PRACTICAL POWER GATING AND DYNAMIC VOLTAGE/FREQUENCY SCALING

Stephen Kosonocky

AMD Fellow

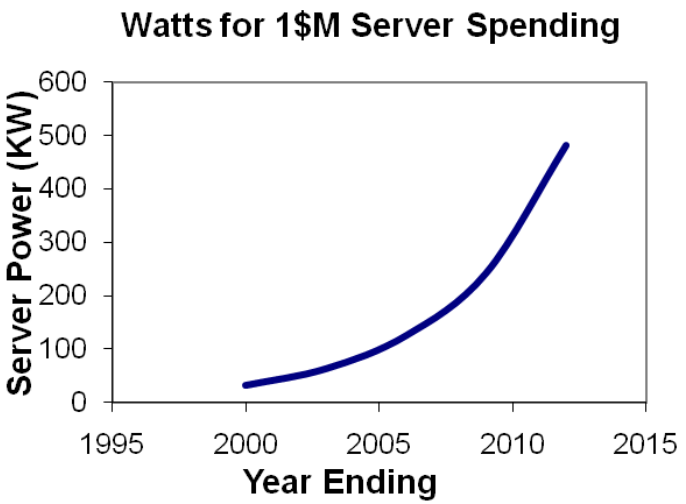
OUTLINE

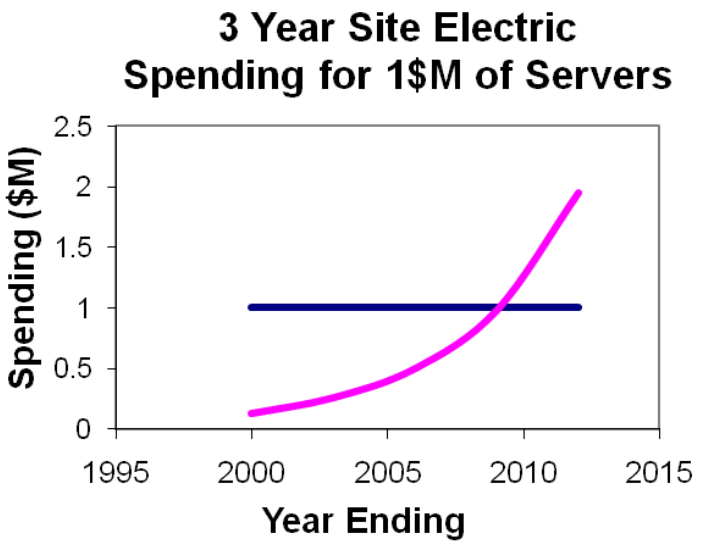
- Motivation for DVFS and Power Gating
- Dynamic Voltage/Frequency Scaling
- Power Gating
- Conclusion

MOTIVATION

SERVER COST TRENDS

- Server compute capacity/$ is increasing steadily
  - Driving up total power/$ spent on servers
  - Also driving up electrical power spending
- Green initiative and focus on energy independence creates additional pressure to lower energy costs
- Increased energy efficiency directly reduces costs

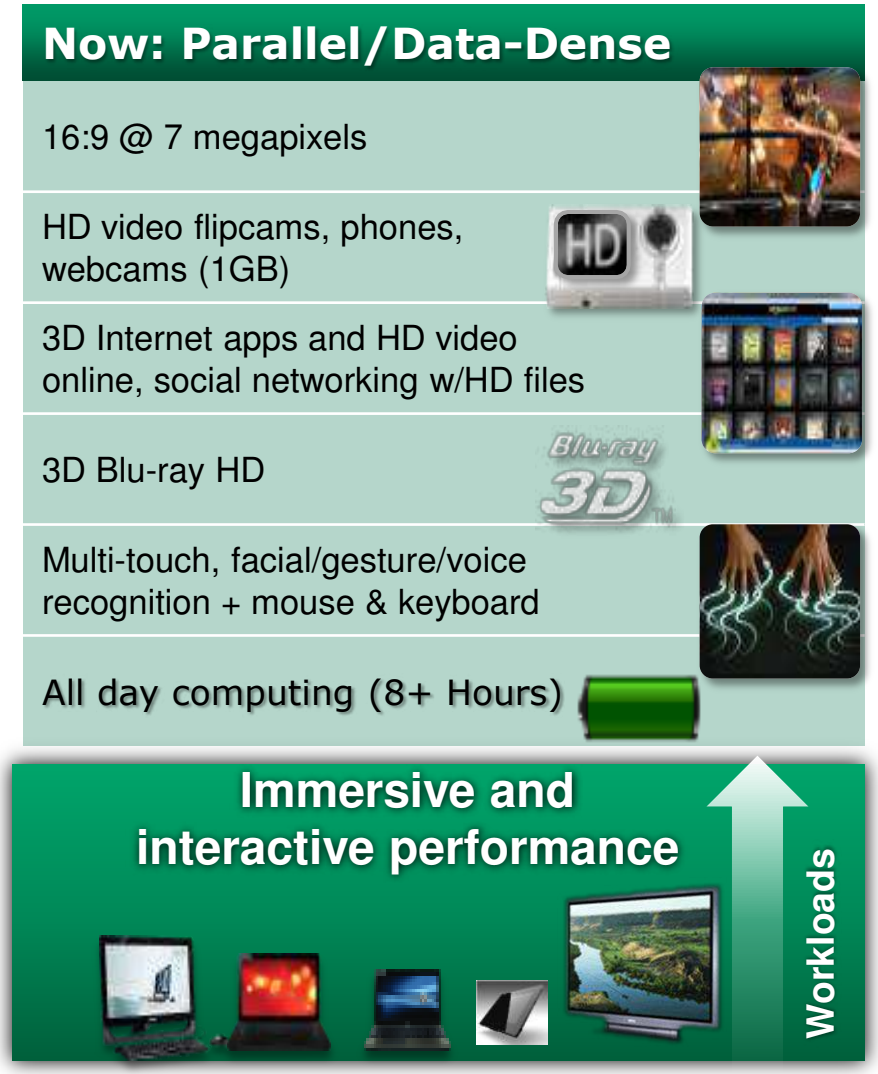




• Source Data: Uptime Institute, 2009

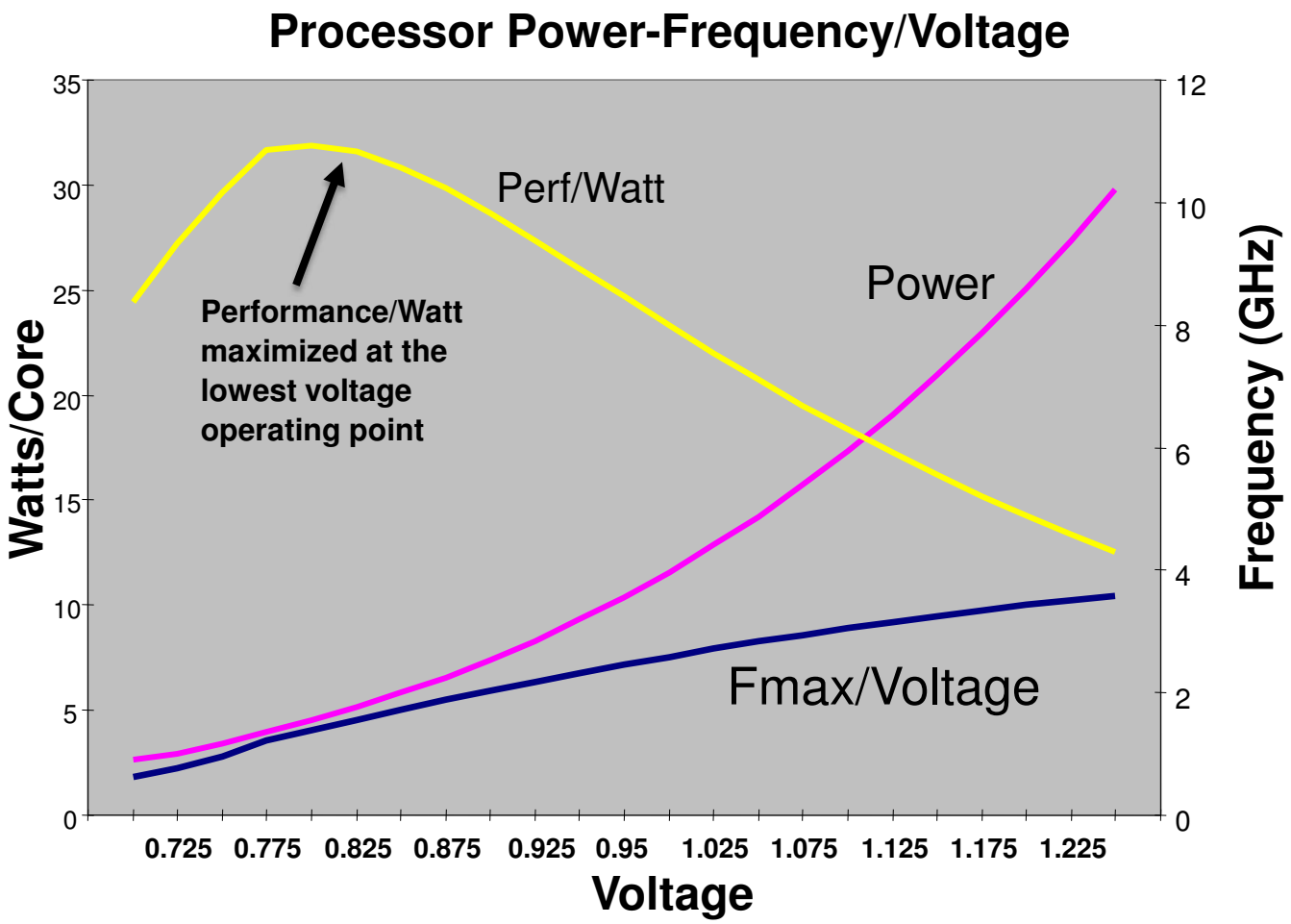
MOBILE AND DESKTOP

- High degree of integration of special functions
- Diverse workloads
  - Data serial, data parallel
  - Streaming
  - Compute intensive in bursts
  - Long periods of inactivity
- Small form factor
- Long battery life
- Instant access of the device
- Autonomous power management



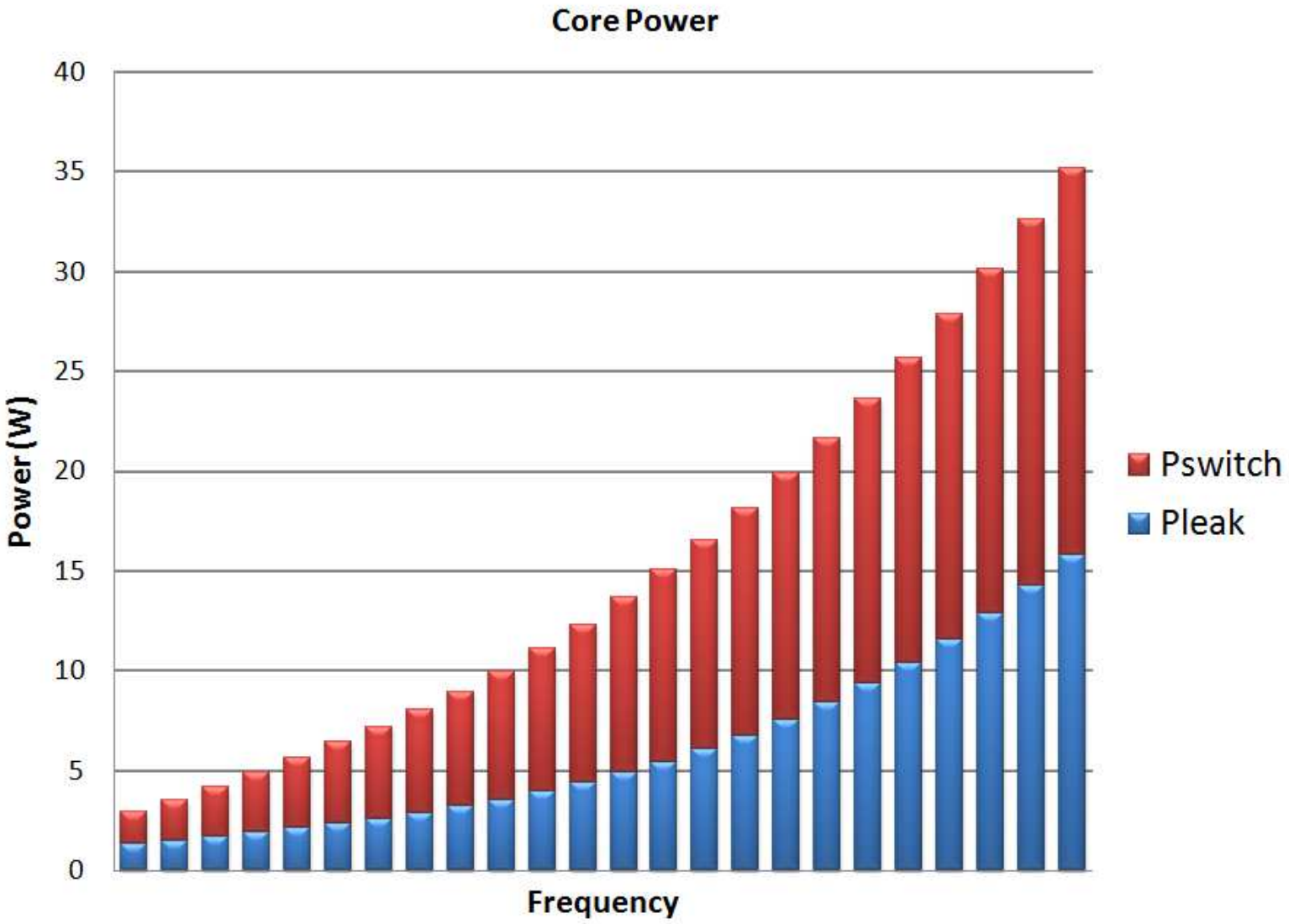
- Samuel Naffziger, VLSI Symposium 2011, Kyoto

POWER/PERFORMANCE TRADE-OFF



Performance/Watt maximized at VMIN

HIGH PERFORMANCE CORE POWER



Depending on process node, leakage ~ 40% total power

Two major knobs have emerged for controlling power

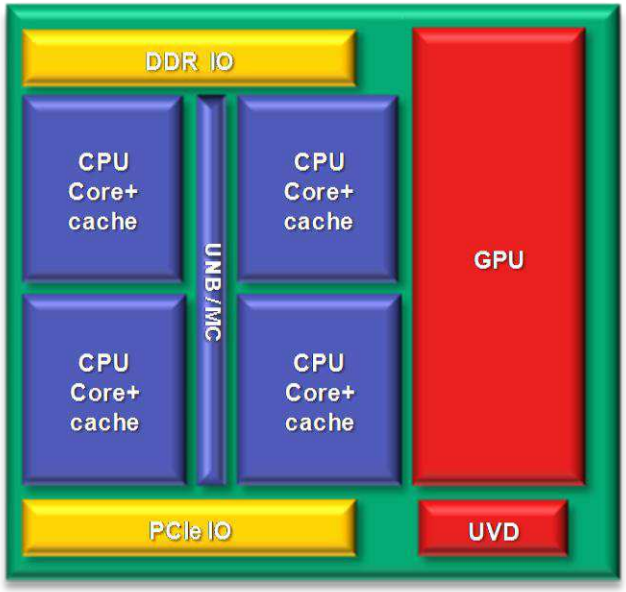
1. Dynamic Voltage and Frequency Scaling
   - Optimize performance for the application while it's running
2. Power Gating
   - Gate power during idle periods

Each present unique challenges for implementation and optimization

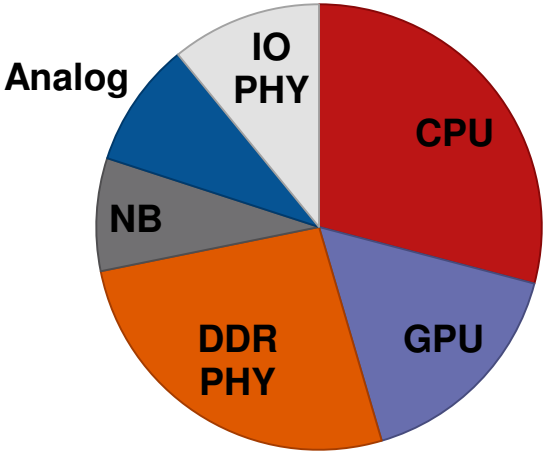
DYNAMIC VOLTAGE/FREQUENCY SCALING

LLANO ACCELERATED PROCESSING UNIT (APU)

- Integration Provides Improvement
  - Eliminate power and latency of extra chip crossing
  - 3X bandwidth between GPU and Memory
  - Same sized GPU is substantially more effective
  - Power efficient, advanced technology for both CPU and GPU
  - Thermal management optimization between CPU and GPU
- Key features for Mobile Mark 07
  - Lower Idle, Active power
  - DVFS / Clock-gating / Power Gating
  - Robust and Flexible Power Management Control



~ Power Breakdown



APU POWER MANAGEMENT

	DDR PHY	IO PHY	CPU	GPU	NB
HW policy	Self-Refresh PreChrg. Pwr-Dn PLL Off	Link width Link suspend PLLs Off	DVFS Deep Pdn PLLs off	DVFS T-put scale Power Dn	DVFS Power Dn
AC/DC policy	Speed, physical interface chg	Link rate change	Max PState	Max PState	Max PState
Boost			DVFS	T-put Scale DVFS	DVFS
OS			P-States Halt		
Driver		Power Dn		Power Dn	Set PStates

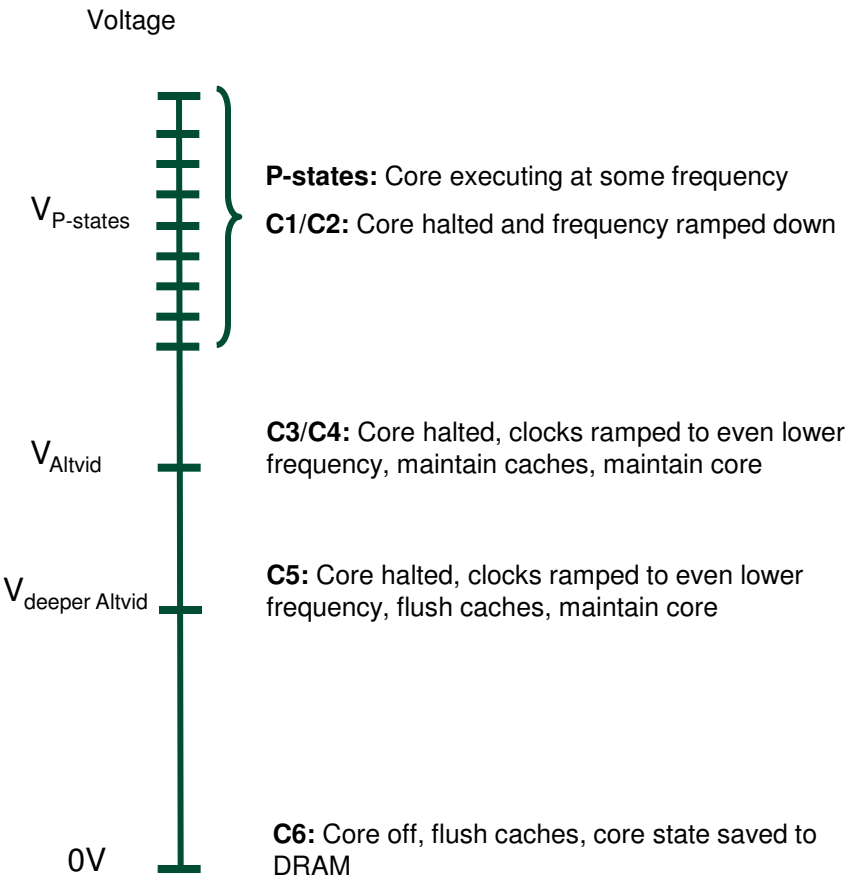
- Sebastien Nussbaum, IEEE Vail Computer Elements, June 22th, 2009

