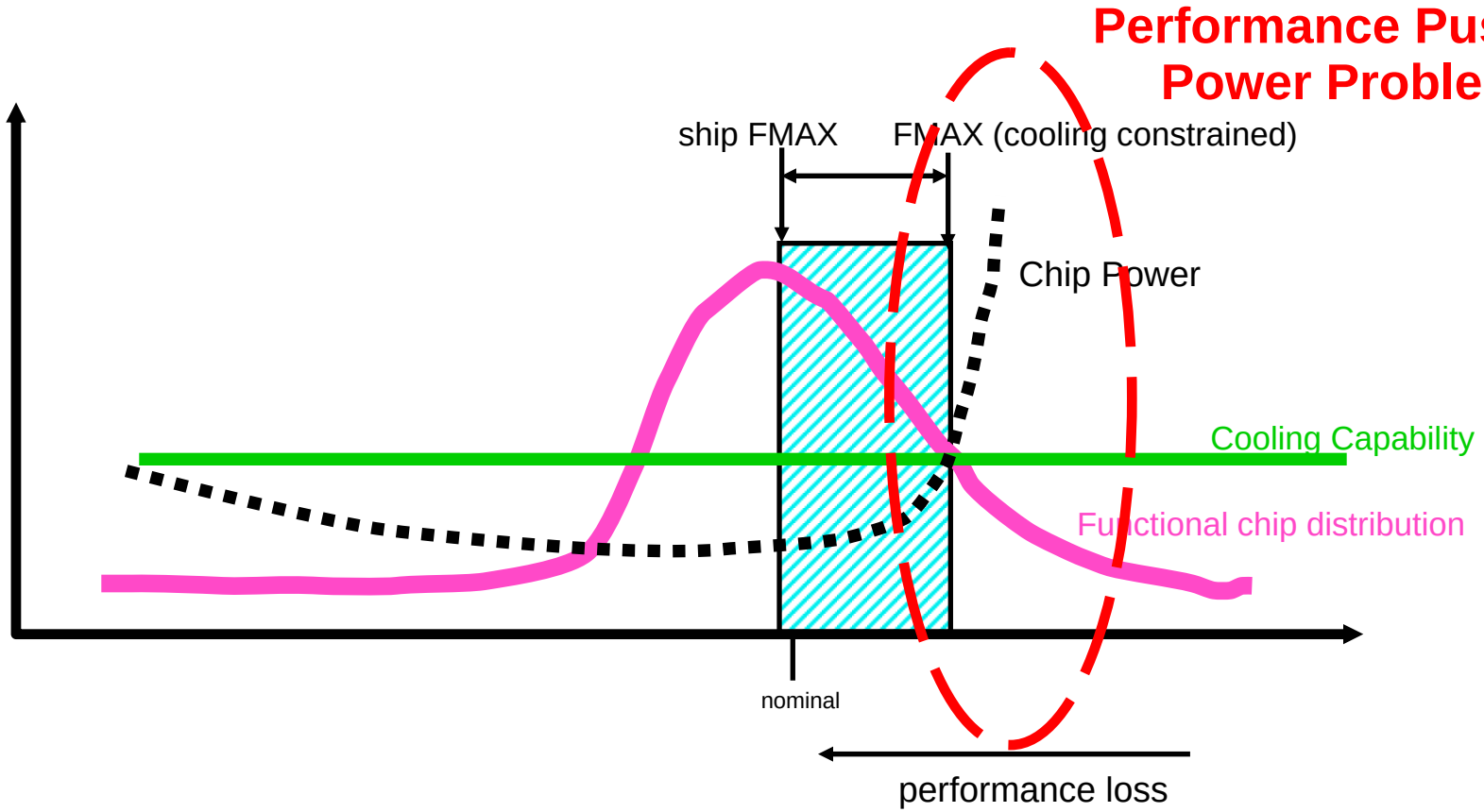
- **Prior ISLPED Focus Areas**
  - Technology advances (materials / process)
  - Voltage scaling
  - Frequency scaling
  - Clock gating
- **System Design Innovation**
  - Sleep mode / nap mode
  - High level OS power monitoring

Challenge: Passive Power Dominates Past 90nm. Technology

- **Fundamental Changes**
  - "Stopping" the chip no longer reduces chip power.
  - One must develop means to literally "unplug" unused circuits.
  - Software must become much more sophisticated to cope with selective shutdowns of processor assets.
  - Scaling produces profoundly different results when attempting to "push" chip speeds



Frequency – Power Balancing Act



Objective: Maximize ship frequency within cooling capability without creating additional product risk!

Designs must optimize “within the box”

- **Chips designs must achieve best performance vs. power dissipation as a primary design goal**
- **Different markets or applications will require unique tradeoffs – workload dependency at a granularity not yet seen**
- **Processor micro architecture must ensure maximum REAL WORK is done by each execution task – this has repercussions on RISC**

Drives requirement for better understanding of the target market and flexibility in design

General Industry Actions

- **Automotive**
  - Sophisticated, closed loop sensor / control systems to manage fuel efficiency
- **Consumer appliances**
  - Advanced electronic controls to manage heating/cooling requirements in air conditioning, cooking, refrigeration, etc
- **Electrical power distribution**
  - Closed loop sense/respond systems to dynamically balance demand, supply, system loss, and switching transients

- **Common Attributes**
  - Energy is only expended when necessary – closed loop control systems becoming mandatory

Electronics Industry Mandate

- **Re-evaluate every facet of design**
  - Device
  - Circuit
  - Micro architecture
  - Sub-assembly
  - System
  - OS / Middleware / Applications SW

- **Goal – manage every electron to create "On-Demand Performance" – and nothing more!**

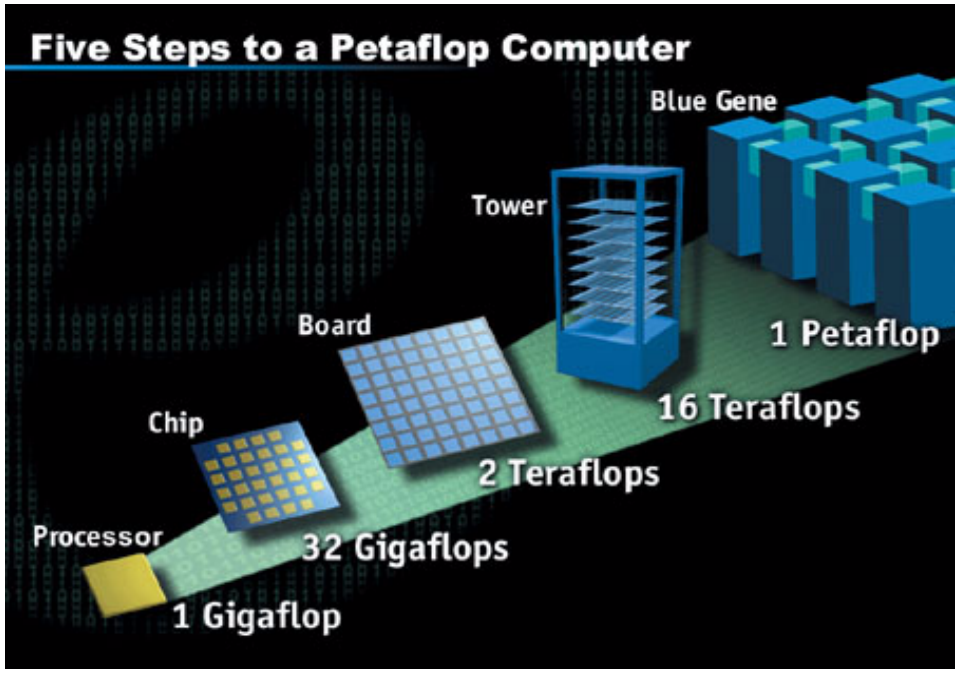
Design Alternative - Example

Blue Gene/L

- **Uses dual core 440 based ASICs**
- **65,535 ASICs sharing 256MB memory each**

Compared to JAMSTEC's Earth Simulator (#1 in June 2004)...