DYNAMIC VOLTAGE SCALING

- Dynamic frequency and voltage scaling (DVFS)
  - Match the frequency of operation with the workload
  - Lower voltage to sustain required frequency

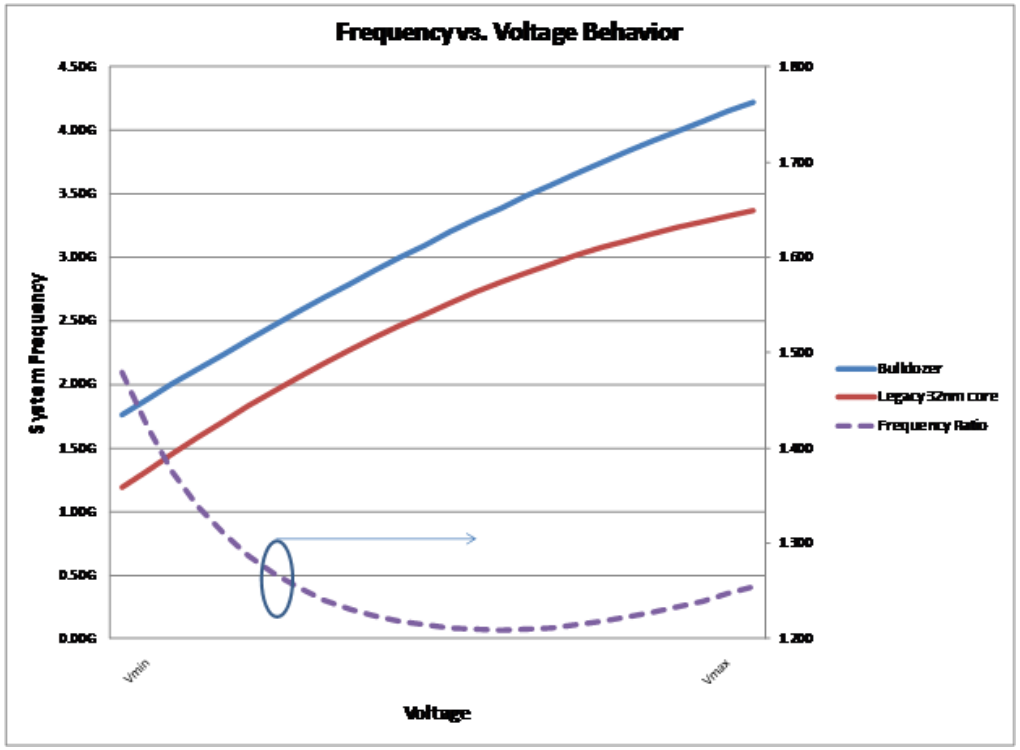
Examples:

- P-states for active control of voltage/frequency
- C-states for degrees of idle control



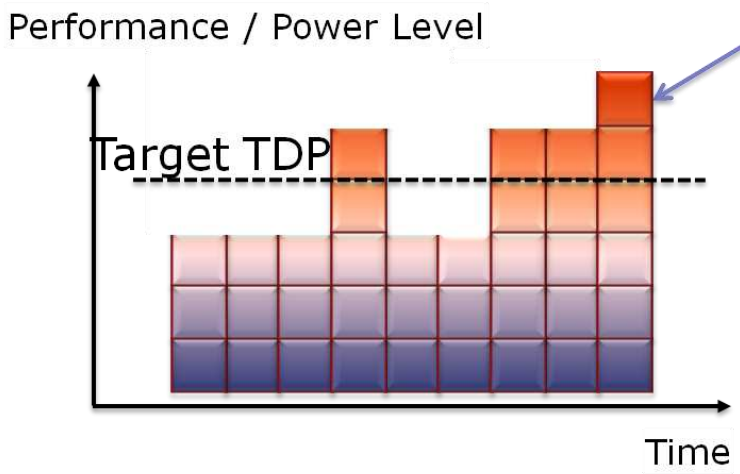
DESIGN FOR DVFS

- Optimization of timing across entire voltage range necessary for efficiency DVFS
- Required two separate timing corners with equally aggressive cycle time goals
  - 0.8V & 1.3V
- Both gate-dominated timing paths and paths with high RC content have been tuned to provide better scaling than the prior design
- Enables the core to thoroughly exploit boost capability



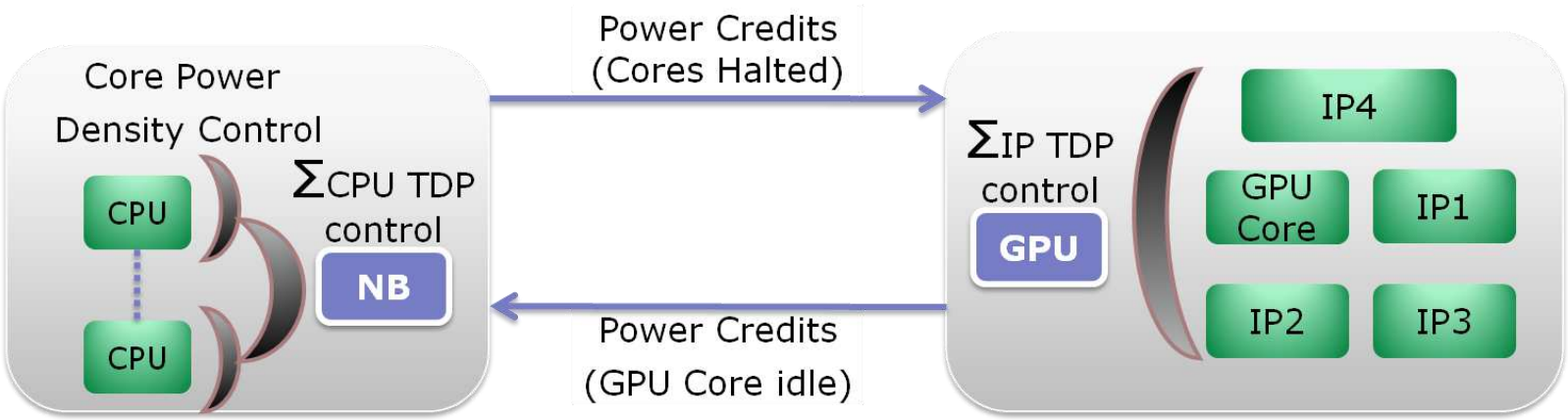
APU DVFS

Performance: low-power applications run at higher frequency



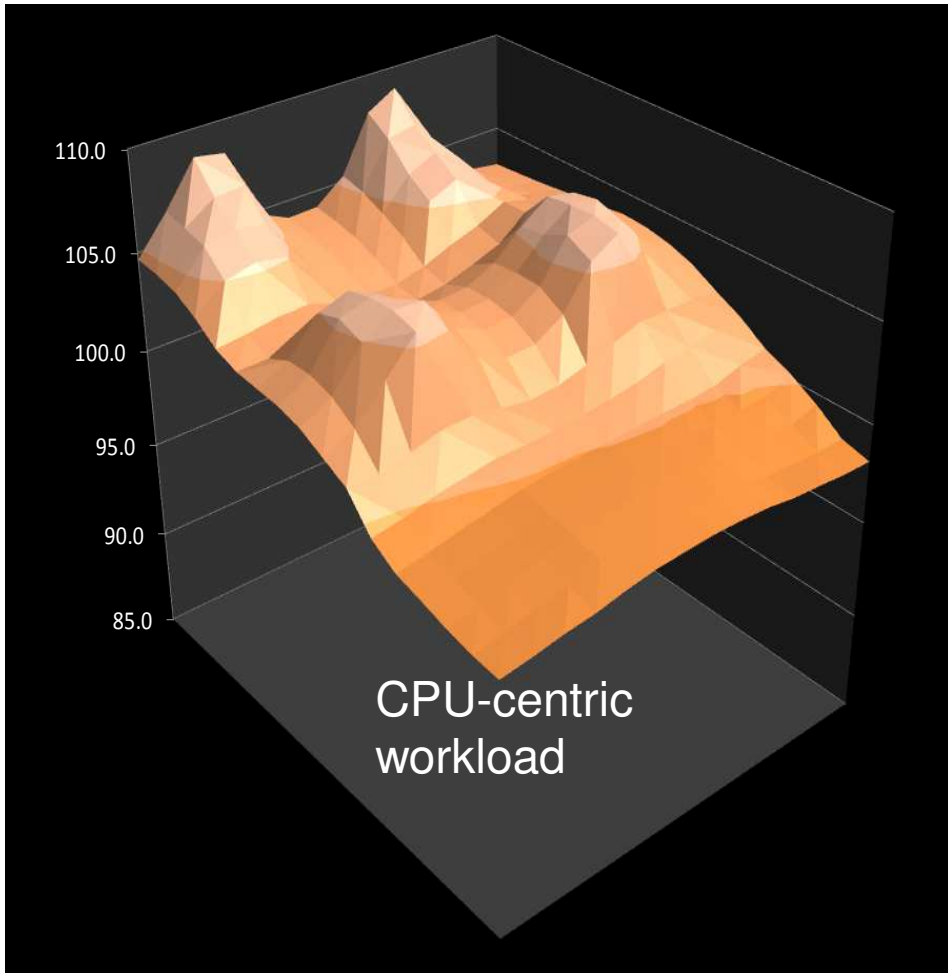
Boost: variable opportunity based on VRM & Pkg constraints, thermal density, and CPU/GPU state

- Chip power estimator
- Thermal history representation
- Lower power apps → better perf. Mechanism ~invisible to application



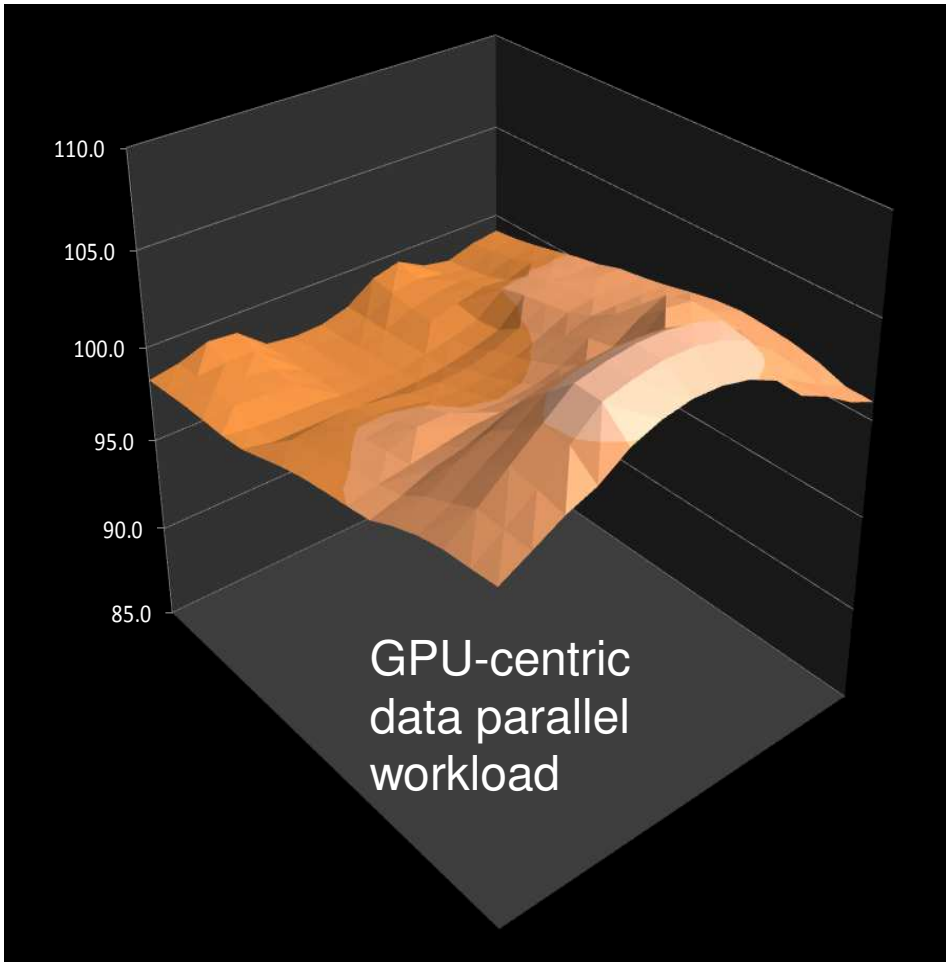
- Sebastien Nussbaum, IEEE Vail Computer Elements, June 22th, 2009

APU THERMAL PROFILE



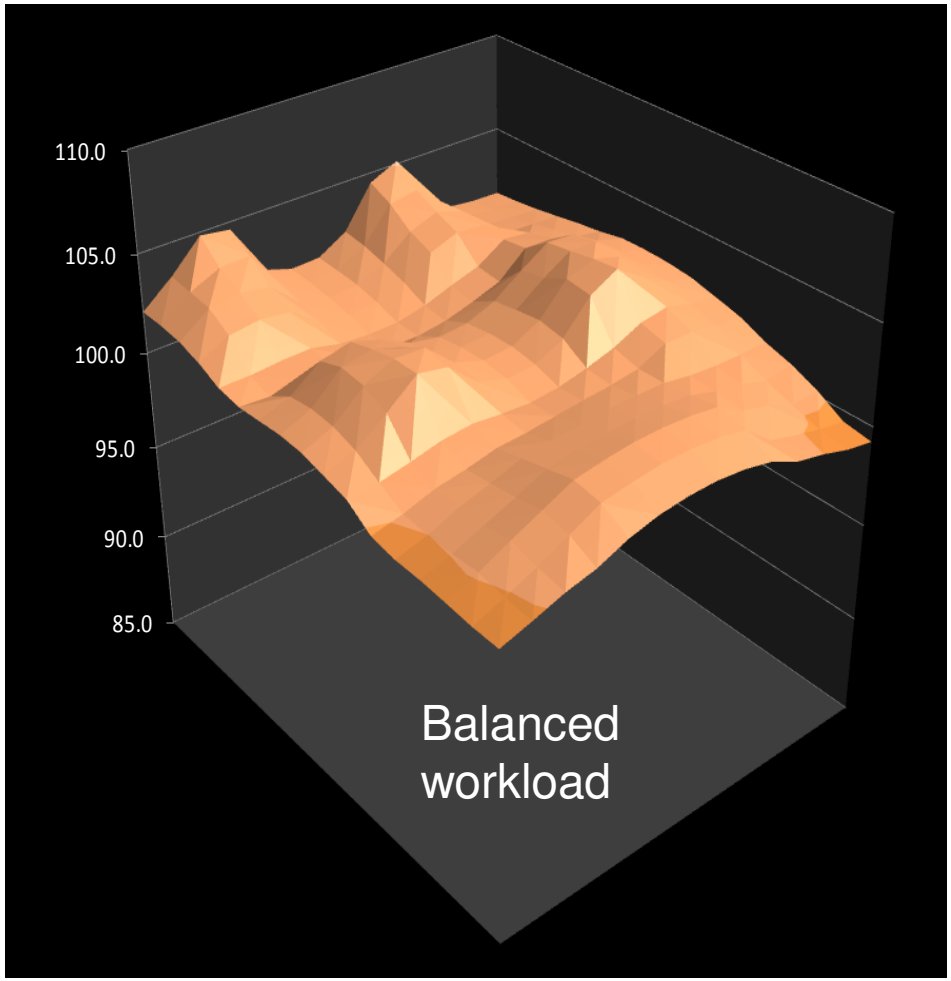
- Samuel Naffziger, VLSI Symposium 2011, Kyoto

APU THERMAL PROFILE



- Samuel Naffziger, VLSI Symposium 2011, Kyoto

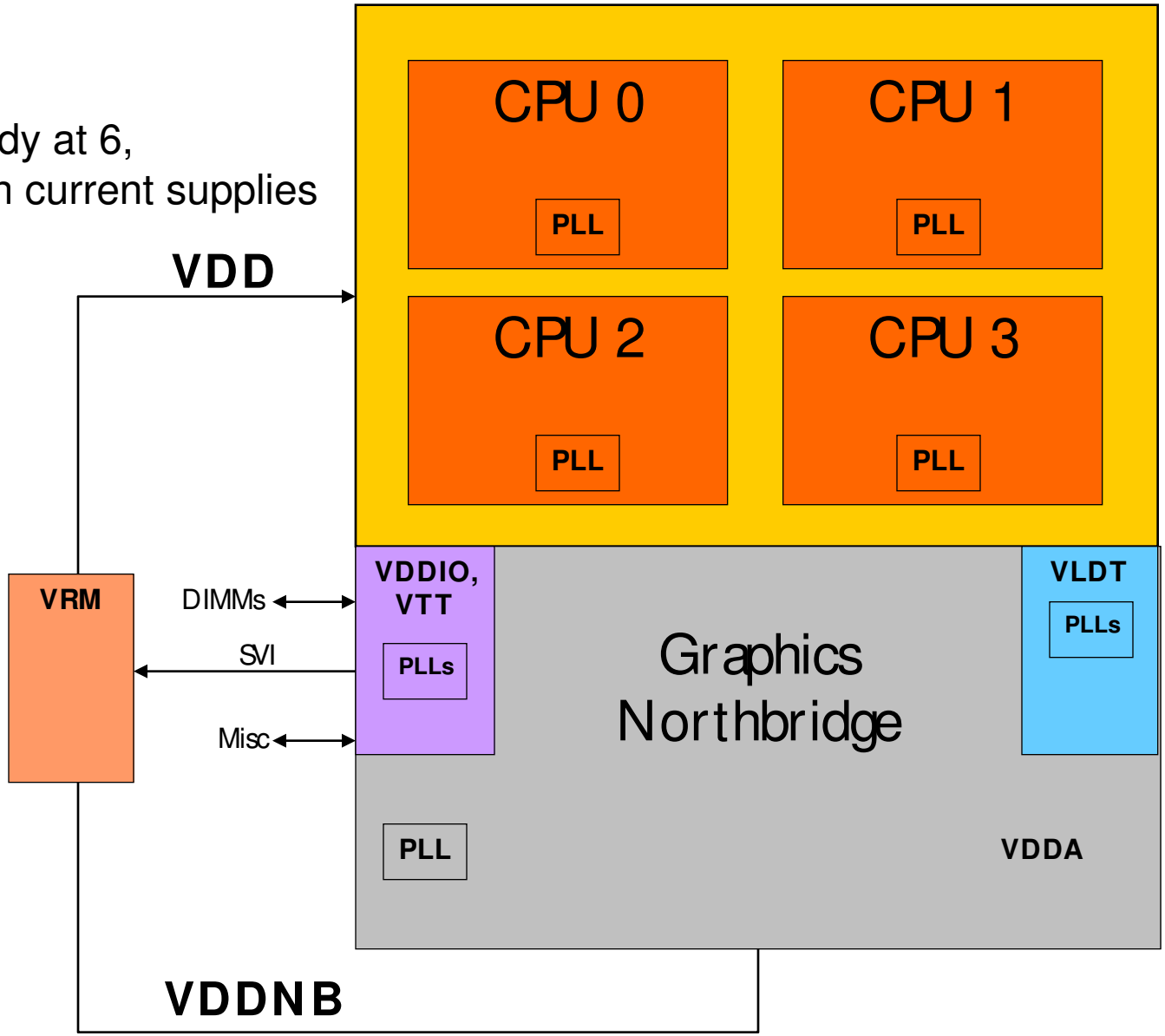
APU THERMAL PROFILE



- Samuel Naffziger, VLSI Symposium 2011, Kyoto

WHAT HAPPENS WHEN WE ADD MORE CORES?

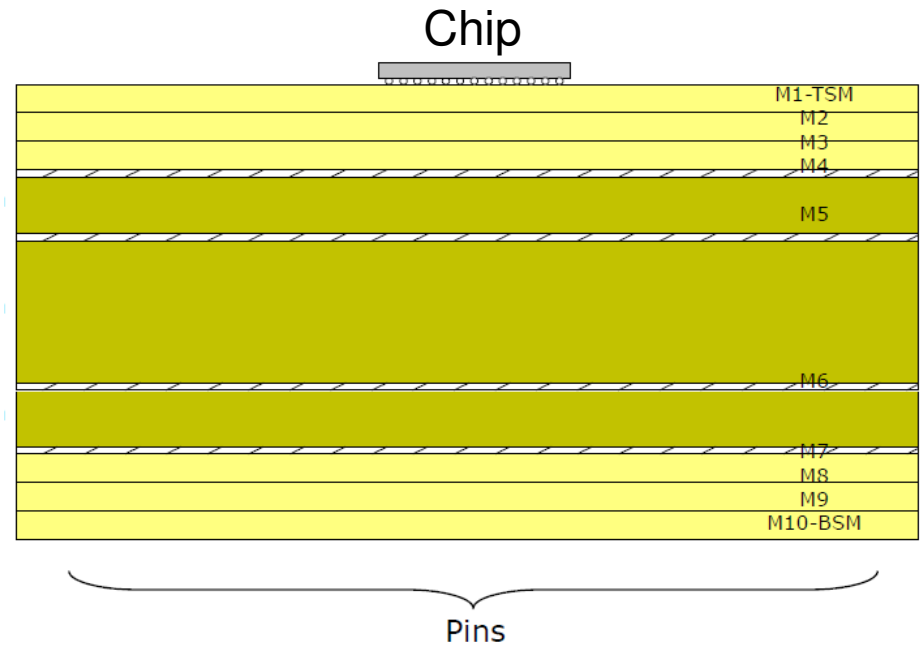
Already at 6,
2 high current supplies



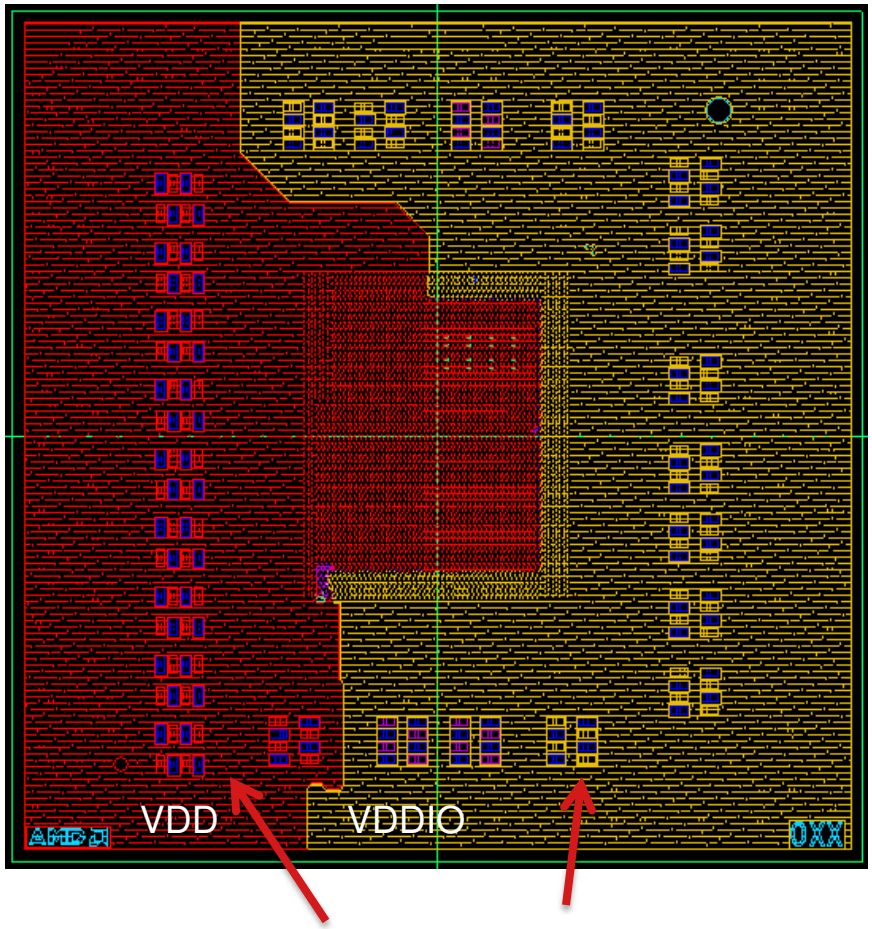
PACKAGE LIMITS

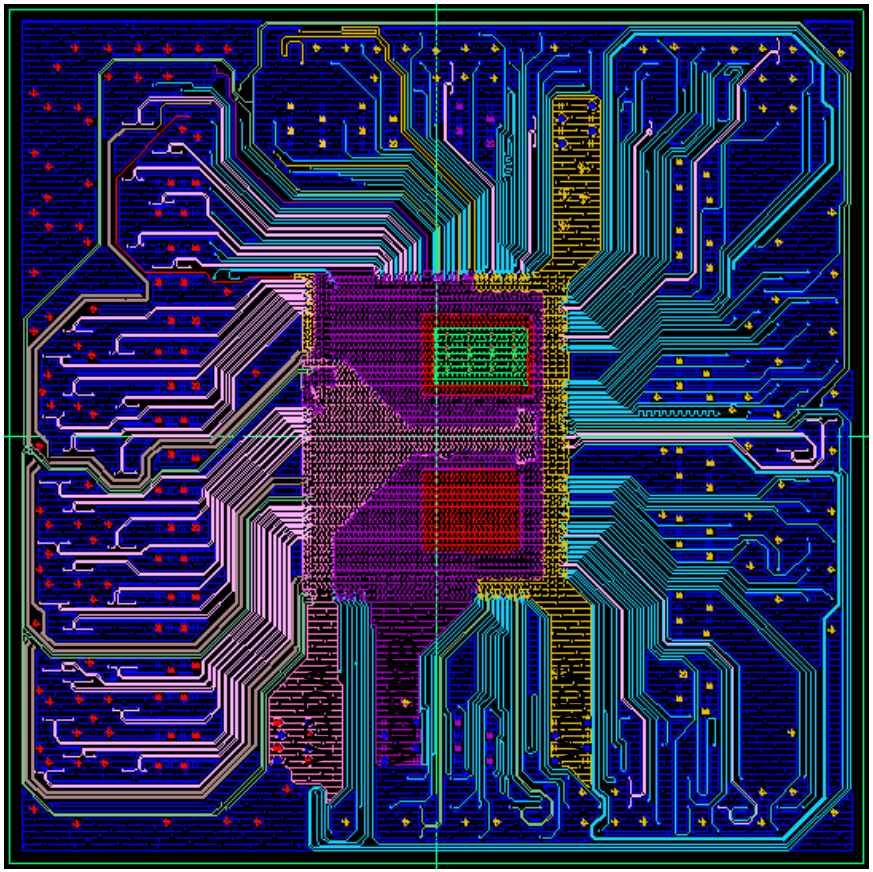
- Packages add constraints to increasing number of discrete supplies
- Typically only use 4 thick layers for high current power distribution
- 3-4 thin build-up layers on each side
  - Used for secondary supply and signal routing to pins
- Difficult to add many more supply rails

Typical low cost package structure



TYPICAL PACKAGE LAYER ROUTING LAYERS





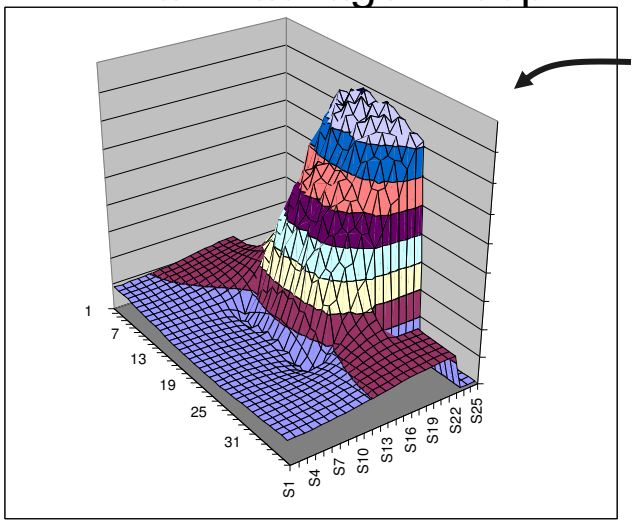
Package capacitors

Space limited for capacitors

Routing congestion

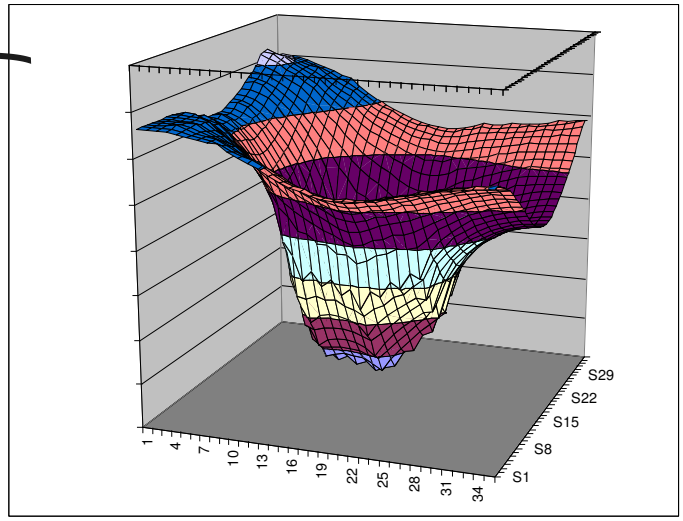
CORE VOLTAGE POWER DELIVERY IN PACKAGE

VDD Max Package Droop



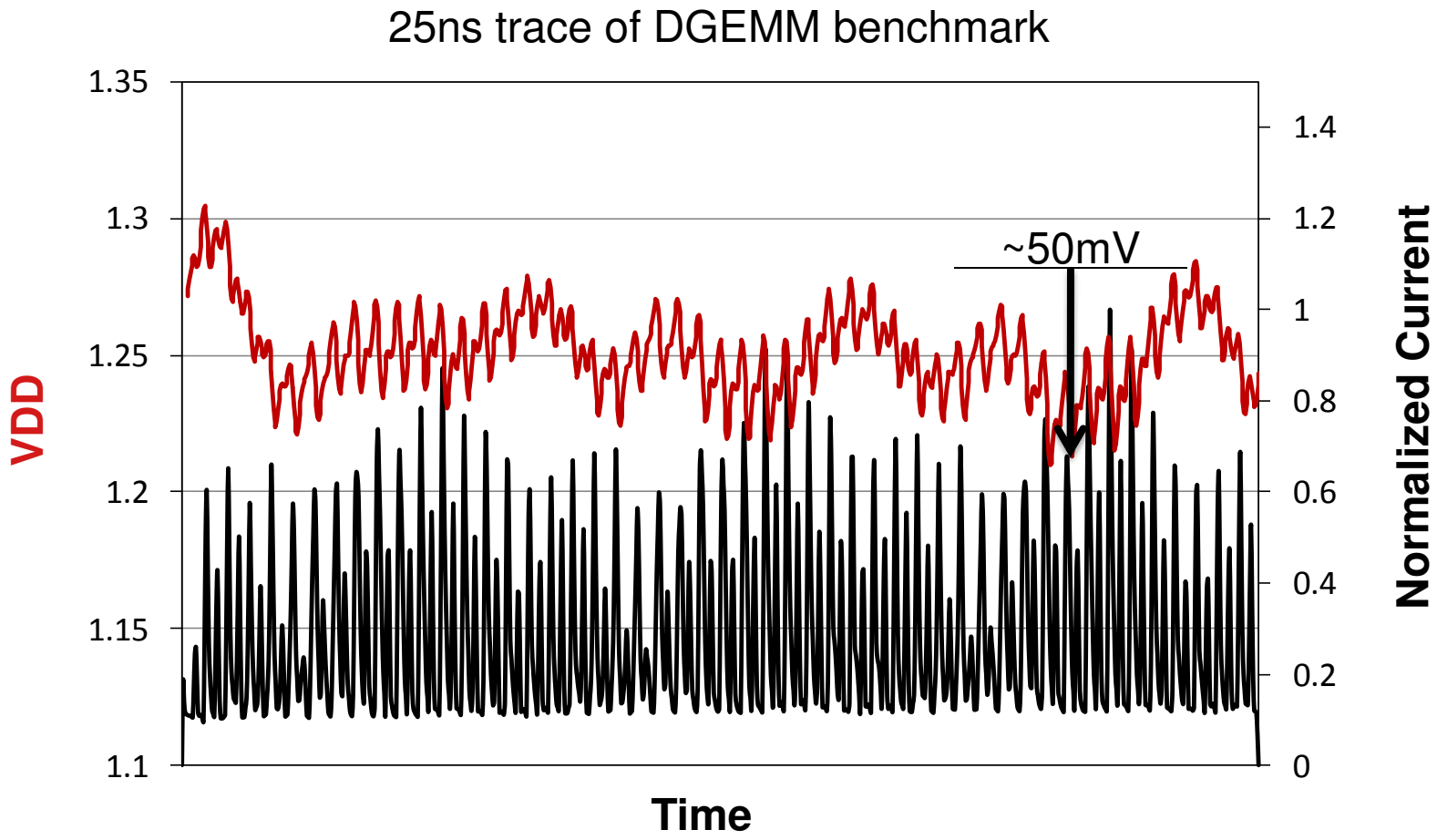
2.4x higher

VSS Max Package Droop



- Package Routing Constraints forces compromises for VDD plane
  - More difficult with increasing VDD planes
- More resources dedicated to VSS
  - Shows less droop

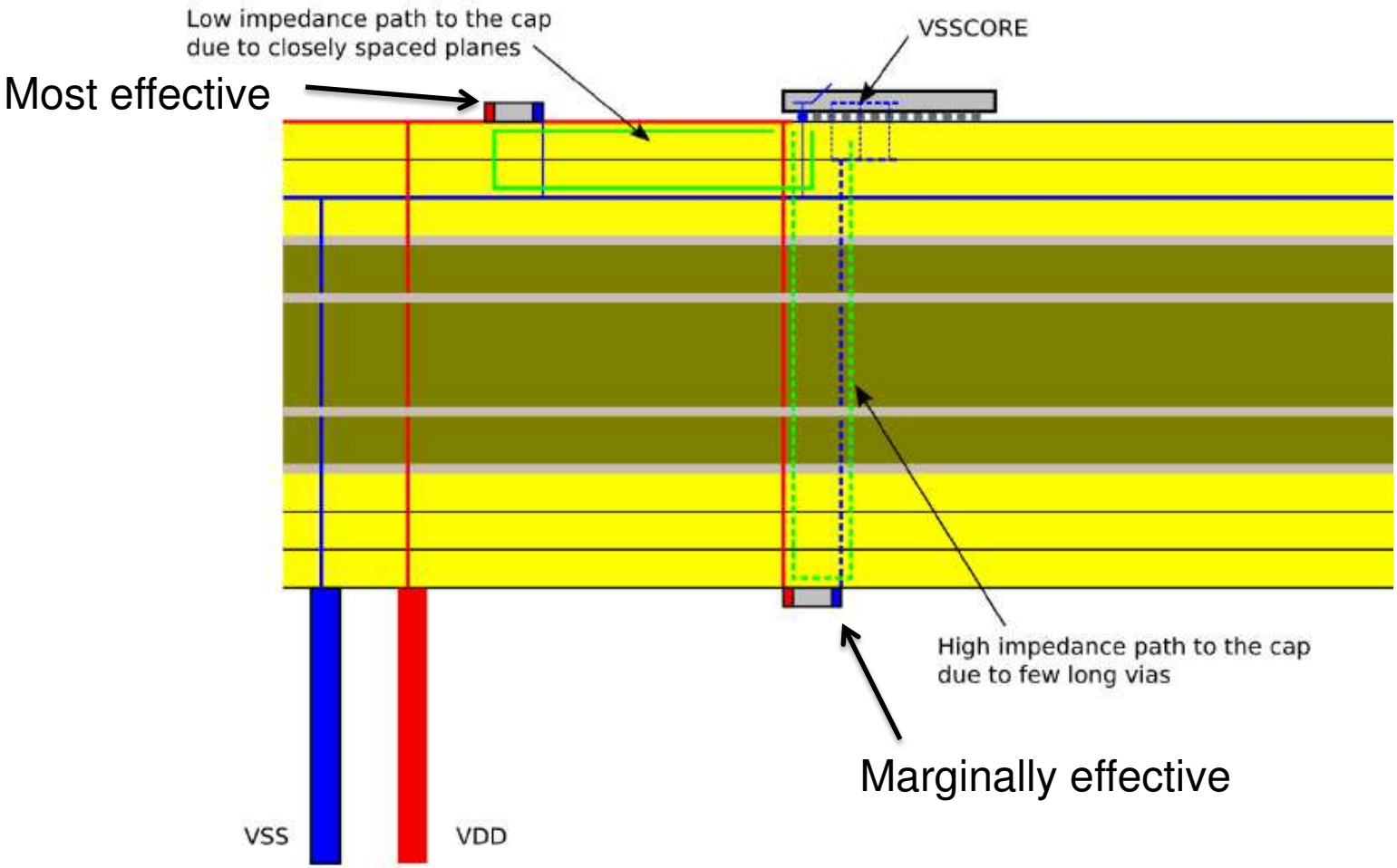
WHAT DOES THE ACTUAL VDD LOOK LIKE?