- **>9x faster peak speed (Tflop/s)**
- **~2x memory bandwidth (TB/s)**
- **<1/10th footprint (sq. ft.)**
- **~1/5th total power (MW)**
- **<1/8 the cost (M$)**



Design tradeoff: reduced performance at each single node for significantly reduced power = world class system achievement

New Directions

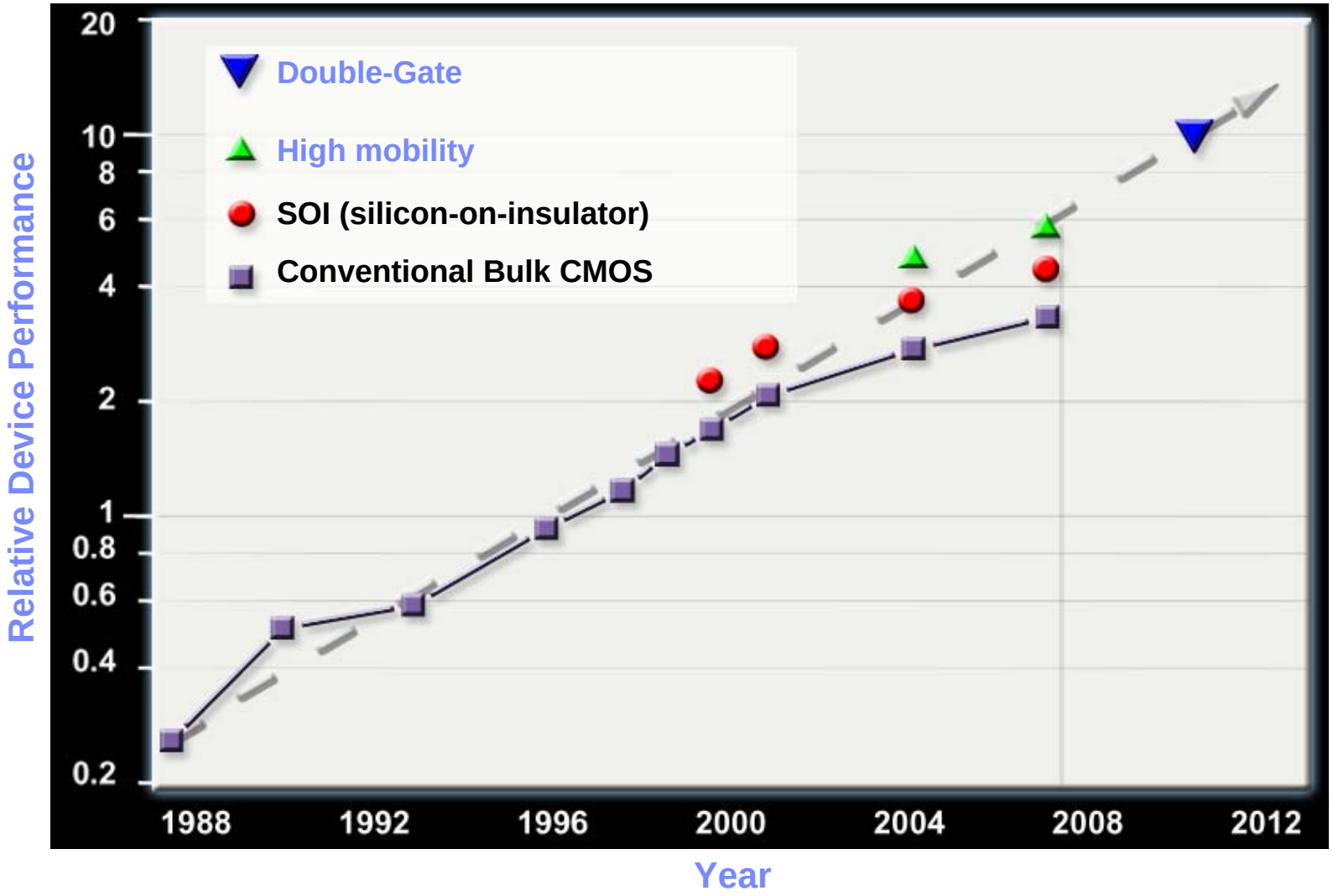
- **Technology**
  - Materials / Process Advances
  - New Device Structures
- **Design Alternatives**
  - Multiple Vt Optimization
  - Fine Grained Voltage Islands
  - Circuit Topology Options
- **Chip core scale-out structure vs. frequency scale-up designs – Impact on System and SW complexity?**
- **Tools and models to balance dramatic increases in design complexity w/o increasing design time or cost**

Low Power Leverage Points - Technology

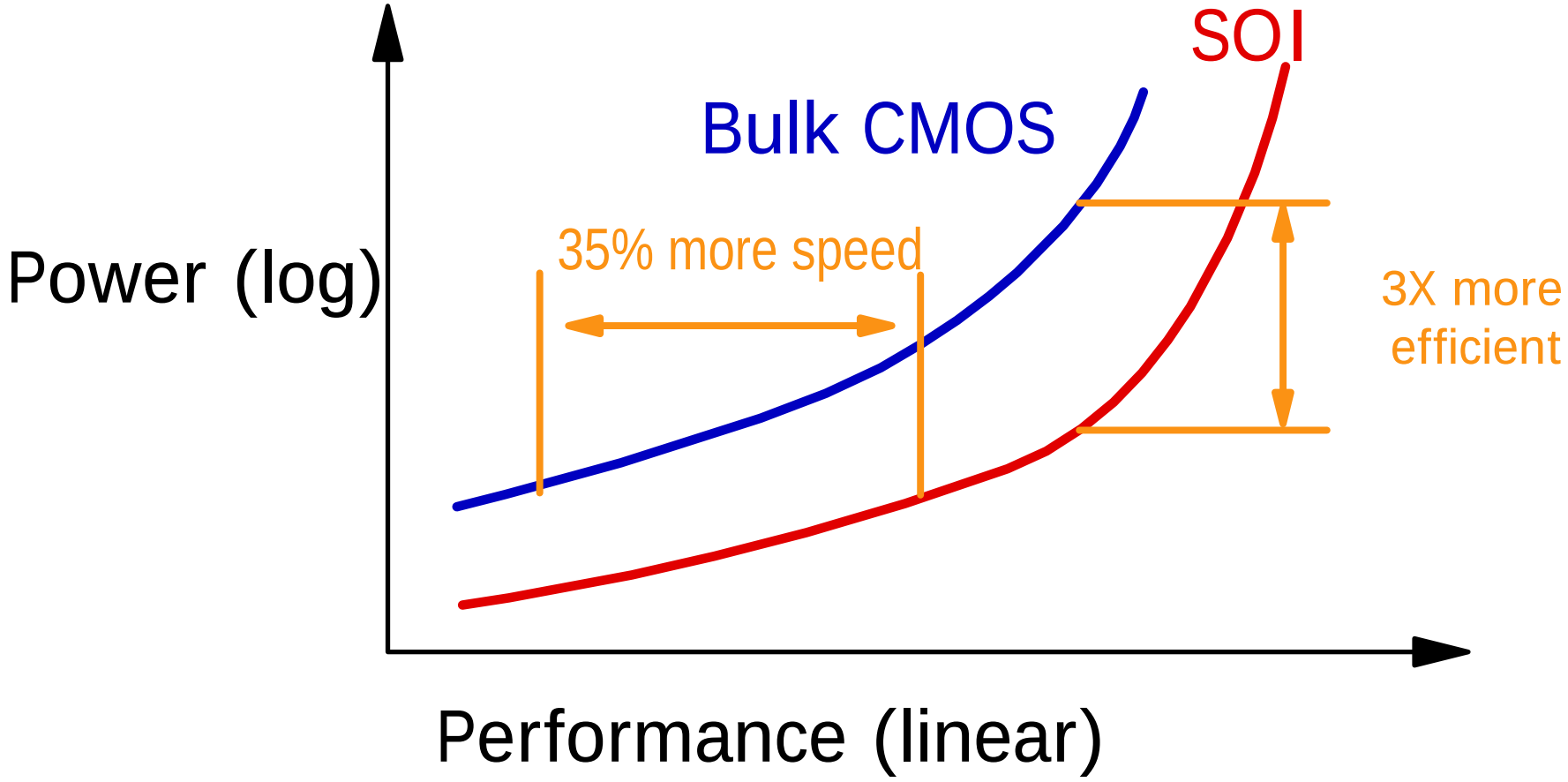
- **Process Technology:**
  - **Leadership power*delay - SOI, strained Si, FINFET**
  - **Low-k BEOL dielectric**
  - **Triple oxide**
  - **Triple well**
  - **High-k, thick gate oxide**
  - **eDRAM**