X86 core Supply simulation w/ package board model

Increased local decoupling cap could help

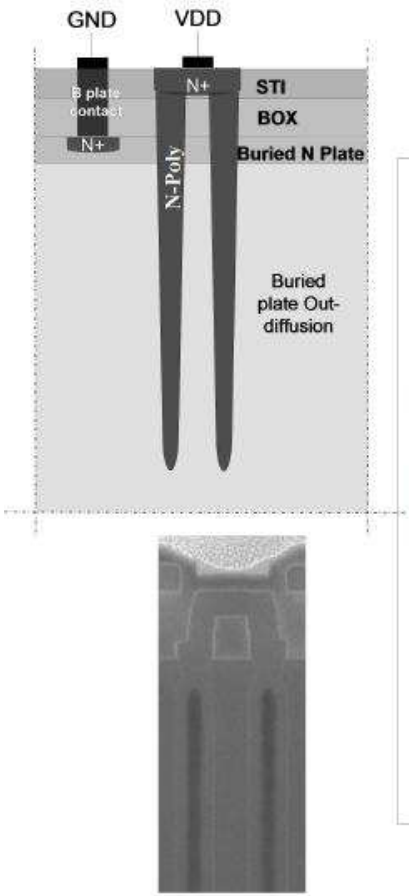
PACKAGE CAPS FOR SUPPLY DECOUPLING



Hard to find places for additional components

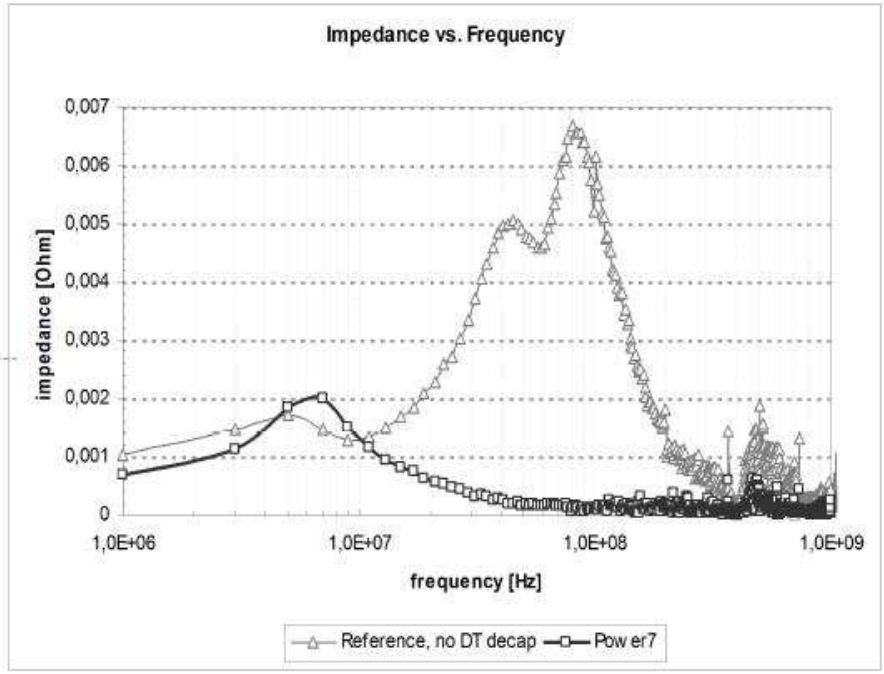
LARGE ON-DIE DECOUPLING CAPACITOR FOR REDUCTION OF SUPPLY NOISE

Trench Capacitor



Cross-section thru deep trench decap

Shift resonant frequency from 80MHz -> 7MHz



- D. Wendel, et.al. “The Implementation of POWER7: A Highly Parallel and Scalable Multi-Core High-End Server Processor”, ISSCC 2010

100fF/um^2 ~ 25x thick-ox

Adds significant process cost

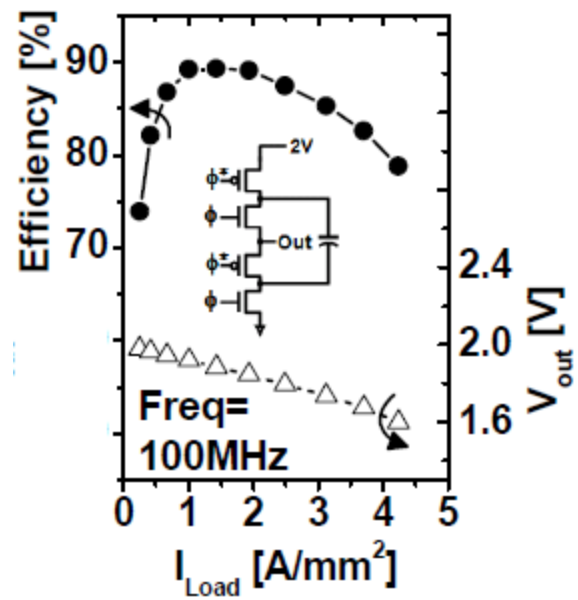
METHODS FOR INTEGRATED VOLTAGE REGULATION

- Switched capacitor regulator
- Buck converter
- Linear regulator
- Switched voltages

INTEGRATED SWITCHED CAPACITOR REGULATOR

- Efficiency dependent on load current
  - Most efficient at 2:1 voltage conversion ratio
  - Limits number of switches in the path
  - ~90% for 2:1, ~70% for other values
- Requires large capacitors
  - Can use trench capacitor for better area efficiency
- Large voltage and current ripple
  - Can mitigate with large decap and/or interleaved converters

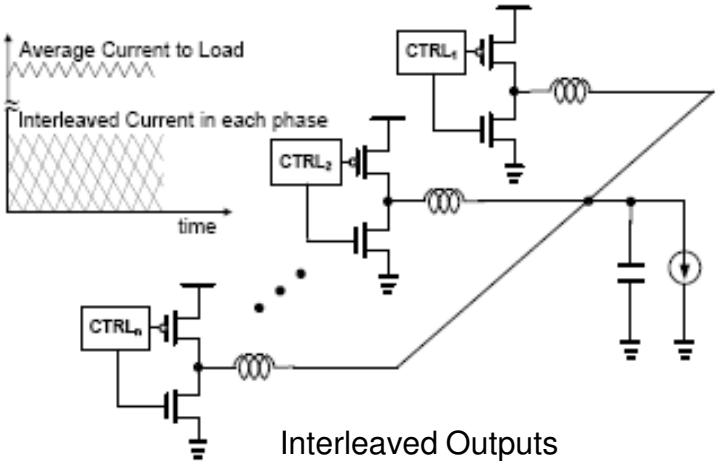
2:1 Conversion



Leland Chang, Robert K. Montoye, Brian L. Ji, Alan J. Weger, Kevin G. Stawiasz, and Robert H. Dennard, "A Fully-Integrated Switched-Capacitor 2:1 Voltage Converter with Regulation Capability and 90% Efficiency at 2.3A/mm2", VLSI Symposium , 2010

INTEGRATED BUCK CONVERTER

- Integrated inductor options
  - ~ 0.33nH possible
- High frequency operation can reduce inductance requirements
  - L proportional 1/Fs
  - Can reasonably approach 1GHz when fully integrated
- Parasitic series resistance is key problem
  - I^2R loss large
- Multi-phase or parallel regulator architecture can help
  - $(I/n)^2 R$ (n=number phases)
  - Current ripple is averaged by output cap
    - Too many phases will reduce efficiency for low current loads

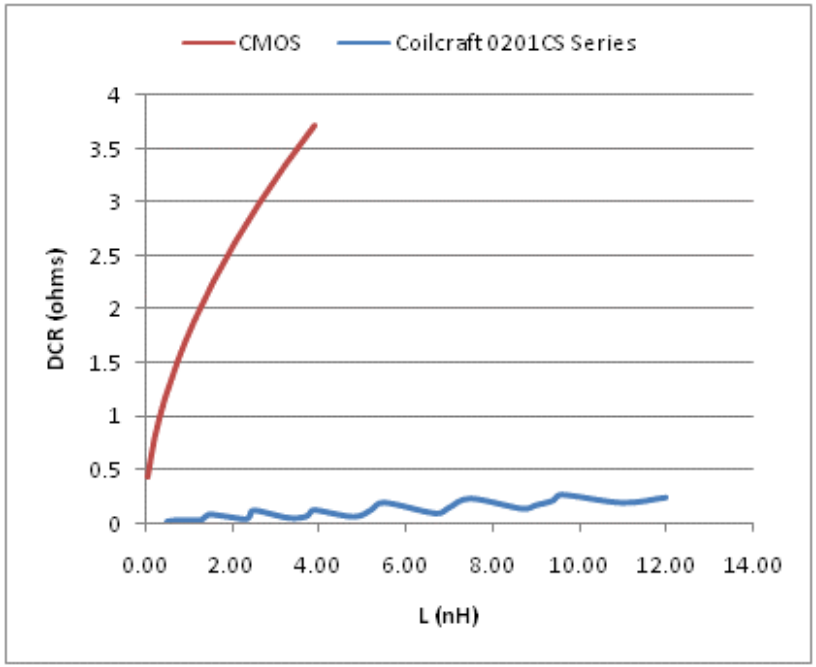


W. Kim, et.al, June, 2007

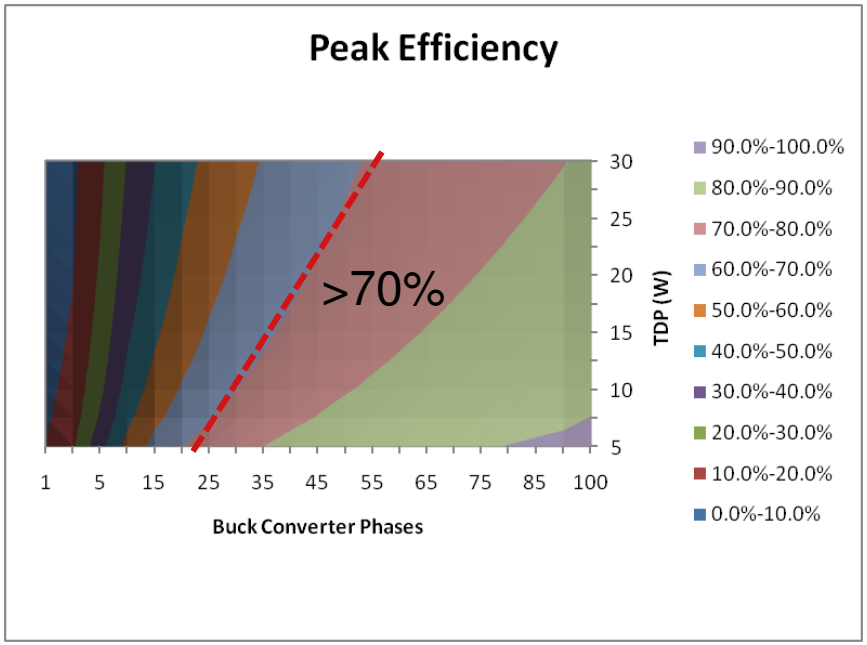
INTEGRATED BUCK CONVERTER KEY LIMITATION: DCR VS. INDUCTANCE

- Embedded inductors needs lower resistance or higher inductance/turn
- Multi-phase or parallel can approach reasonable efficiency only with very high number of paths
  - Reduces current in each inductor
- Fully embedded approach really needs a specialized integrated inductor technology

CMOS Inductor vs. Small Package size components



Peak Efficiency considering inductor resistive loss alone



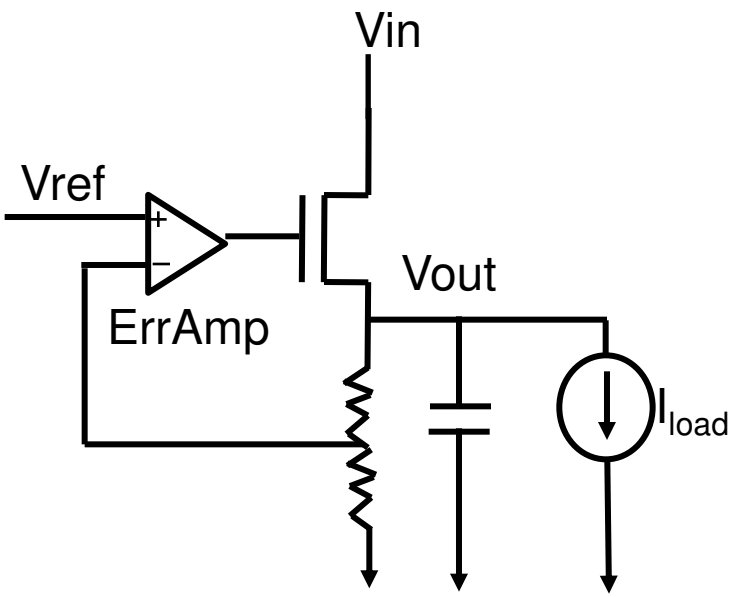
Discrete inductors consume valuable package resources

INTEGRATED LINEAR REGULATION

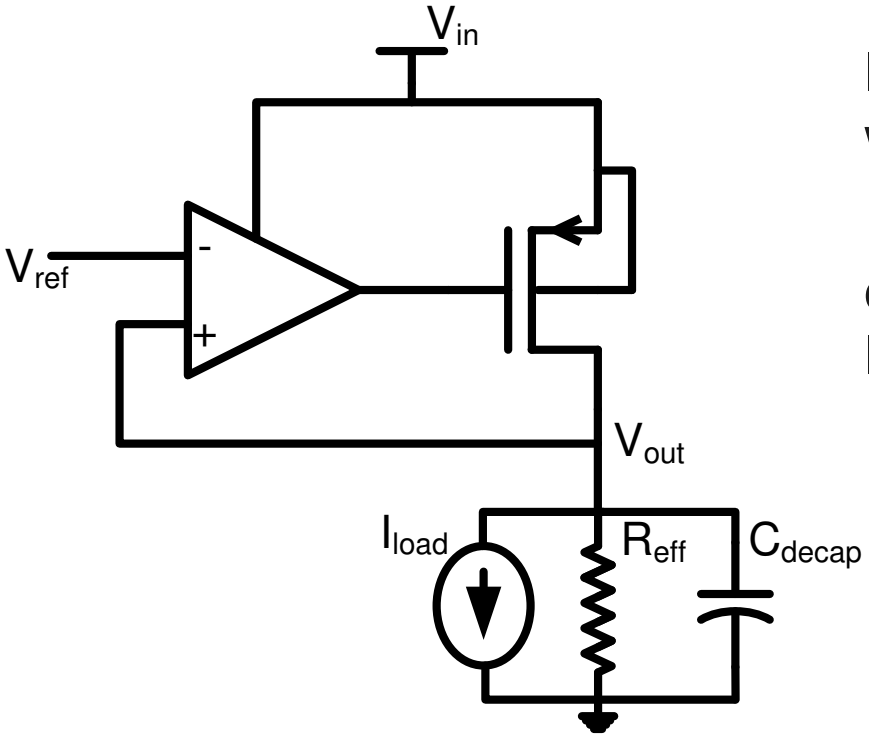
- Error between divided output and reference voltages is and fed to the pass transistor to correct the error
- NMOS type higher performance
- PMOS type (Low Drop Out)
- Efficiency good only for small (Vin – Vout)

  Eff = (Power Out)/(Power In)

  = ~ Vout/Vin
- LDO design can be more efficient (PMOS output)
  - Stability & speed of feedback become an issue



LINEAR REGULATOR (PMOS)

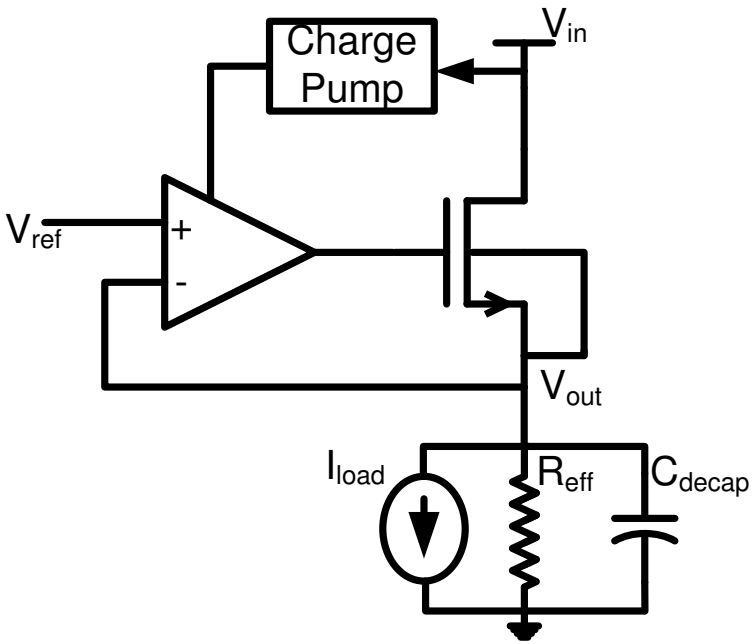


PMOS based linear regulators offer low dropout voltage while operating within (0-V_{in}) range

Required gain-bandwidth product challenging high for processor loads

- BW ~ 1GHz needed
- Hurts efficiency, power wasted in Op-amp

LINEAR REGULATOR (NMOS)



Source-follower topology

- Will require charge pump to provide gate overdrive.
- Accurate voltage setting.
- Supply rejection – dropout voltage tradeoff

High BW response makes it a good candidate for integration w/o large Capacitors

Requires charge pump or high voltage supply >Vin for good efficiency

Intrinsically good frequency response

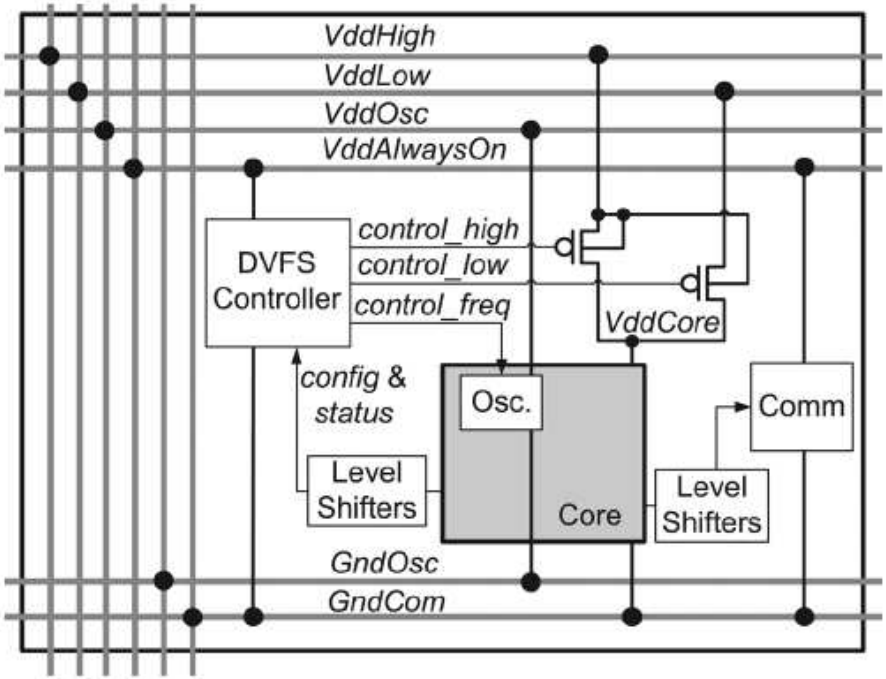
Charge-pump (or low current supply) allows Low drop out

Higher Efficiency

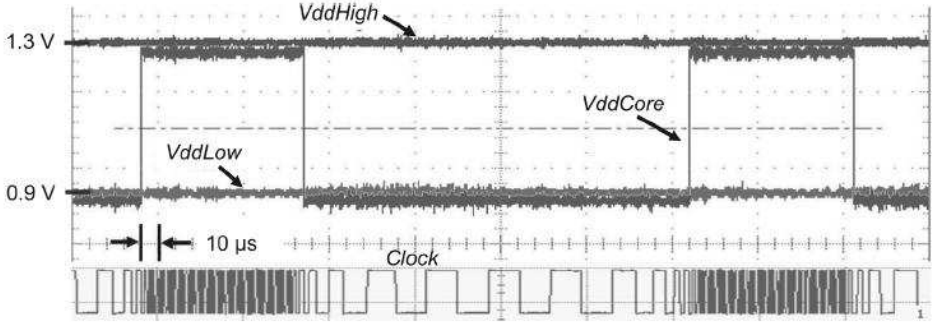
- Amplifier is high gain, low bandwidth.
- Bandwidth and efficiency depends on desired slew rate
- $0V_{th}$ devices offer additional efficiency

DVFS SYSTEM IMPLEMENTED WITH DUAL VOLTAGES

- Fully integrated167 processor system
- Each processor tile contains a core that operates at:
  - A fully-independent clock frequency
    - Any frequency below maximum
    - Halts, restarts, and changes arbitrarily
  - Dynamically-changeable supply voltage
    - *VddHigh* or *VddLow*
    - Disconnected for leakage reduction
    - Each power gate comprises 48 individually-controllable parallel transistors



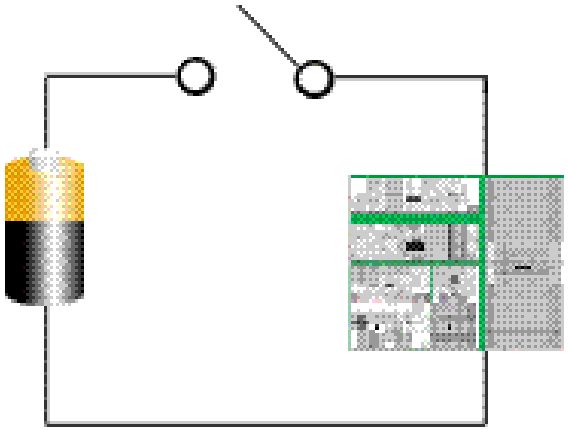
Dean N. Truong, Wayne H. Cheng, Tinoosh Mohsenin, Zhiyi Yu, Anthony T. Jacobson, Gouri Landge, Michael J. Meeuwsen, Christine Watnik, Anh T. Tran, Zhibin Xiao, Eric W. Work, Jeremy W. Webb, Paul V. Mejia, Bevan M. Baas, "167-Processor Computational Platform in 65 nm CMOS", JSSC, Vol. 44, No. 4, April 2009



POWER GATING

WHAT'S THE PROBLEM?
JUST TURN IT OFF WHEN YOUR NOT USING IT!

- Many choices and options
- Need to analyze your particular application for the optimal solution
- Considerations:
  - Design complexity
  - Power savings
  - Frequency/performance impact
  - Wake-up time
  - Area overhead
  - Verification complexity
  - Return on investment (ROI)



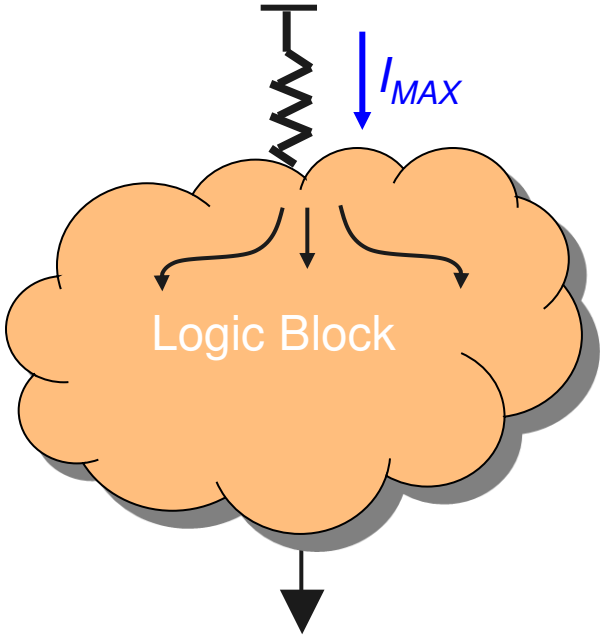
POWER GATING MODES (ACTIVE)

During active mode

- Natural low gate activity factors inside of logic block allow sharing of power gate device for many logic functions with gated logic block
  - Reduces resistance requirements of power gate device
  - Works to improve Ion/Ioff ratio requirements
- Design objective
  - Minimize resistance of cut-off device
- Penalties:
  - Increased logic delay
  - Requires higher external supply voltage for same frequency
    - Consumes more power during active mode

$$V_{LOGIC} = V_{dd} - V_{IRPG}$$

$$V_{IRPG} = R_{ON} \times I_{MAX}$$



POWER GATING MODES (SLEEP)

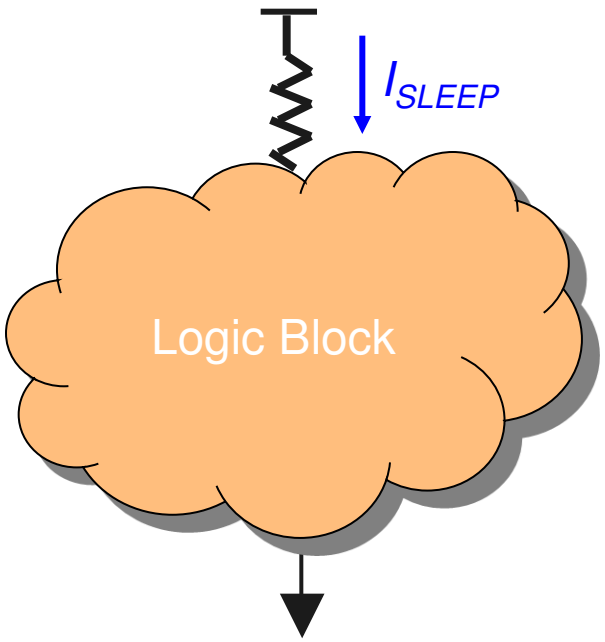
<u>During sleep mode</u>

- Steady state
  - Want to maximize OFF resistance of cut-off device
  - Reduce stand-by power with lower leakage

    (10x -> 1000x possible)
- Transition to cutoff
  - Creates supply *di/dt* for small I_{SLEEP} vs. large I_L

$$P_{SLEEP} = V_{dd} \times I_{SLEEP}$$

$$\frac{dI_{CUTOFF}}{dt} = \frac{(I_{SLEEP} - I_L)}{\Delta t_{CUTOFF}}$$

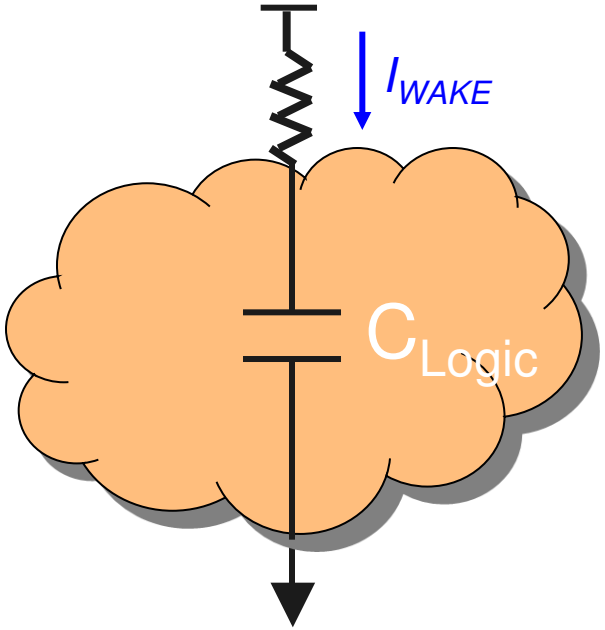


POWER GATING MODES (WAKE)

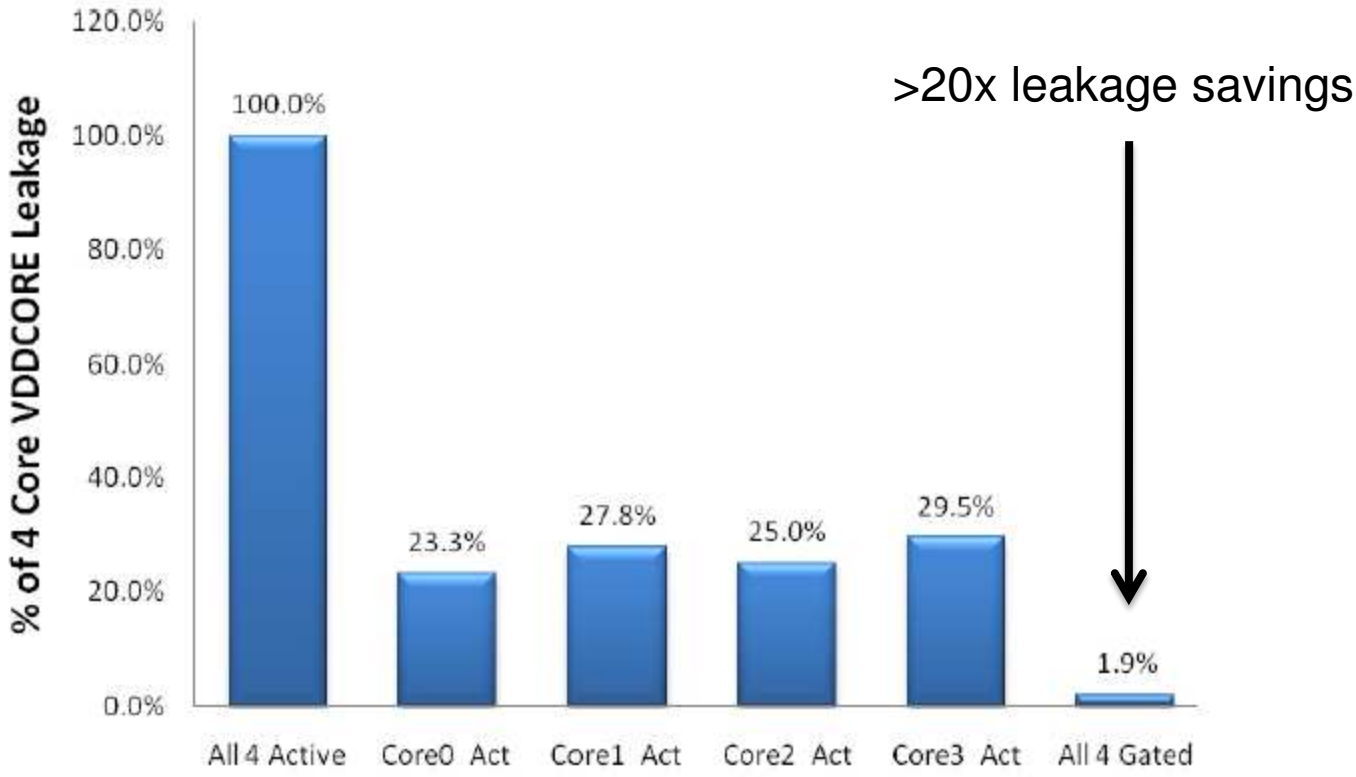
During wake-up mode

- Need to control *in-rush* current to charge logic devices to VDD
  - Logic capacitance is discharged to nV_{dd} *during sleep mode*

$$I_{WAKE} \approx \frac{(1-n)V_{dd}}{R_{WAKE}}$$



LEAKAGE SAVINGS WITH POWER GATING



Early Llano Measured Results

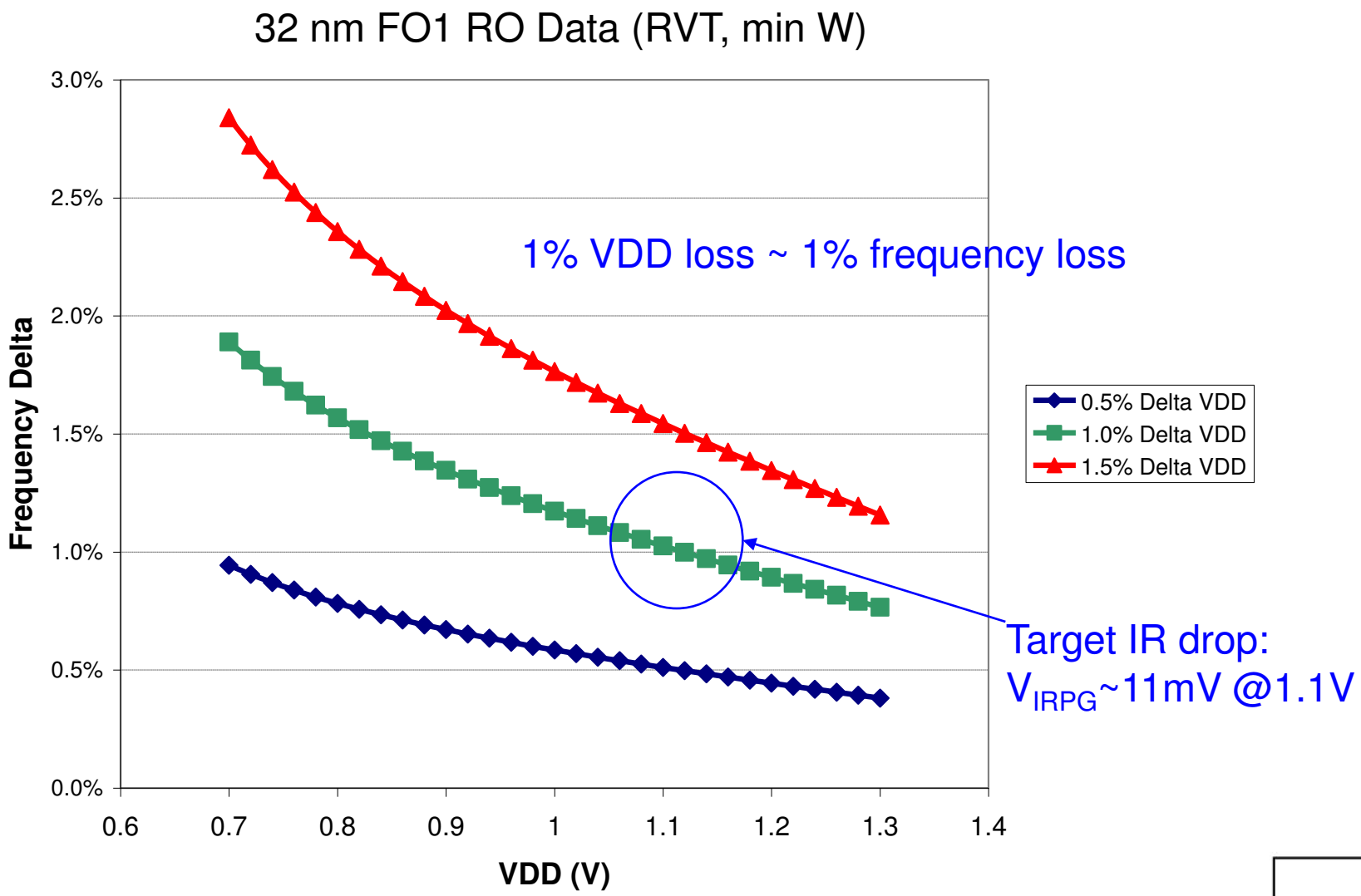
Very effective method for controlling power during idle periods

Ravi Jotwani, Sriram Sundaram, Stephen Kosonocky, Alex Schaefer, Victor F. Andrade, Amy Novak, Samuel Naffziger, "An x86-64 Core in 32 nm SOI CMOS", JSSC, January, 2011

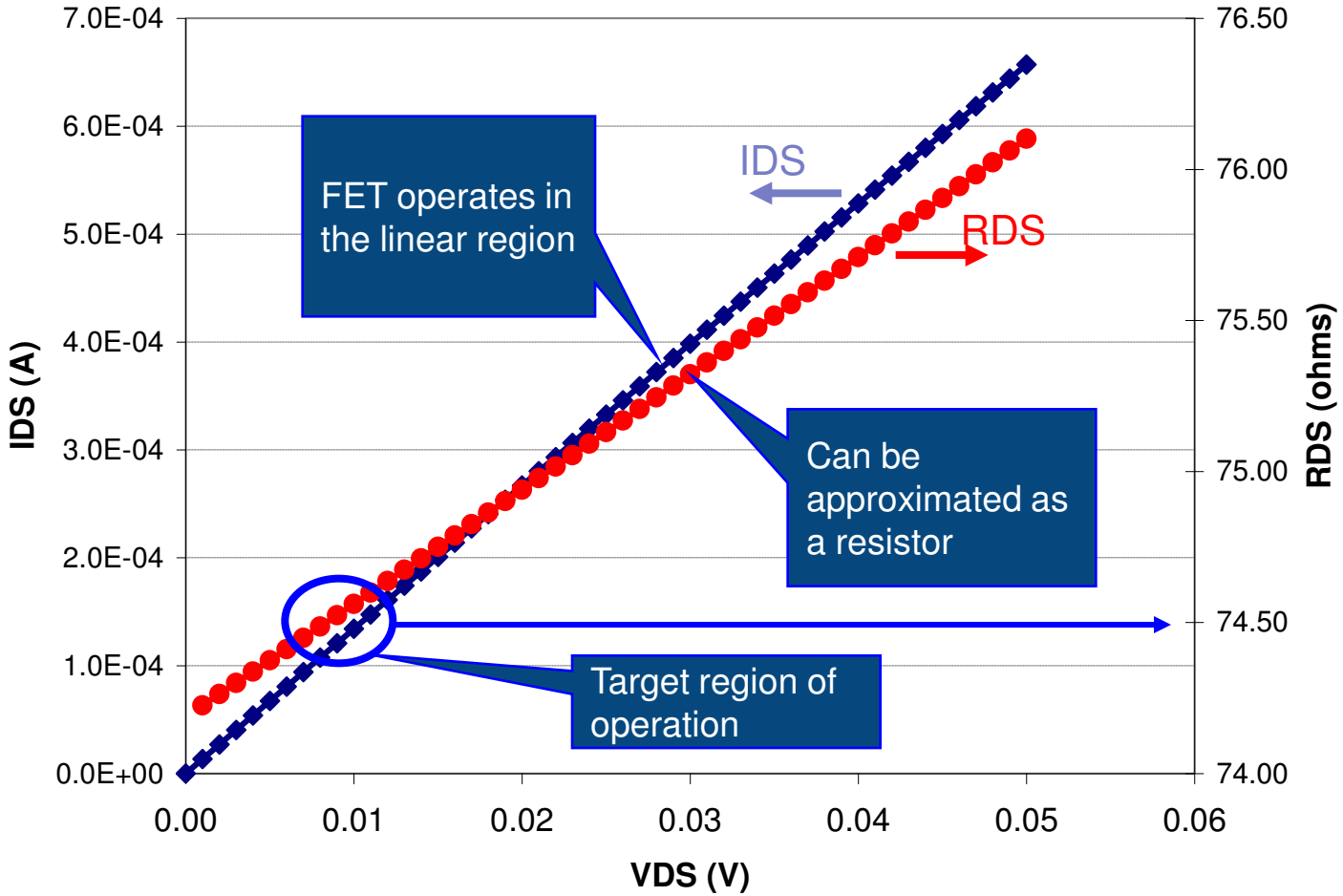
DETERMINATION OF FREQUENCY IMPACT DURING ACTIVE MODE

- Determine tolerable voltage drop for active mode operation
  - De-rate timing models by fixed maximum value to account for voltage loss in power switch
  - Increase external voltage to compensate for voltage drop
- Use critical path simulations or representative RO data for voltage/frequency trade-off
- Design worst-case power gate and grid resistance to meet specification target
- Analyze drop at highest operating voltage and highest power consumption
- Keep voltage drop due to power gating small to minimize uncertainty at power island boundary crossings
- 2nd order effects:
  - Circuit area growth due to wire blockages caused by power switch
  - Decreased available quiet decoupling capacitance due to increased resistance to implicit and explicit charge reservoirs on the die