CMOS Device Performance

New Device Structures are Needed to Maintain Performance

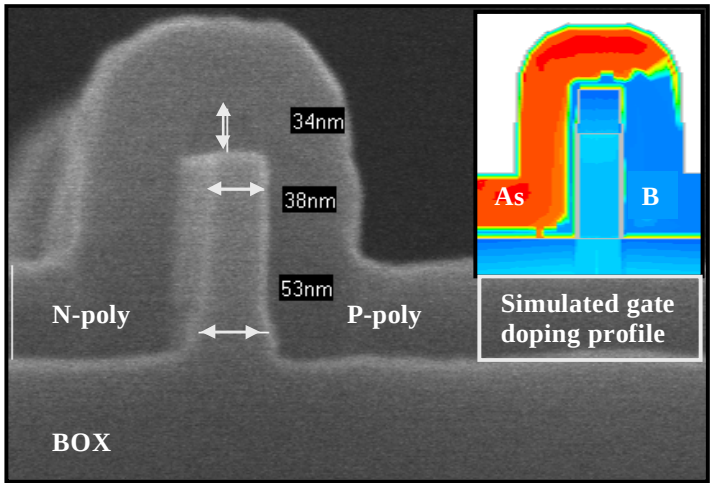


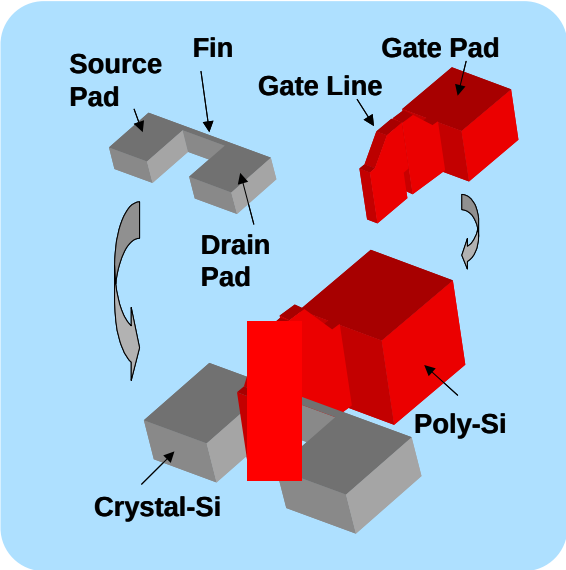
SOI Power vs.. Performance

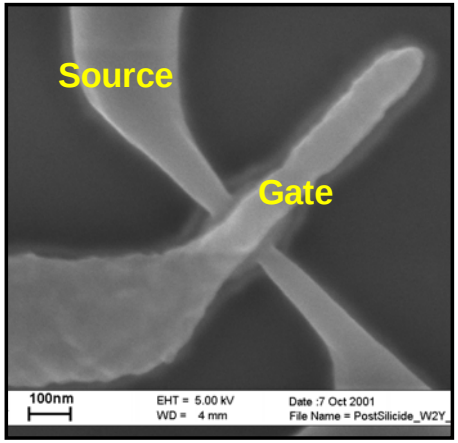


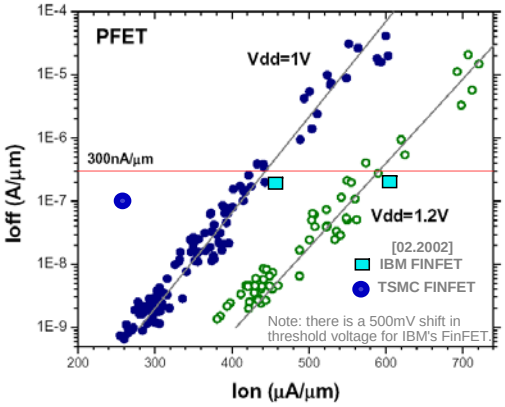
Double-Gated FET-FinFET

- **World's first Vt centered double - gate FinFET for pFET and nFET**
- **Record high current for any double gate**
- **Faster switching time than single gate bulk or SOI**
- **Working CMOS inverters and ring oscillators**



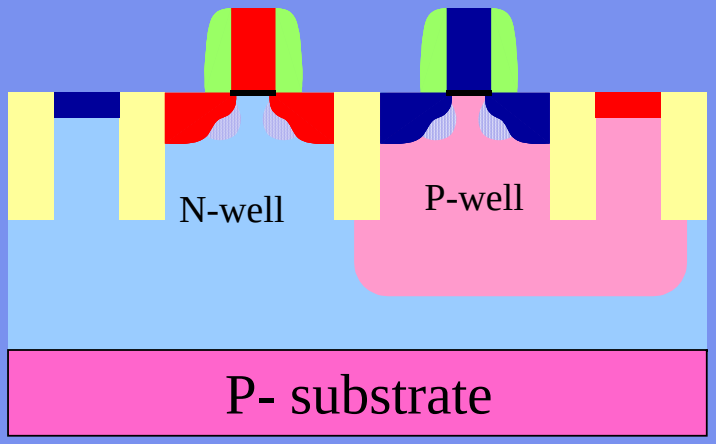


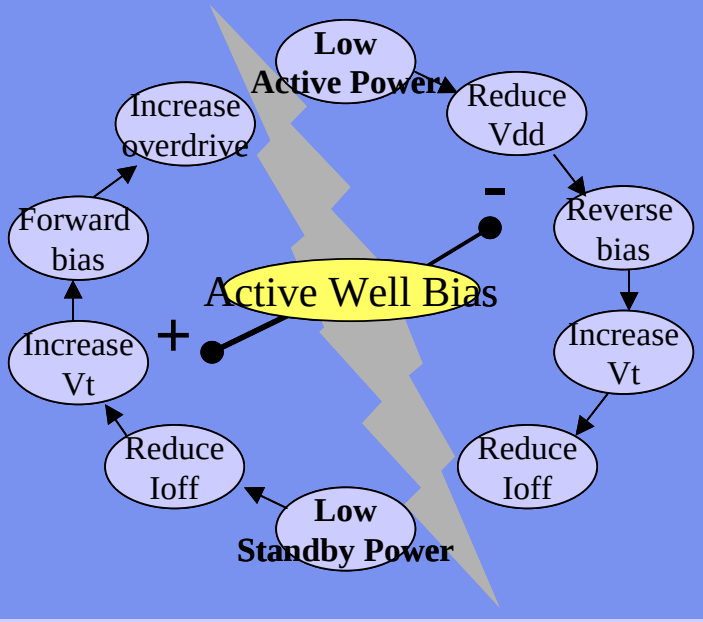




Innovative design of dual gated CMOS structures, and their reduction to practice in MANUFACTURABLE formats, is another area where invention supplants scaling for future generations of CMOS

Active Well for Low Power





- **Active Well CMOS**
  - Forward body bias for performance
  - Reverse body bias for low power
- **Advantages**
  - Up to 35 % speed improvement at same off-state power
  - Better SER protection if used in SRAM cells
  - Isolated well beneficial for analog applications
- **Challenges**
  - Circuit design style change to fully utilize advantages
  - Requires additional well contacts
  - Requires deeper trench isolation and optimized well design

Holistic Design – Atoms to Applications Software

- **Only the simultaneous optimization of materials, devices, circuits, cores, chips, system architecture, and system software provides an effective means to optimize for both performance and power**

- **Innovation will be necessary**
  - Asset virtualization
  - Fine grained clock gating
  - Dynamically optimized multi-threading capability
  - Dynamic reconfigurable circuit technology

- **Open (accessible) architecture for system optimization and compatibility is necessary to allow collaborative innovation**