FREQUENCY IMPACT ESTIMATED WITH RING OSCILLATOR

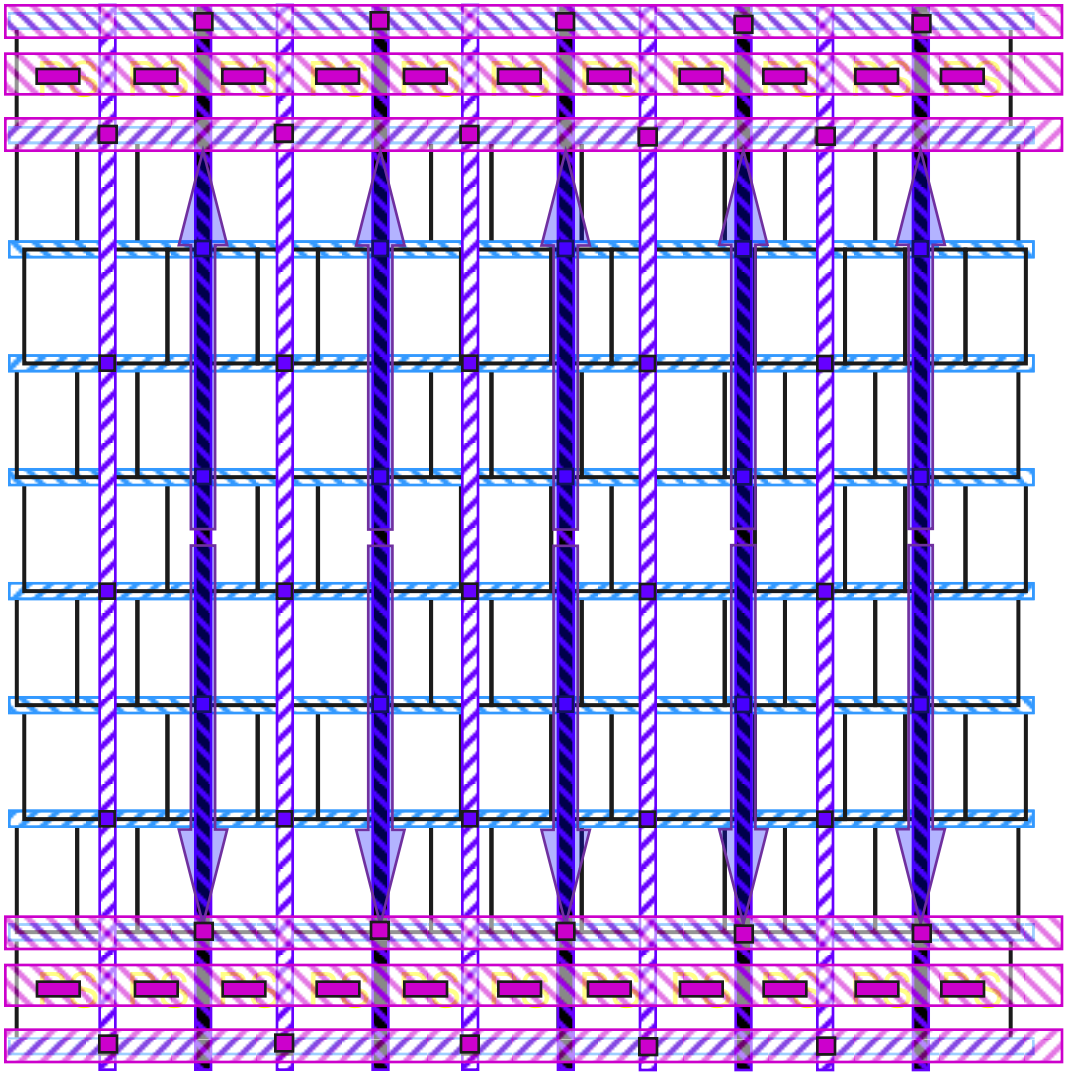


POWER GATE CELL I-V CHARACTERISTICS



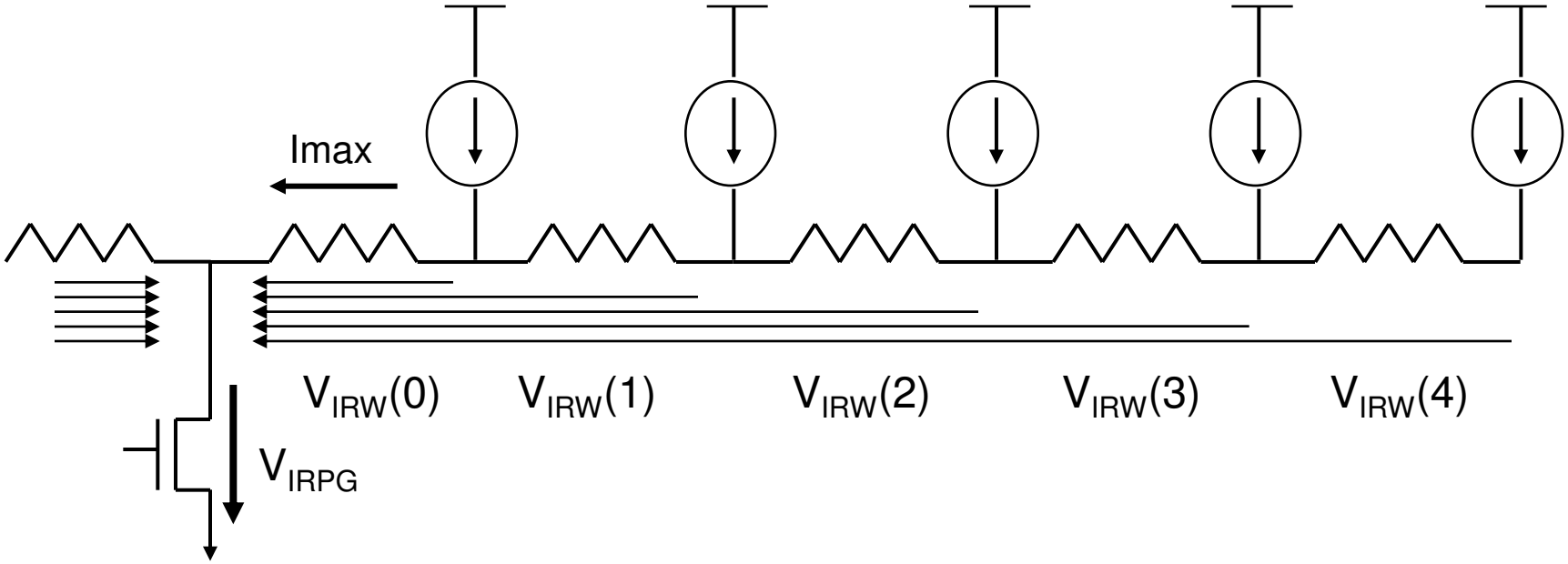
EMBEDDED ROW-BASED POWER SWITCH PLACEMENT

Current Flow



Pre-populate logic area with power gate cells at uniform locations prior to place and route

MAXIMUM POWER GATE SPACING
(ROW-BASED PLACEMENT)

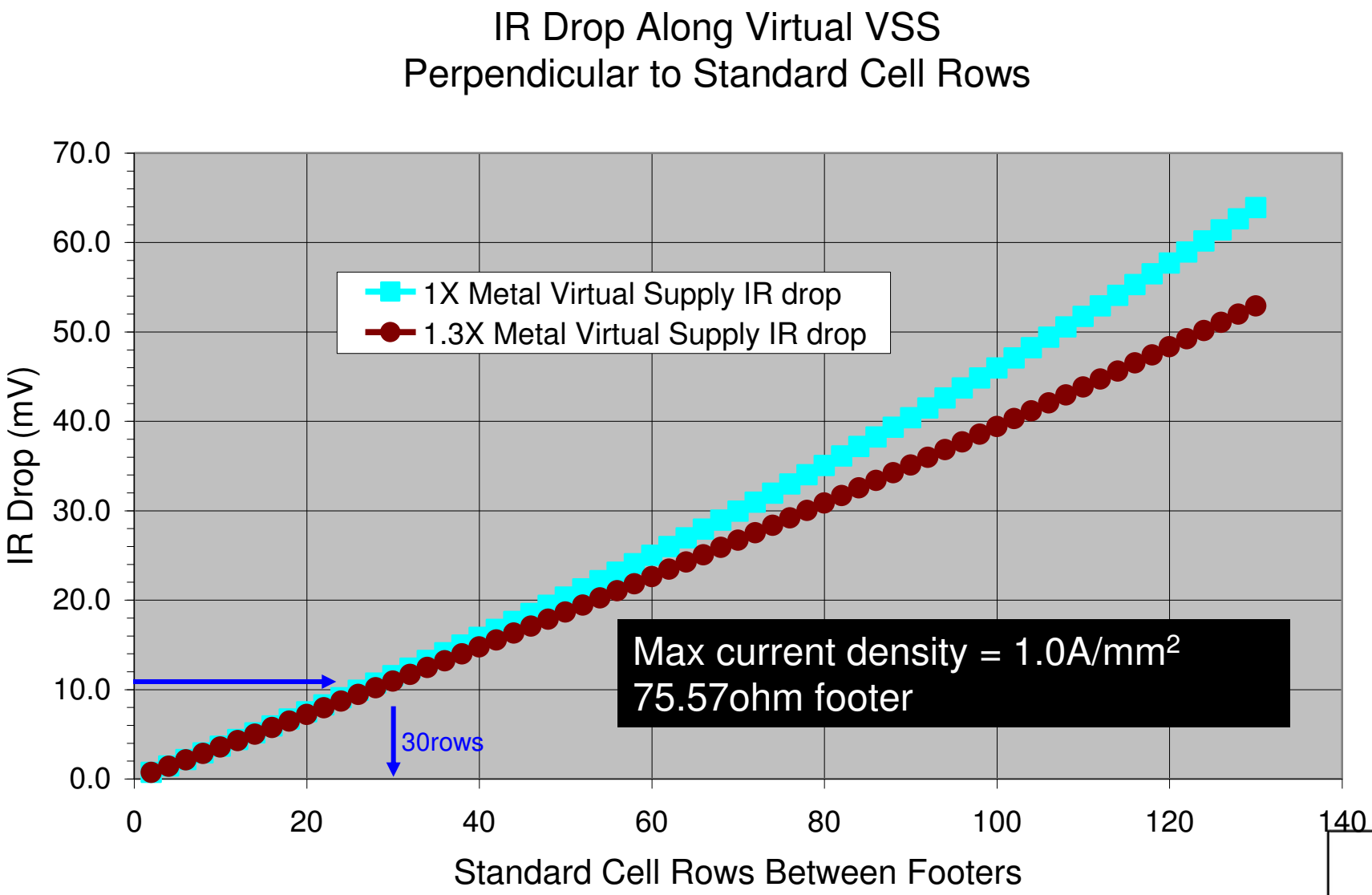


- Assume uniform current density
- Can estimate power gate spacing with respect to IR and EM limits

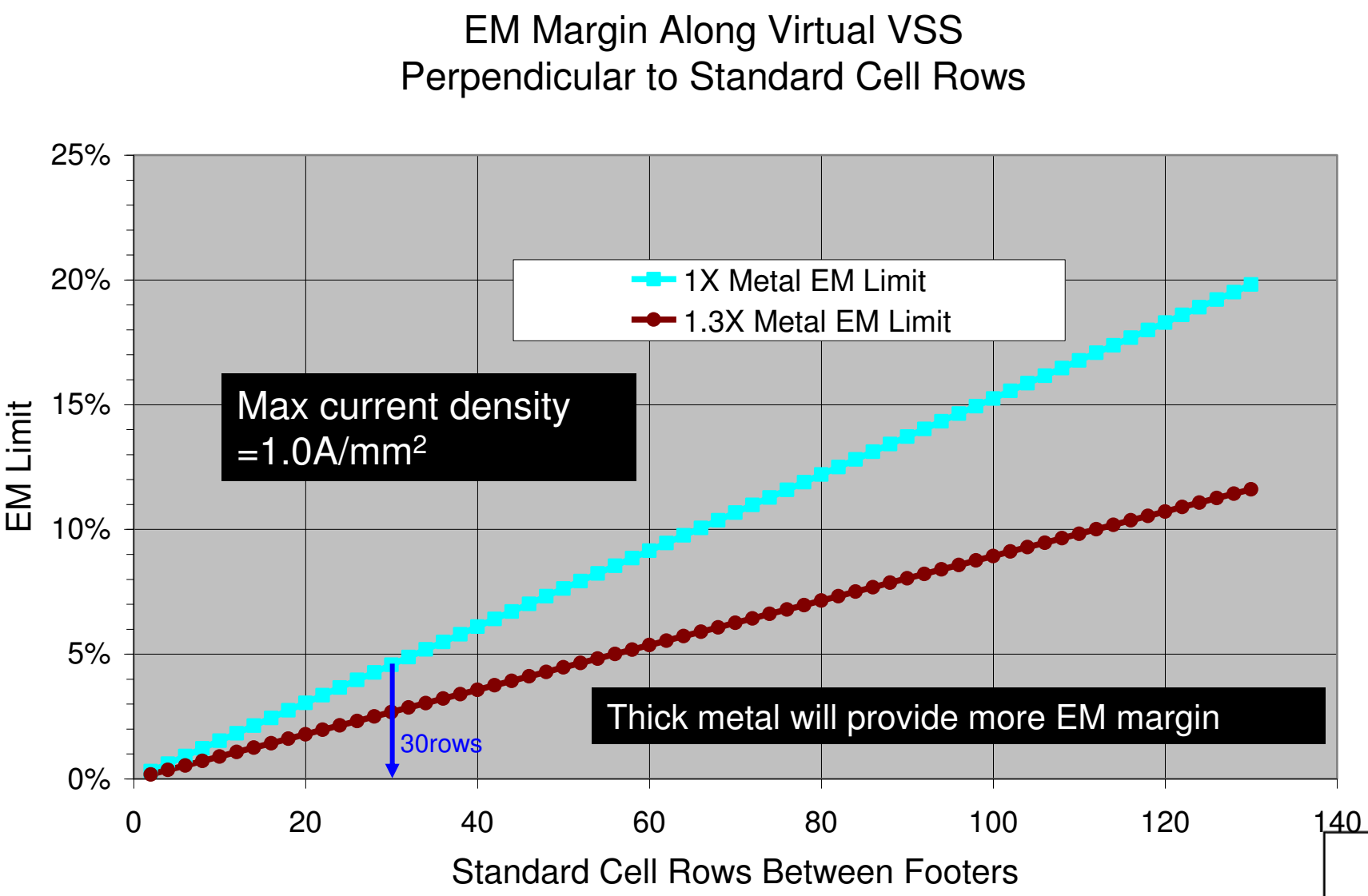
$$V_{IRPG} = 2 * rows * I_{ROW} * R_{PG}$$

$$V_{IR\_WC} = V_{IRPG} + \sum V_{IRW}$$

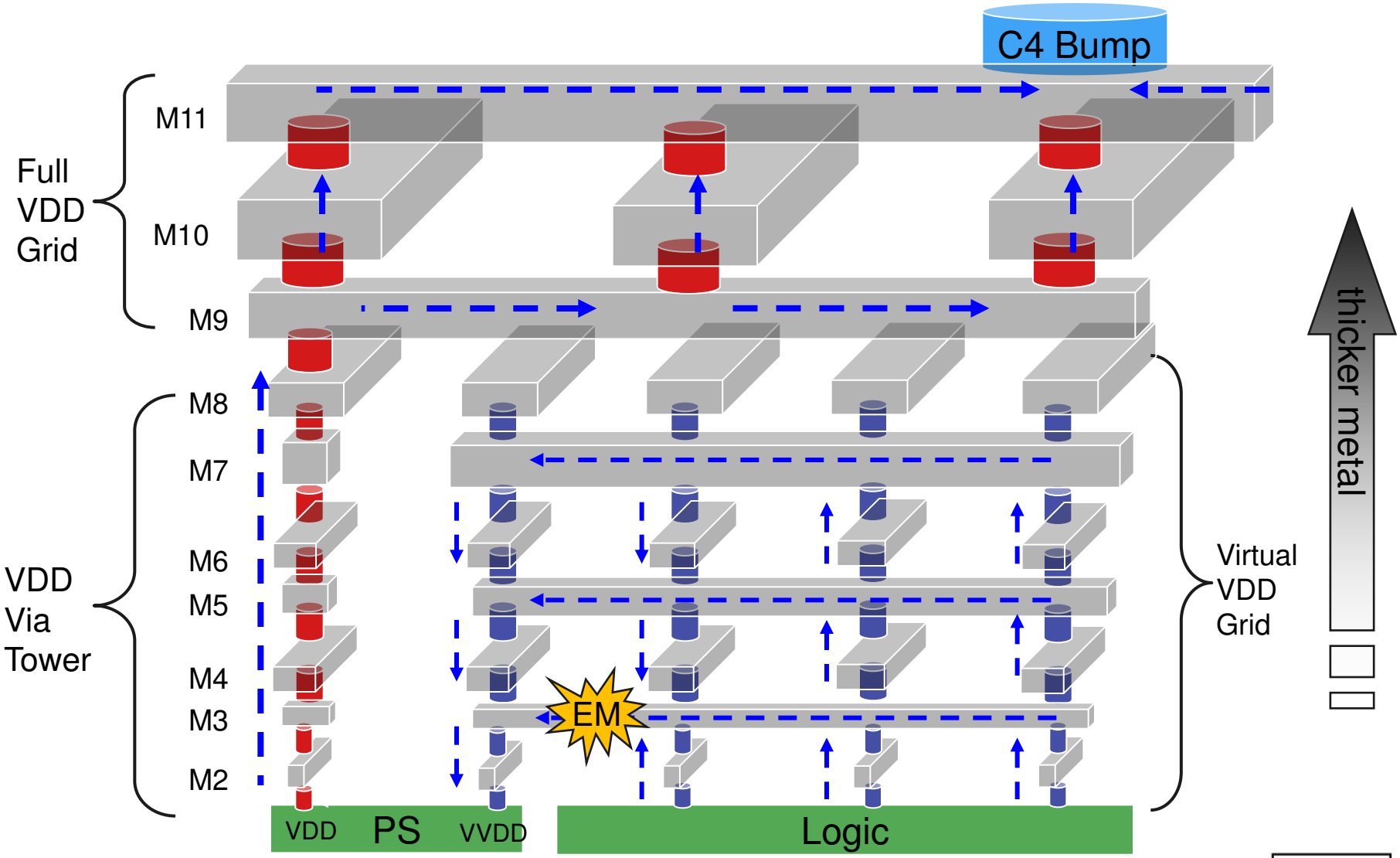
IR DROP EXAMPLE (UNIFORM POWER DENSITY)



EM EXAMPLE (UNIFORM POWER DENSITY)



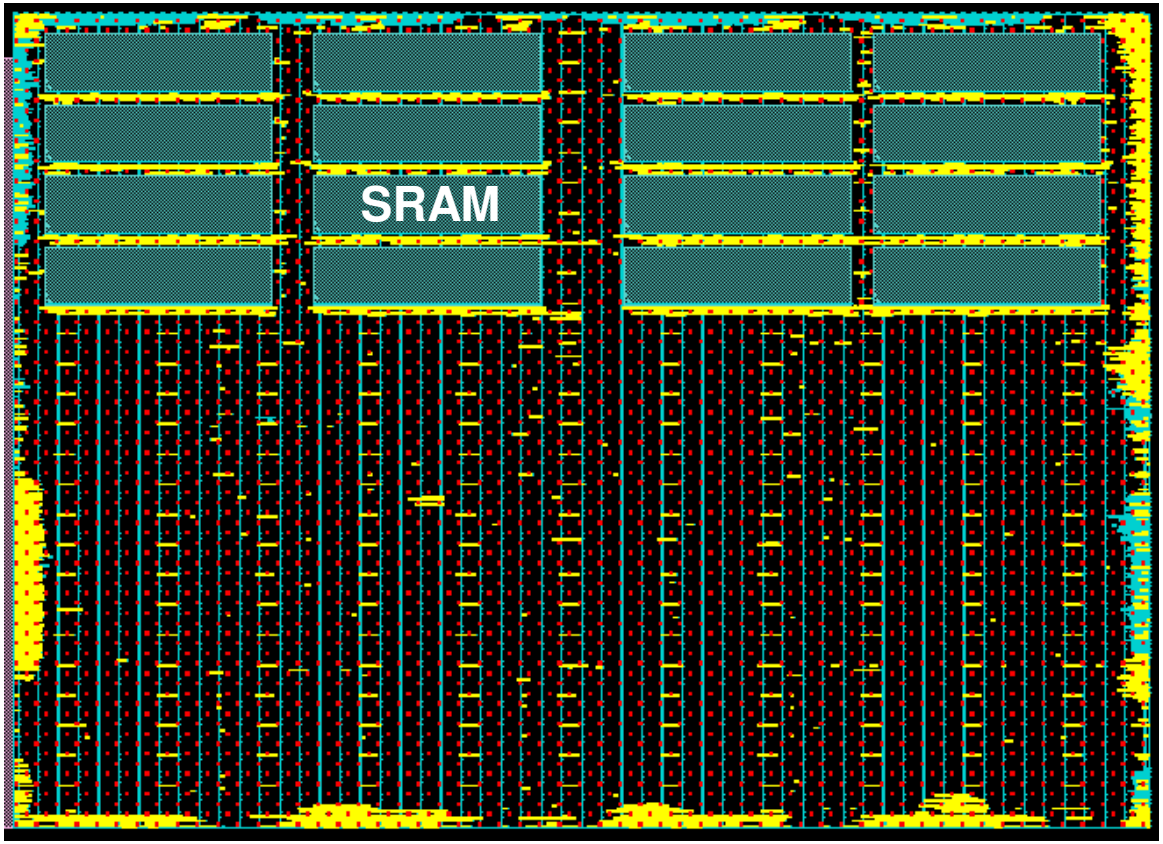
VERTICAL CROSS SECTION



EM hazards can occur in high power density regions

EMBEDDED POWER GATE EXAMPLE
STYLE USED IN LLANO GPU

AMD GPU Functional Unit Power Gate Tile



Power gate:

Checkerboard pattern

Always-on cells:

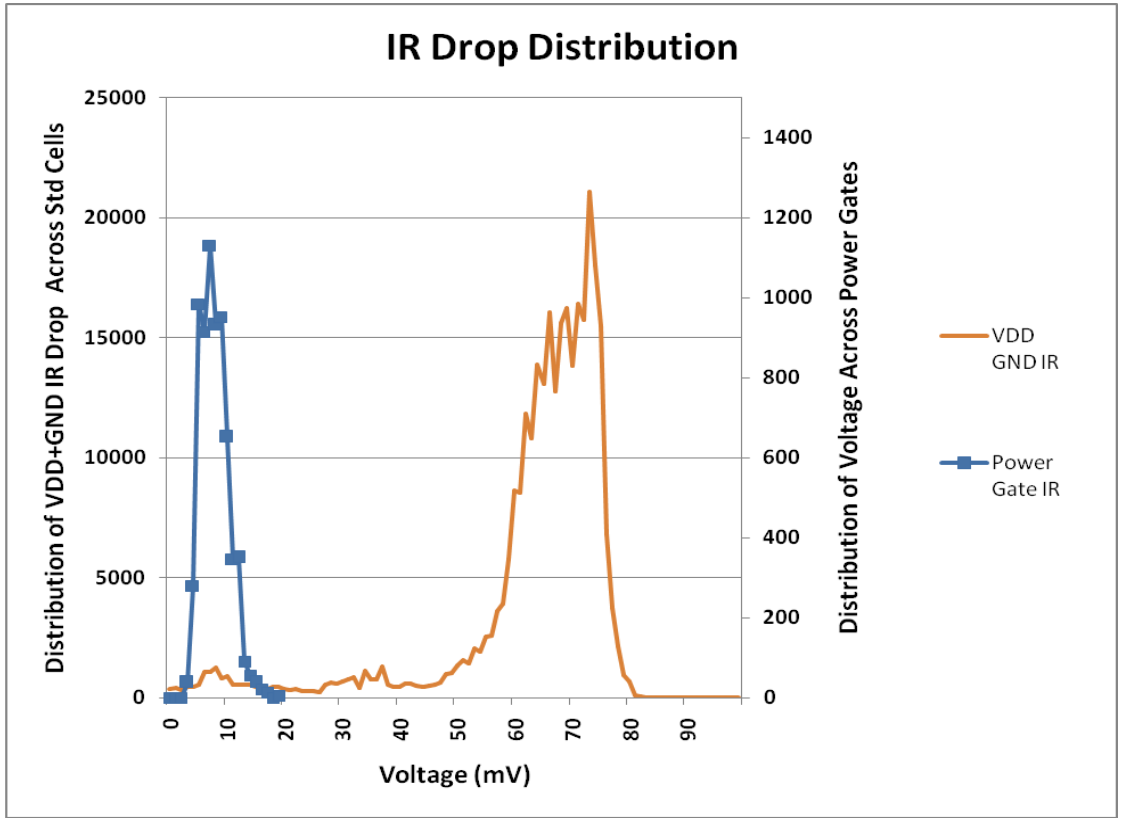
FeedThru cells, Power control logic, Clamp cells for macro input pins and tile output ports.

Input I/O buffers

Power gate cells must squeeze around arrays
Providing power to always-on cells can be problematic
IP I/O protection typically needed to isolate individual regions
SRAMs need their own custom power gating

EMBEDDED POWER GATE EXAMPLE

AMD GPU Functional Unit Power Gate Tile

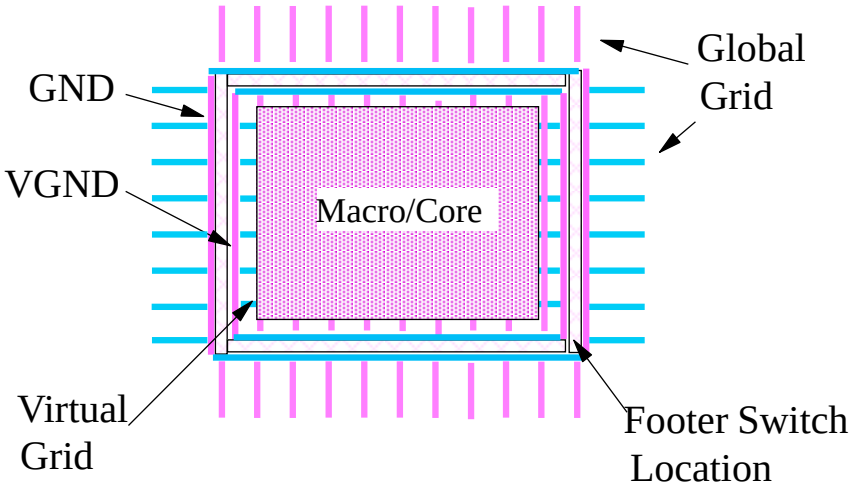


Simulated with Apache Redhawk in vectorless dynamic analysis mode

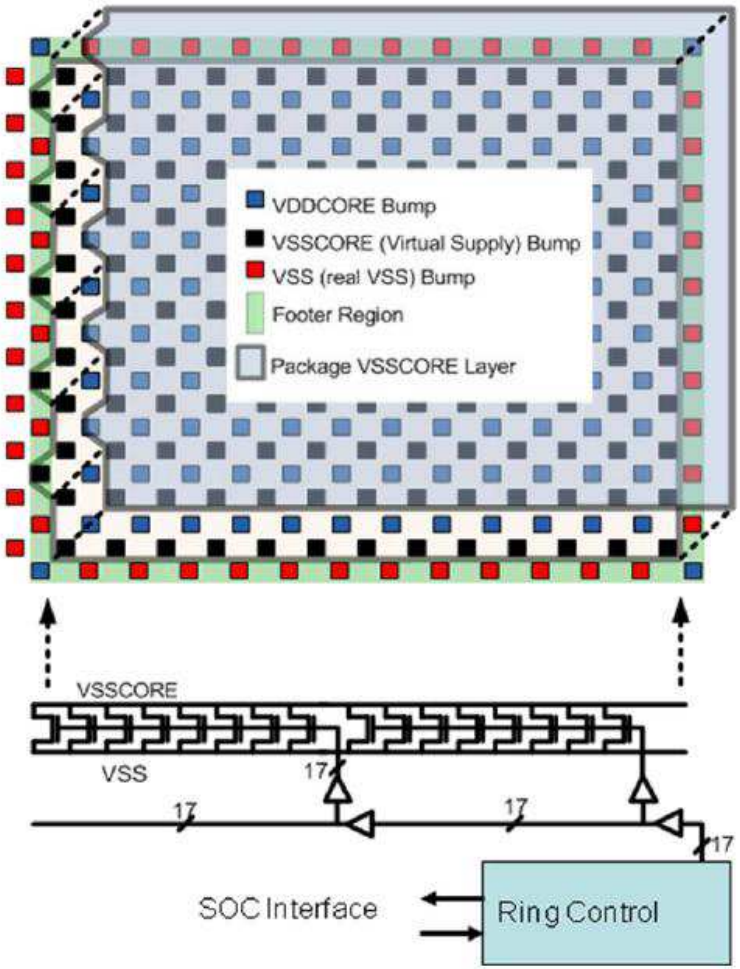
Power gating increases distribution of IR drop on grid
Can cause problems with critical circuits and paths

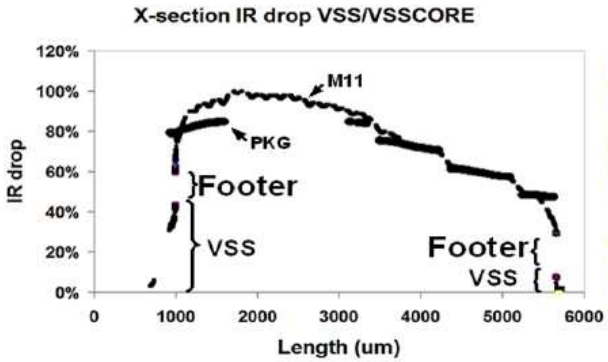
EXTERNAL RING STYLE GATING

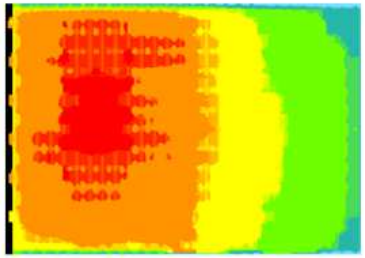
- Interrupt global supply grid to power-gated IP block
- Requires careful chip floorplanning to accommodate supply interruption
- Increased sharing of power gate devices can ease power gate sizing
  - Large IP power gating can greatly ease worst-case current analysis
- Can be difficult to create always power-on islands with power-gated region



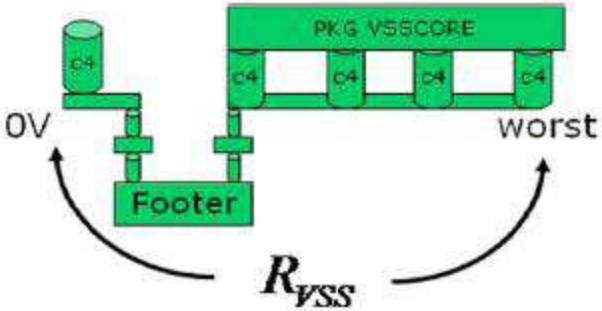
LLANO CPU CORE RING GATING WITH PACKAGE LAYER ASSIST







Virtual voltage spreads uniformly
Bumps near hot spots can exceed max limits



Added Effective VSS Resistance

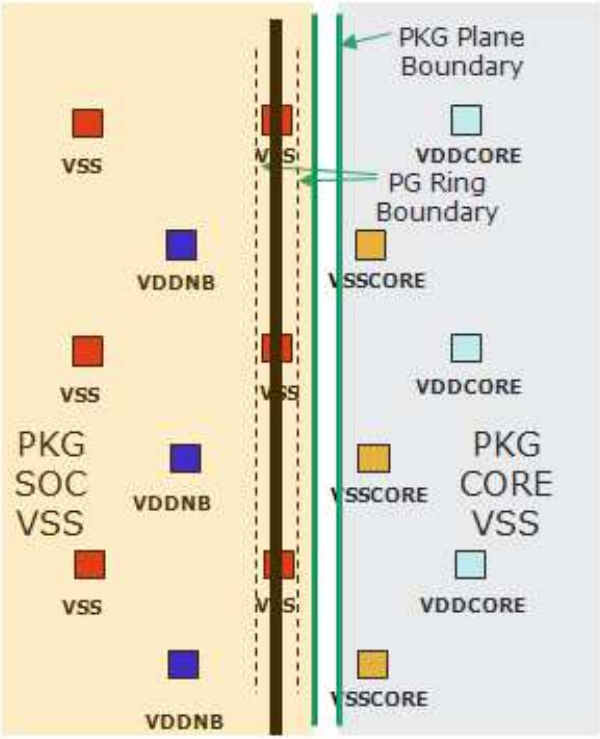
$$R_{VSS} = 1.1m\Omega = \frac{VSS_{Droop}}{I_{TDP}}$$

High R_{VSS} can create noise issues

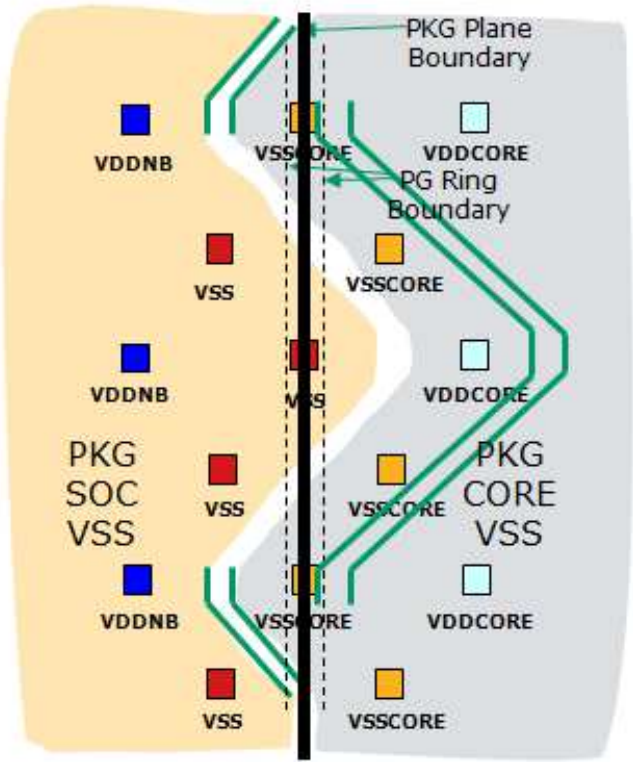
Ravi Jotwani, Sriram Sundaram, Stephen Kosonocky, Alex Schaefer, Victor F. Andrade, Amy Novak, Samuel Naffziger, "An x86-64 Core in 32 nm SOI CMOS", JSSC, January, 2011

STRAIGHT EDGE & ZIG/ZAG EDGE PACKAGE PLANE BOUNDARIES

Less restrictions on bump locations
- Saves Area



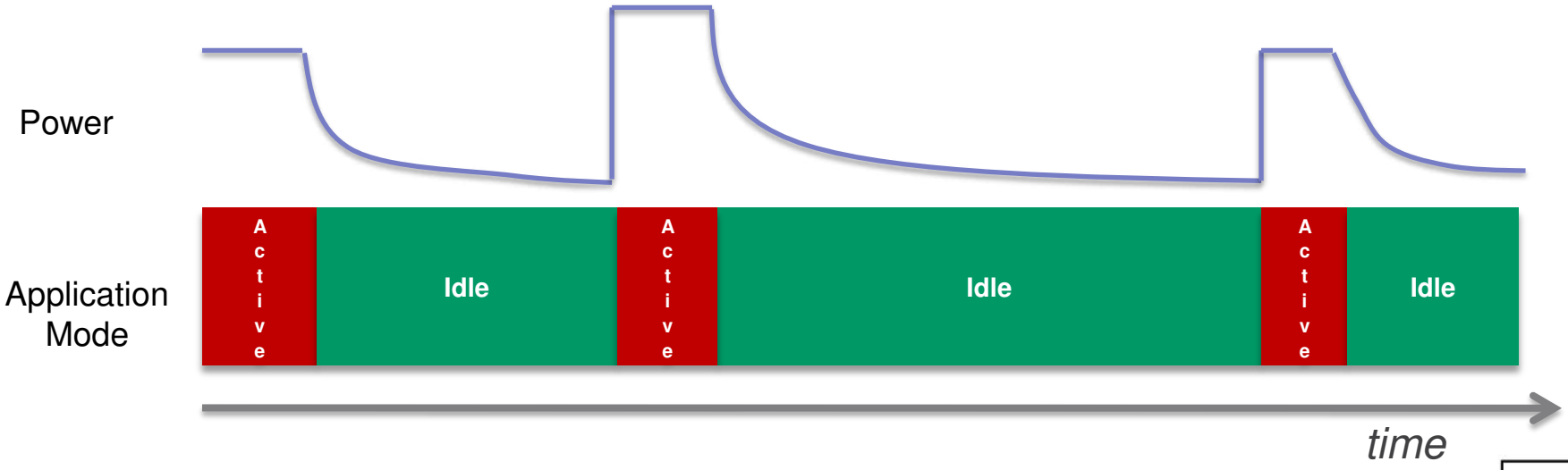
Allows 1.5x peripheral C4 boundary between VSS & VSSCORE
- Higher max current capacity