Example: Power-Performance Tradeoff

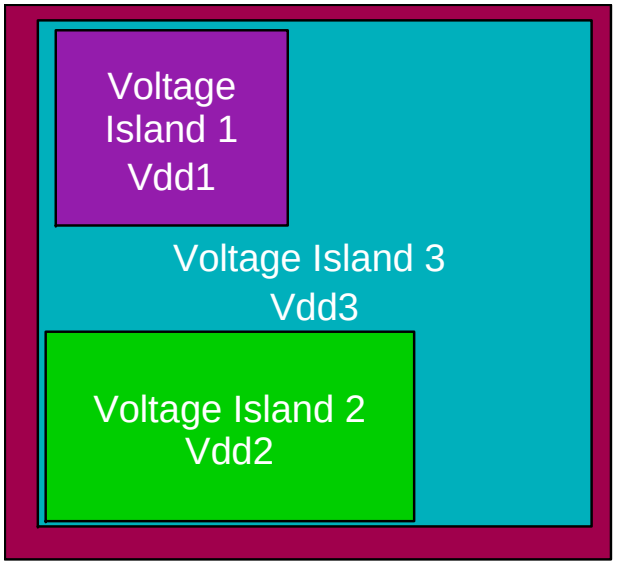
- **Power-performance tradeoff techniques :**
  - Multiple-supply voltages
    - Very effective in reducing power (power α Vdd^2)
    - Overhead of integrating multiple supply voltage PD scheme in methodology
  - Multiple-threshold voltage
    - Very effective in increasing performance with very little impact on dynamic power
    - Exponential increase in leakage power ($P_{leakage} \alpha e^{vdd-vt}$)
  - Gate sizing
    - Very little reduction of power
    - Linear reduction in leakage

Multi-Threshold Library: Limitations

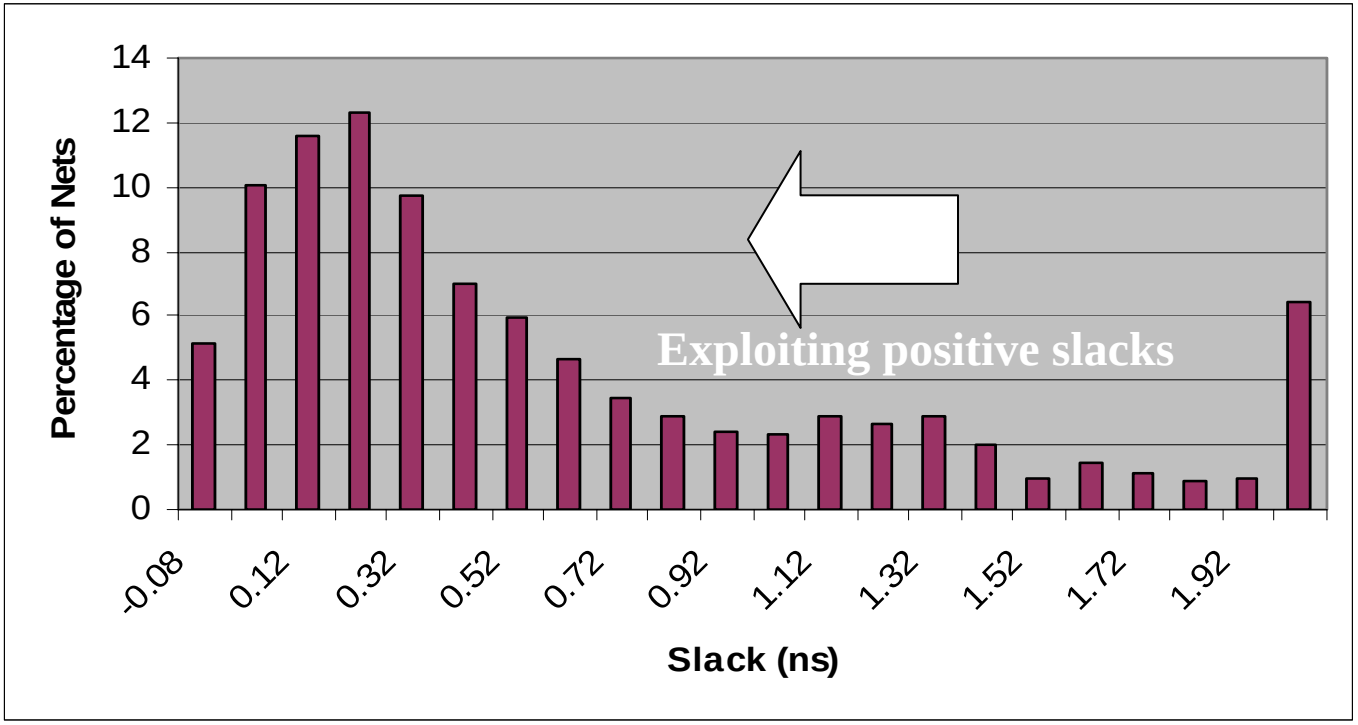
- **Each threshold option requires two additional masks**
  - One for the pfet, one for the nfet
  - Added expense due to wafer processing costs
- **Added design complexity**
  - Each Vt implant is independent
  - Additional source of process variation
    - All transistors won't track exactly the same
    - Must be accounted for in chip design
  - Currently requires additional timing runs – Time/$
- **Does this fit the Success Metrics for your product??**

Why Voltage Islands?

- **Voltage Islands**
  - Regions (logic and/or memory) on chip supplied through separate, dedicated power feeds
  - How to best exploit voltage islands to achieve real-world improvements in both AC and DC power



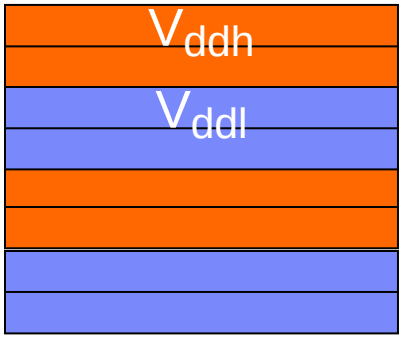
Low Power Opportunities



- **Slack distribution of PPC core after timing closure: over 60% cells have slack > 320ps**
- **Assign non-critical cells to a lower V_{dd} => lower power (both dynamic and leakage)**
- **Idea is easy, but implementation is not**

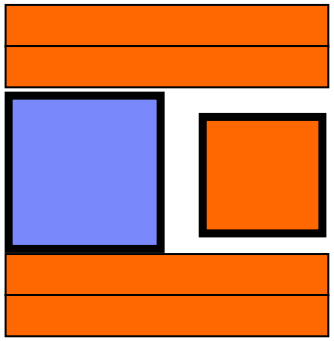
Physical Design Styles with Multiple Voltages

Circuit Rows



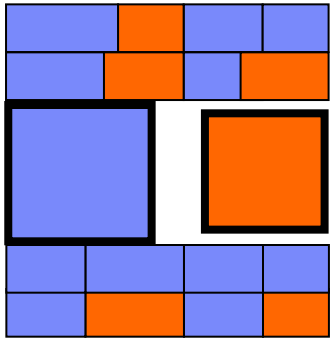
Usami
JSSC'98

Coarse Grained
Voltage Islands



Lackey
ICCAD'02

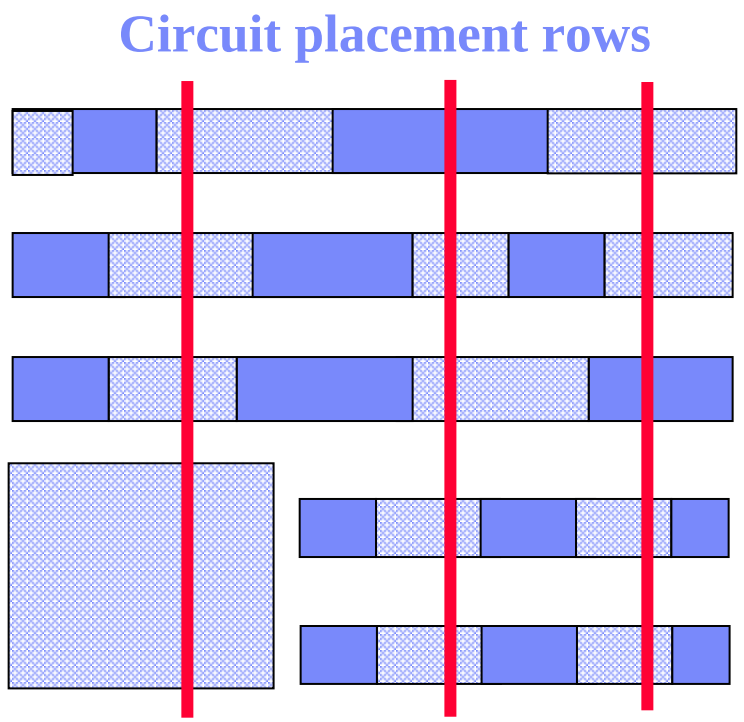
Fine Grained
Voltage Islands



Puri
DAC'03

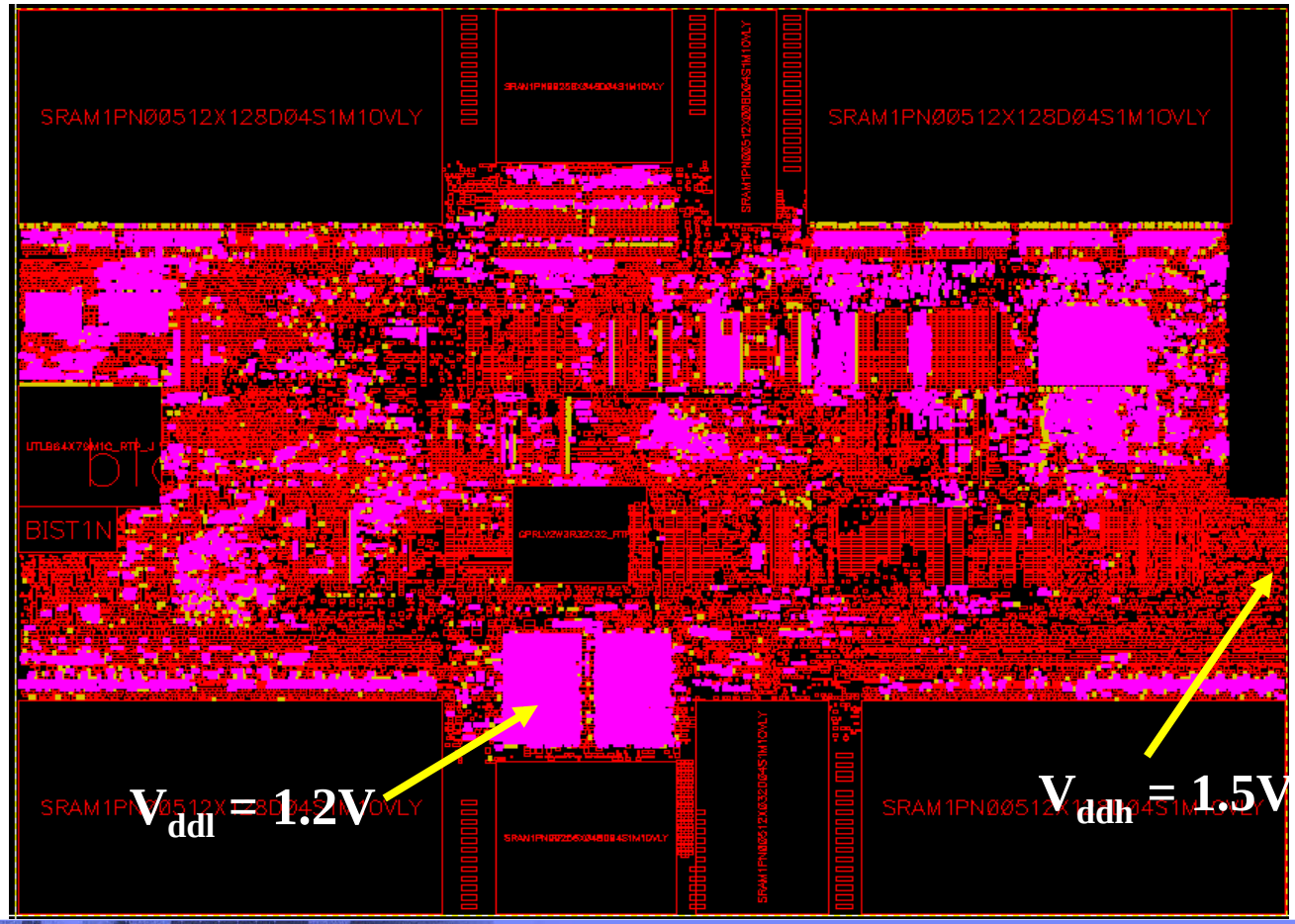
Fine-Grained Voltage Islands

- **What is a fine-grained voltage island?**
  - Identification of circuits or groups of circuits that based upon physical proximity can share a common power connection
    - In the theoretical limit, this can be a single circuit
- **Why is this important?**
  - Placement, especially in bit stacked physical architectures, such as microprocessors, is critical to timing, wiring, area, and power



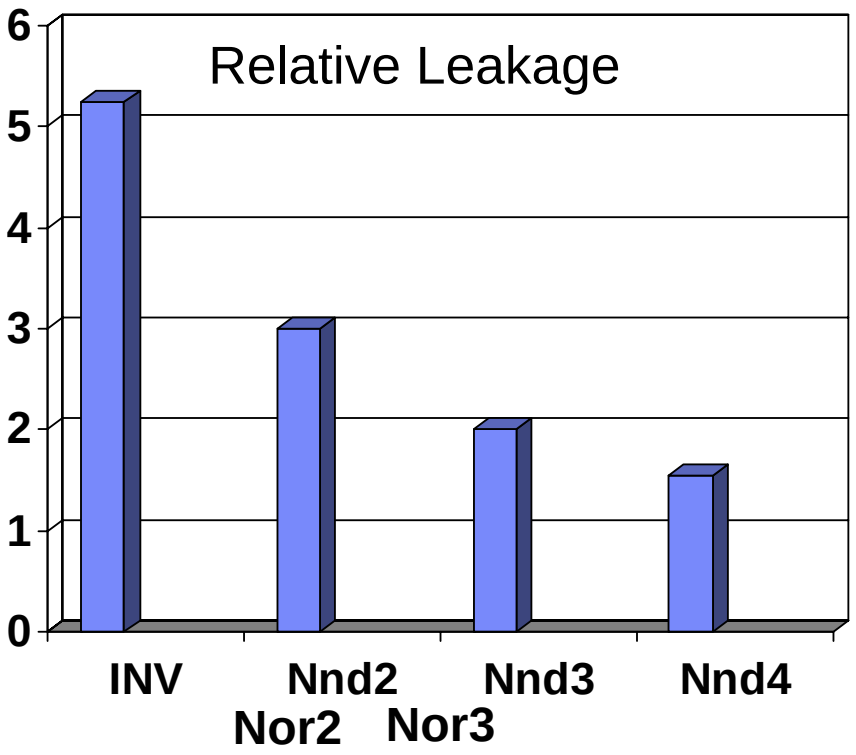
Example: Fine-Grained Voltage Island in uP Core

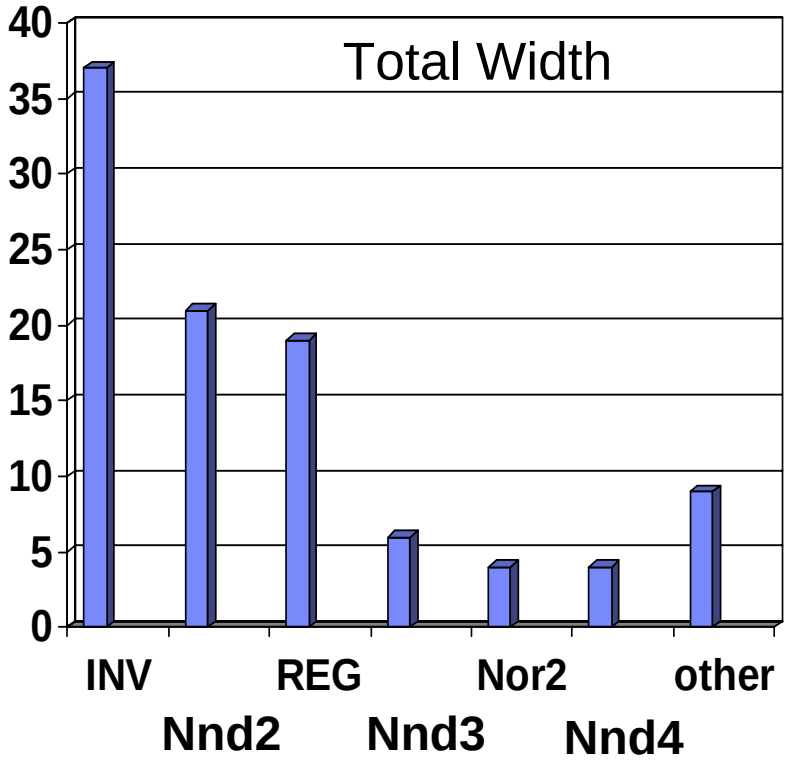
- **No timing degrade, and no area increase for the core**
- **Power saving about 5-10% (due to various system and circuit constraints, only 25% cells are selected to be V_{ddl}, although 60-70% can potentially be V_{ddl})**



Leakage Contribution as a Function of Topology

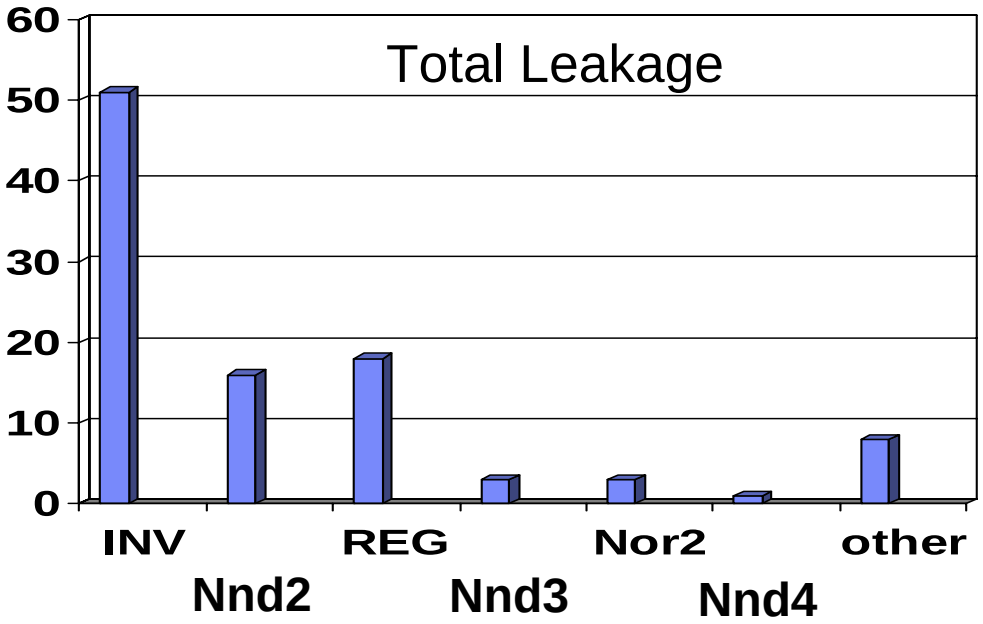
- **Leakage has a strong dependence on circuit topology.**
  - Inverters are the leakiest topologies





Leakage Impact – Design Options

- As a result of high relative leakage and the highest total width, inverters constitute over 50% of the total leakage in a typical design

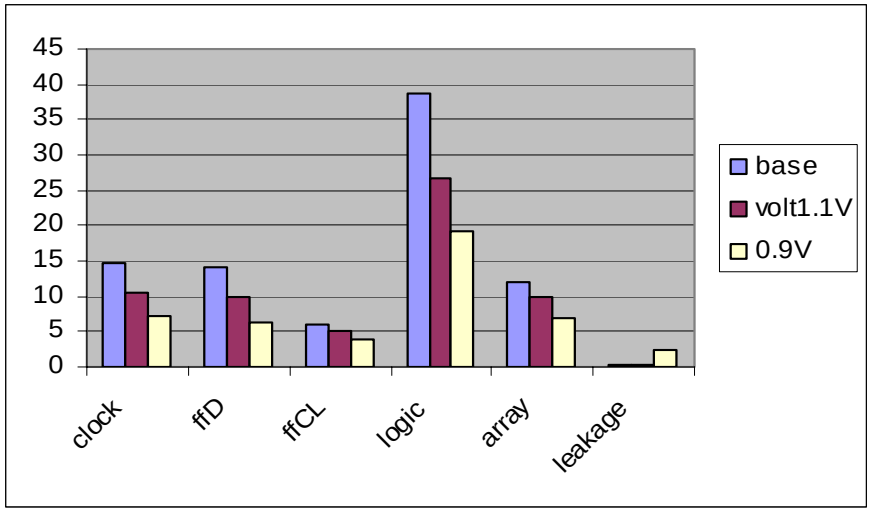


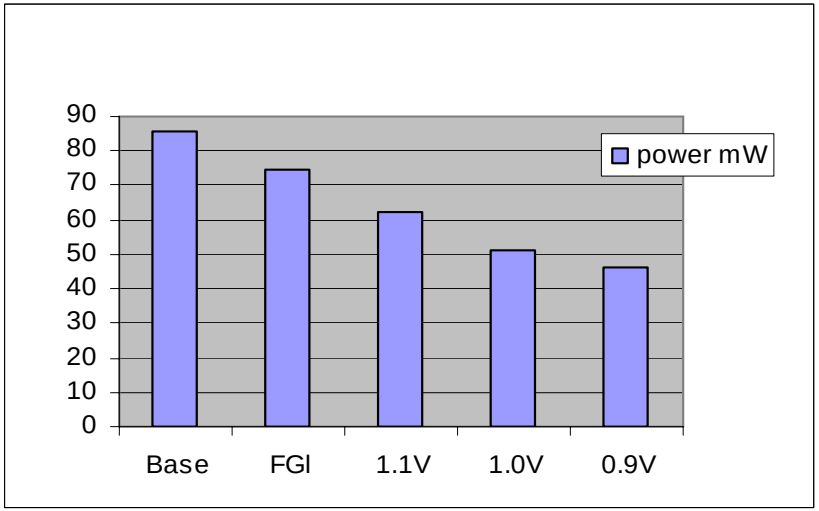
- Can leakage power be minimized by improved algorithms for circuit topology selection? **Are we using all design levers that are available?**

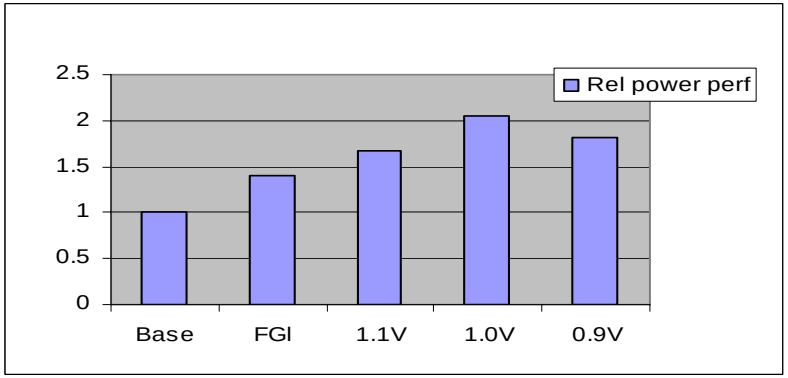
Examples – What Can Be Achieved

- **ASIC**
- **Embedded Processor**
- **High Performance Processor**

Case Study: ASIC Power Optimization







- **Optimizations**
  - **Fine Grained Library**
  - **Voltage Scaling**
    - **1.2V to 0.9V**
  - **Multi threshold at 1.0V and 0.9V**
- **Power: 85 to 46 mW**
- **PowerPerf : 2x**

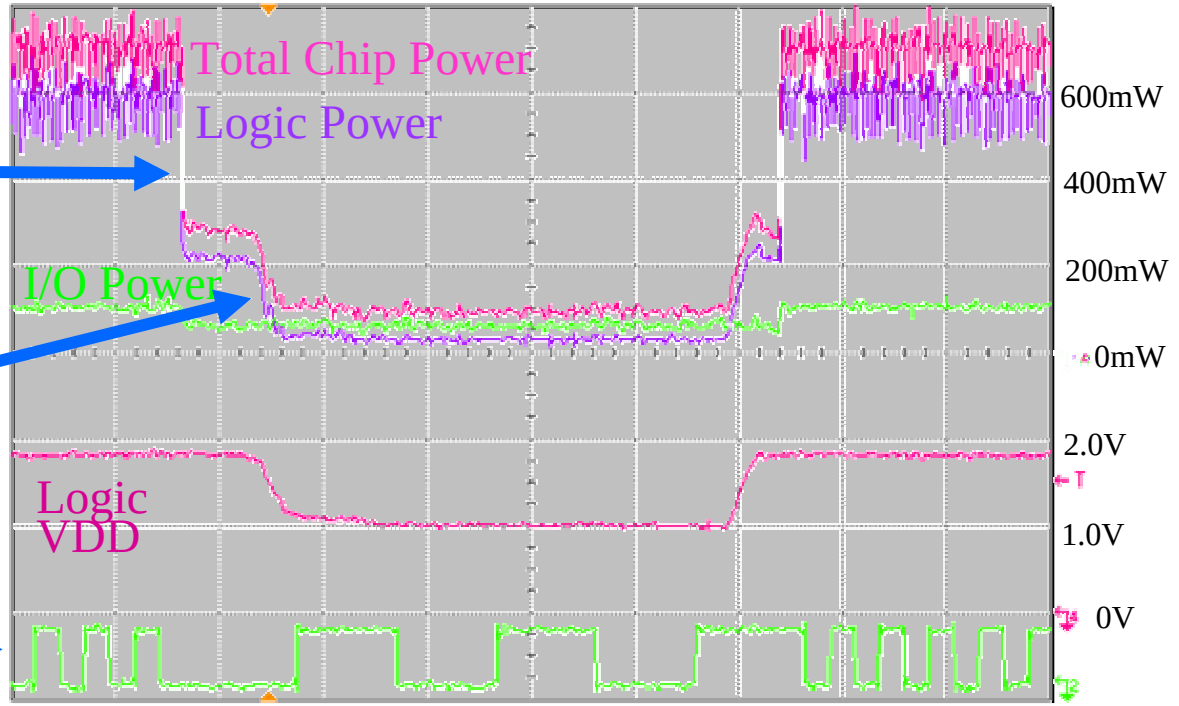
“Conventional” Power Management

PowerPC 405LP Example

Dynamic Frequency Scaling
266MHz CPU to 66MHz CPU

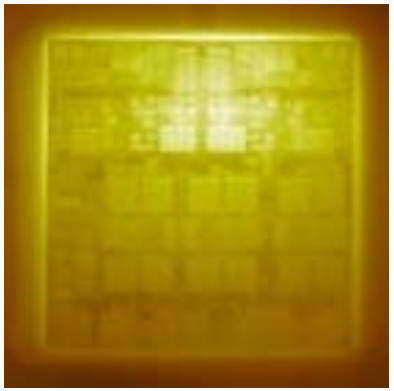
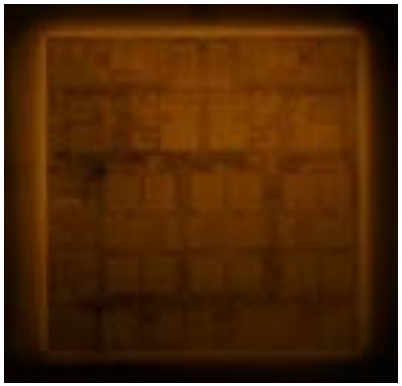
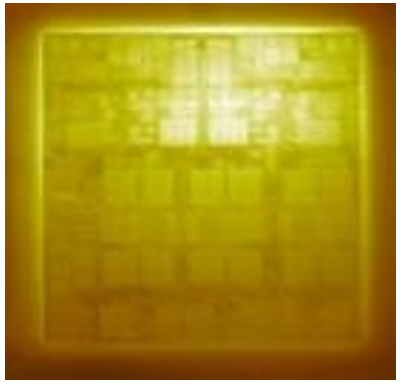
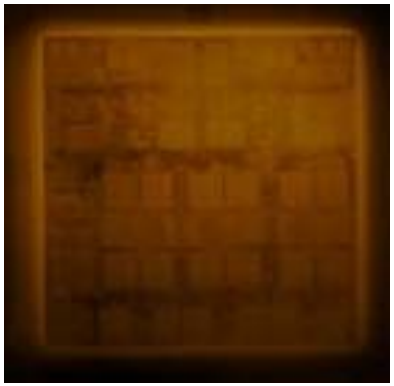
Dynamic Voltage Scaling
1.8V --> 1.0V at up to 1V/100us

Uninterrupted Operation
Linux 2.3.17 Running
Dhrystone 2.1 code
400 loops per cycle .



Adaptive Power Management Through Frequency and Voltage Scaling

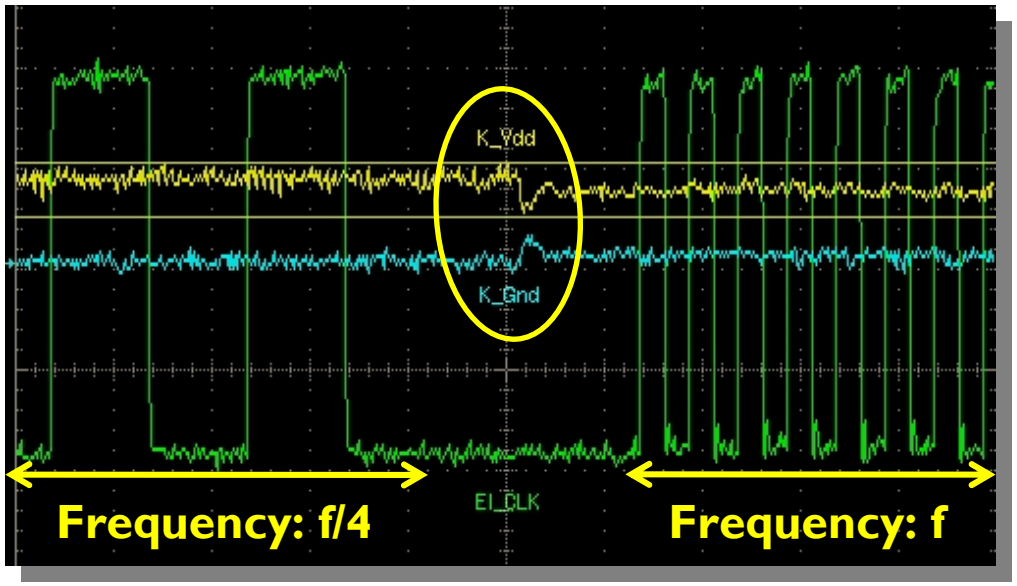
POWER5 Dynamic Power Management

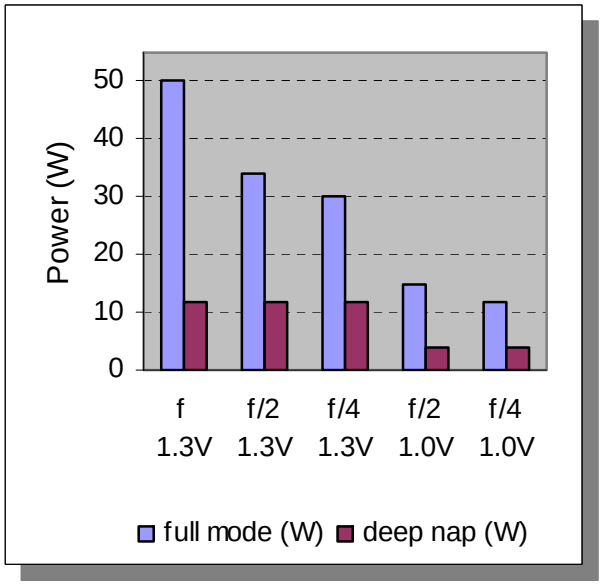
	No Power Management	Dynamic Power Management
Single Thread		
Simultaneous Multi-threading		

Photos taken with thermal sensitive camera while prototype POWER5 chip was undergoing tests

Simultaneous Multithreading with dynamic power management reduces power consumption below standard, single threaded level

PowerPC 970® PowerTune Technology





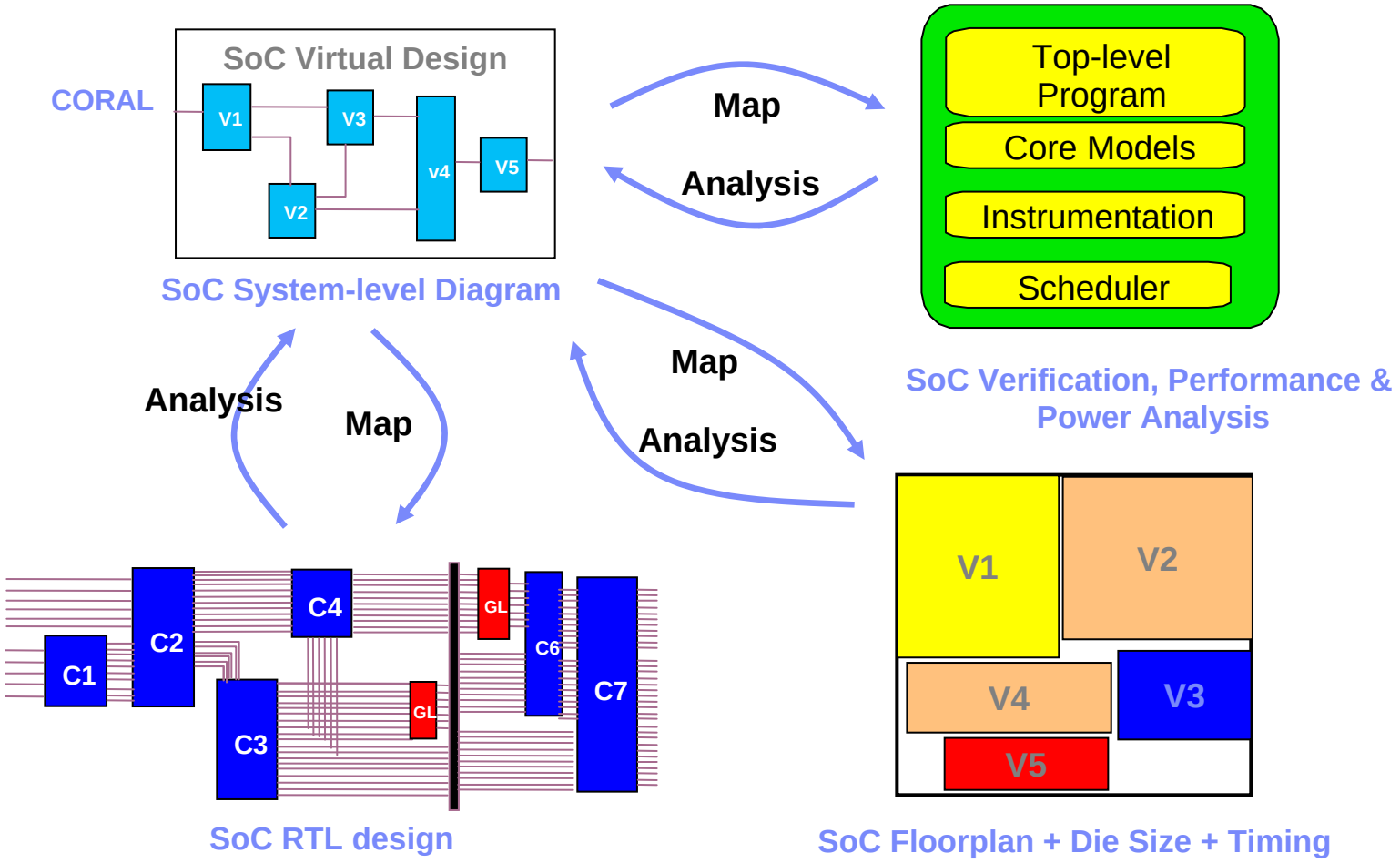
- **Seamless voltage / frequency scaling under program control**
  - Operates at full frequency, half and quarter frequency
- **Employs clock dithering to reduce dI/dt**

Power Management Design Tools

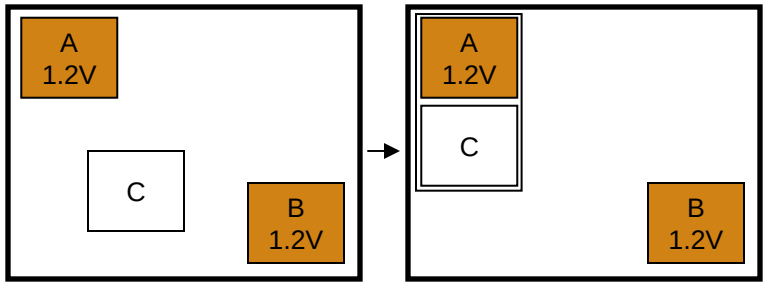
- **Design complexity is increasing**
- **Time to Market pressures are greater than ever**
- **Costs to tape-out a complex SoC design are escalating rapidly – it is unacceptable to respin the design multiple times to achieve performance/power goals**

- **Improved Design Tools are a MUST**
  - Realistic power estimates
  - Design concept tradeoff assessments
  - Easily inserted into industry standard design flows

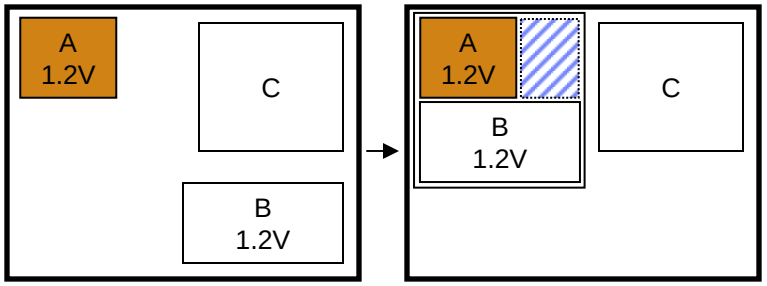
Example – System for Early Analysis of SoCs (SEAS)



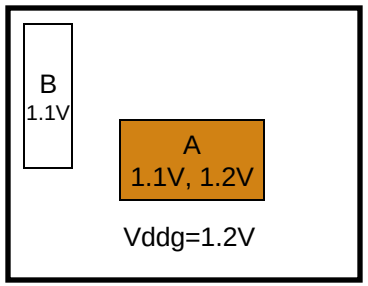
Motivation: Floor planning for Voltage Island Creation



Preplaced blocks



Dead space



Proximity to power pins

AND, there are always other issues
-Timing
- Routing congestion
- …

- *Very hard to come up with a manual solution – Automation clearly needed,*
- *Creation of Voltage Islands (or Partitioning a Design into Voltage Islands) has to be Physically Aware.*

Power Must Be Controlled at All Levels

- **Technology**
- **Chip Design**
- **System Design**
- **Software Design**
- **Manufacturing**

- **If any facet of your final product is not “Power Aware”, and your competition is, you will not succeed in the market!**

Summary

- **Historical frequency improvements are slowing as power densities are increasing**
- **Power management techniques integral to successful system architecture**
- **Collaborative innovation will be required to keep performance on its historical exponential curve**

Acknowledgements

- **IBM Research**
  - R.Joshi
  - P.Bose
  - R.Puri
  - Y.Shin
  - L.Stok
  - J.Darringer
- **IBM Electronic Design Automation**
  - N.Dhanwada
- **IBM System and Technology Group**
  - M.Papermaster
  - A.Correale
  - M.Sherony
  - L.Su
  - R.Schmidt
  - G.McElveen
  - T.Parker

IBM Microelectronics Division

1580 Route 52, Bldg. 504

Hopewell Junction, NY 12533-6351

The IBM home page can be found at http://www.ibm.com