Peak currents flowing into bumps at periphery can be alleviated using zig/zag approach

MAXIMIZING POWER GATING OPPORTUNITIES

- To maximize power savings, we would like to go into a power gating mode as soon as possible to reduce leakage power
- Practical constraints limit opportunity
  - Energy break-even point
  - In-rush current induced noise
  - Time to restore state of power gated region



ENERGY BREAK-EVEN POINT

Power Gate Example:

T0: Logic block starts inactivity

T1: Control circuit decides to power gate

T2: Power gate signal propagation complete

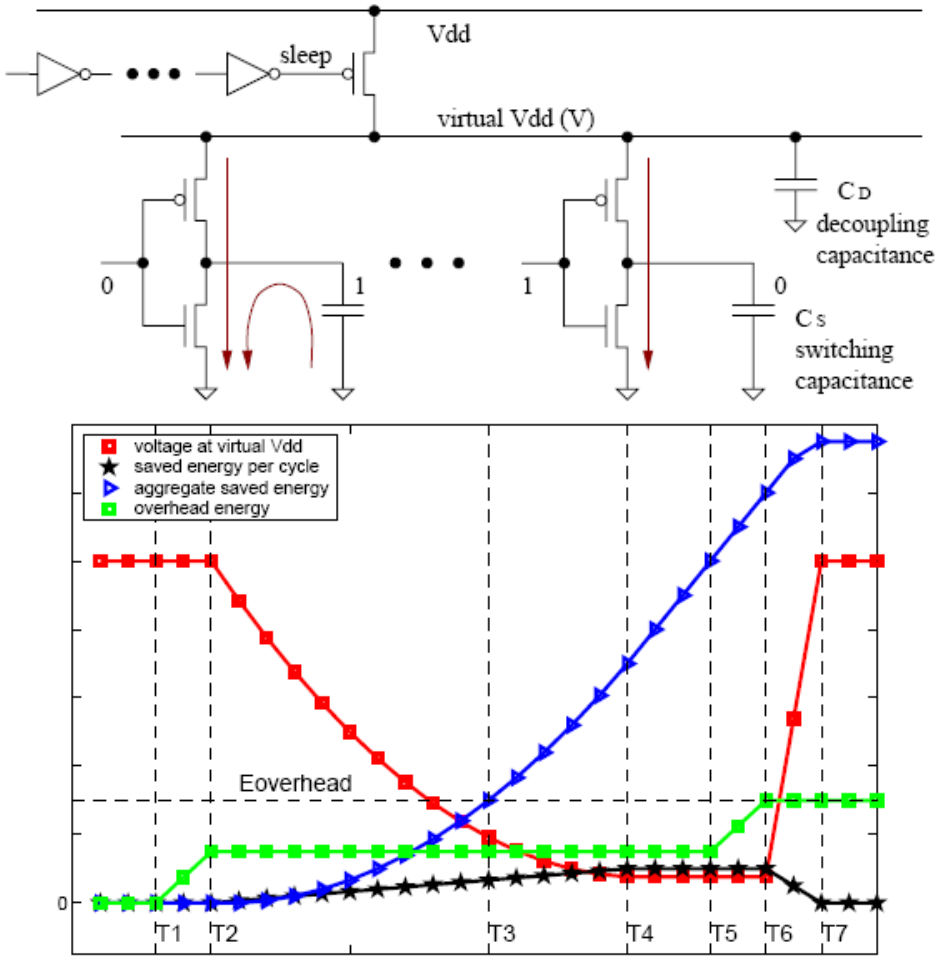
T3: Energy break-even point

T4: Full discharge of virtual supply

T5: Control circuit decides to re-power logic

T6: Power gate signal propagation complete

T7: Gated logic is fully charged and ready for activity



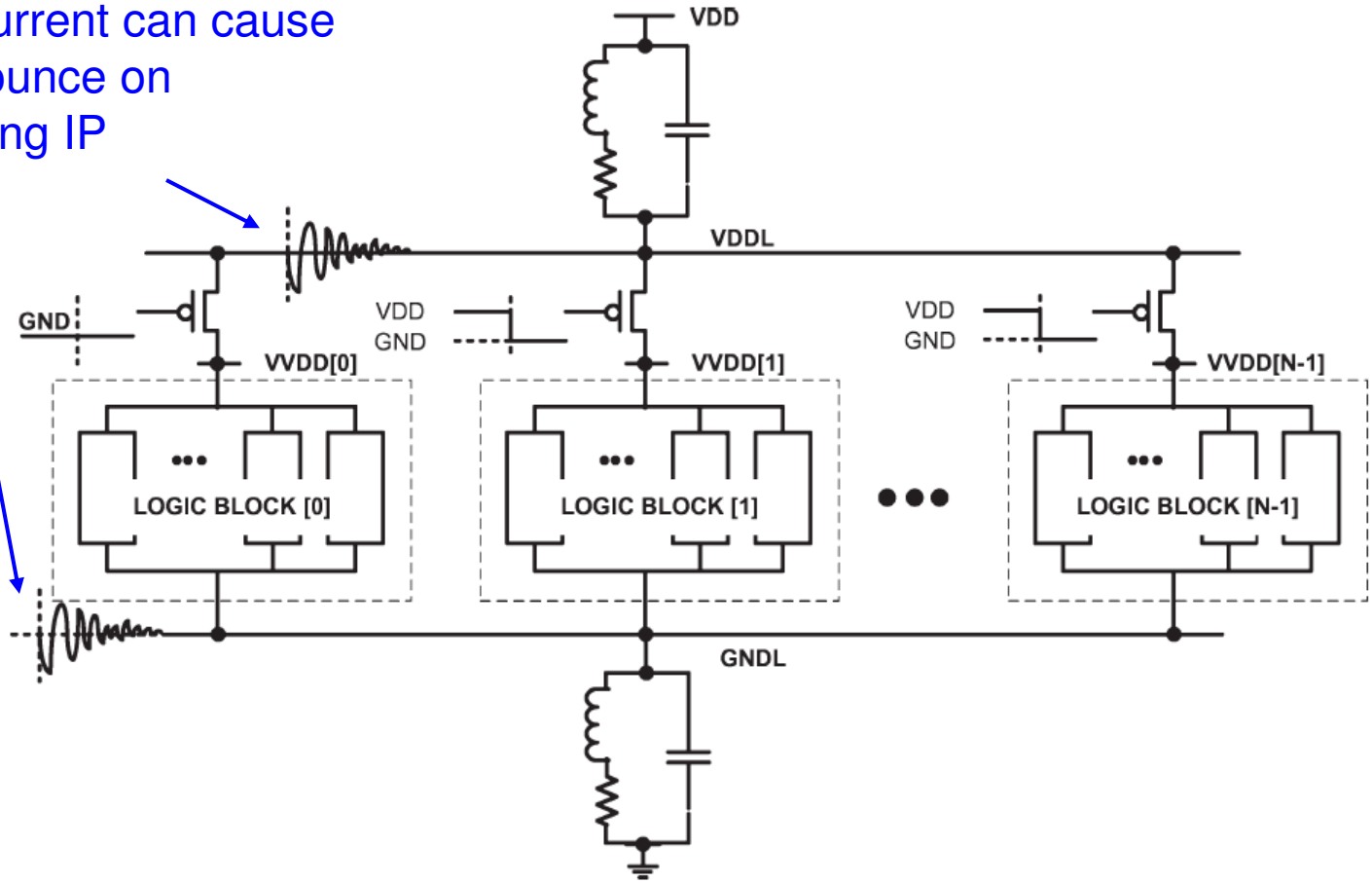
Z. Hu et al., "Microarchitectural Techniques for Power Gating of Execution Units," ISLPED'04, August 9–11, 2004, Newport Beach, California, USA.

IN-RUSH CURRENT CONTROL

- During wake-up of the power gated region it's critical to control the in-rush current
- The effective resistance of parallel combination of footers is very small
  - Can create excessive currents
- Transition to cutoff mode can also create similar *di/dt* events
- Hazards
  - Supply bounce from large *di/dt*
  - EM violations
  - Short circuit currents from imbalanced nodes in power gated region during wake-up

UNDERSTANDING SUPPLY BOUNCE

In-rush current can cause supply bounce on neighboring IP



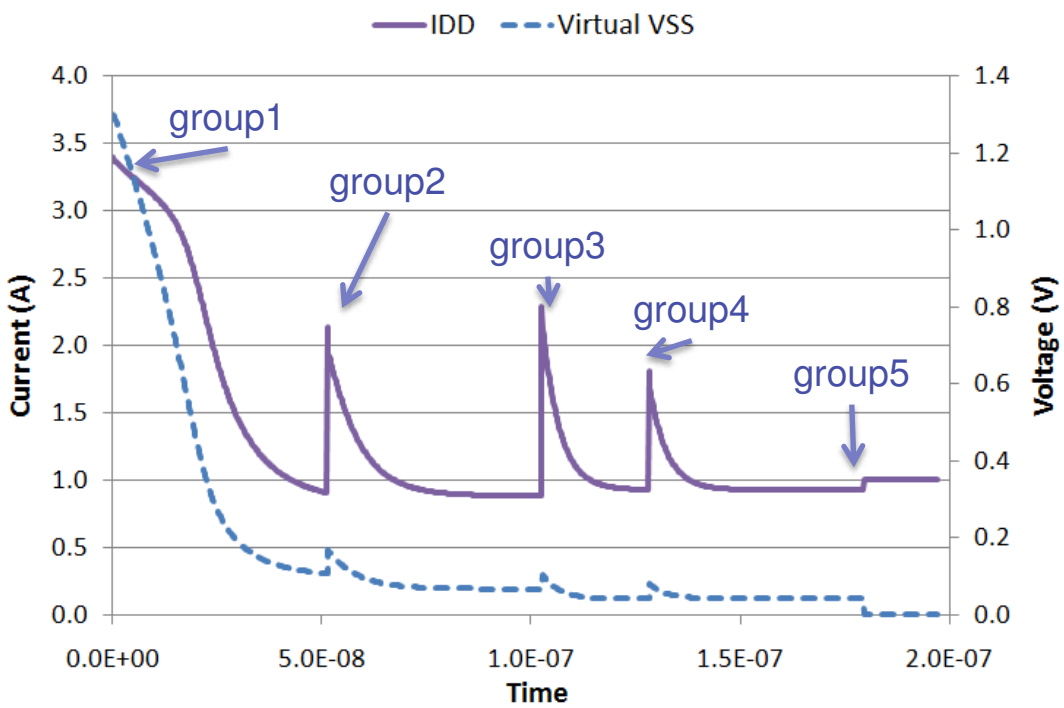
Suhwan Kim, Chang Jun Choi, Deog-Kyoon Jeong, Stephen Kosonocky, Sung Bae Park, "Reducing Ground-Bounce Noise and Stabilizing the Data-Retention Voltage of Power-Gating Structures", IEEE Transactions on Electron Devices, Vol. 55, NO. 1, January 2008

IN-RUSH CURRENT SIDE-EFFECTS

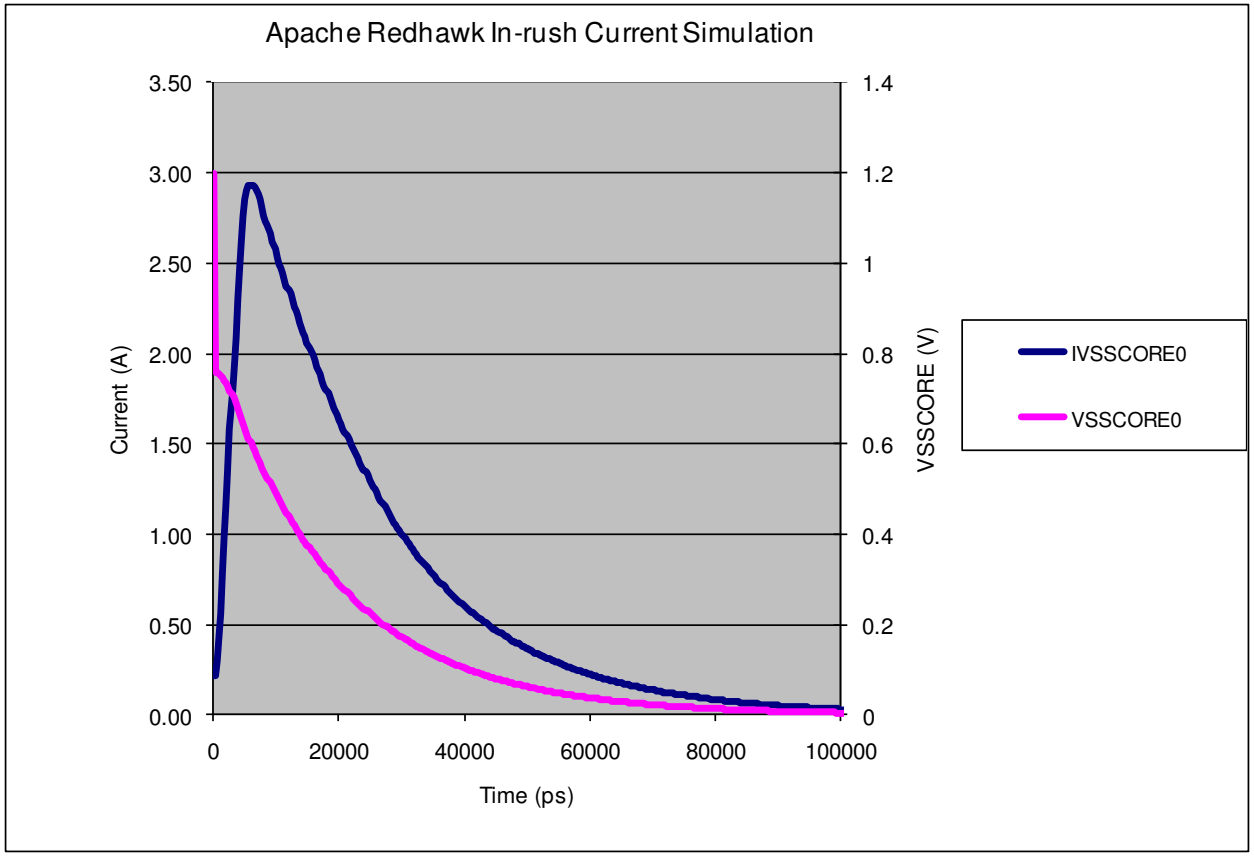
- Disturb neighboring circuits
  - Retention flops, adjacent IP logic, SRAM
    - Supply bounce, ground bounce
- Over stress gate-oxides due to voltage excursions
- Un-even distribution of wake signals during power-up
  - Creates excessive short circuit currents when some gates are charged and others are not
- Large di/dt can cause EM failures

PROGRAMMABLE IN-RUSH CURRENT CONTROL

- When exiting power gating, power gates are gradually enabled in groups with progressively larger effective device widths
  - Controlled by FSM
  - Fuse programmable strengths and duration for post-silicon tuning
  - Analytical R-C model w/o inductance demonstrating control

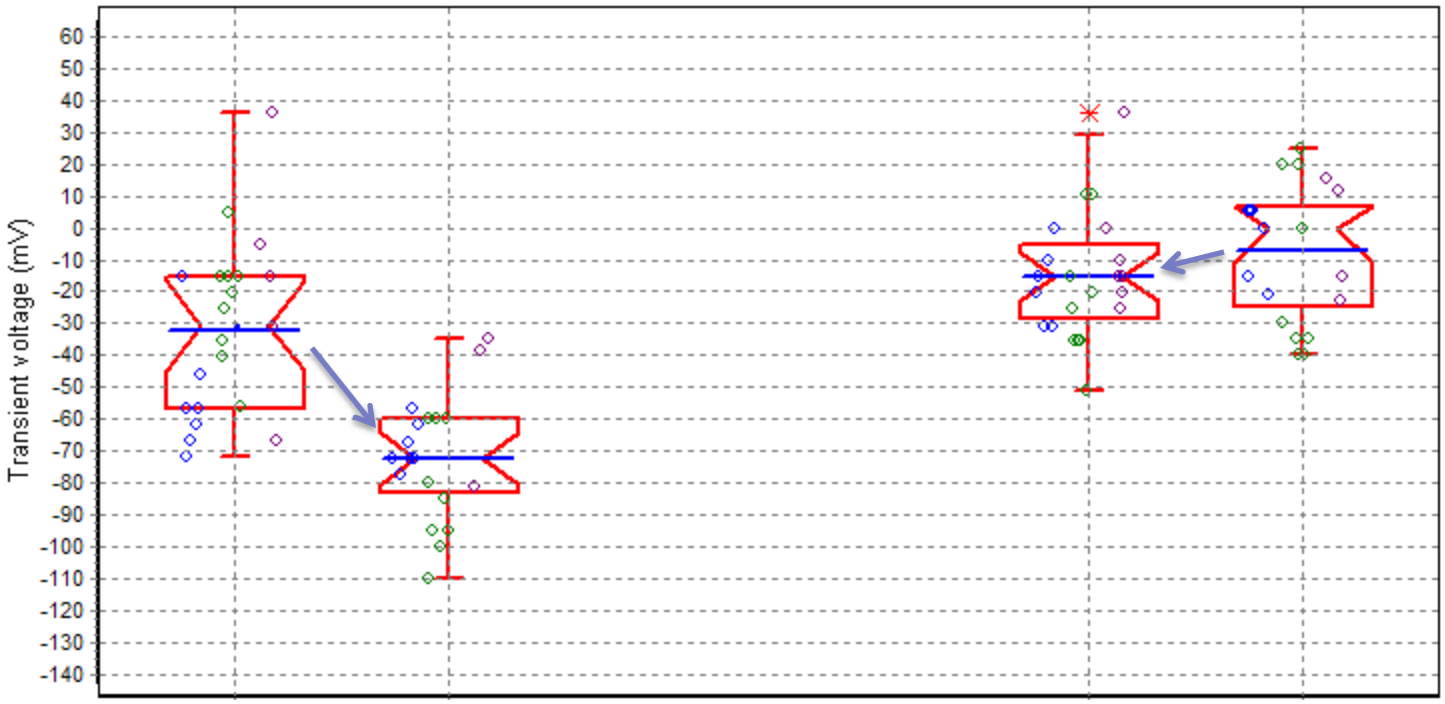


LLANO CPU CORE IN-RUSH CURRENT ON CC6 EXIT



Fuse programmable in-rush control built-in
Sample setting shows a 3A peak during CC6 exit

HW DATA SHOWING IN-RUSH EFFECT ON OTHER CORES



1 cores in/out CC6
3 core max freq

3 cores in/out CC6
1 core max freq

1 cores in/out CC6
3 core max freq

3 cores in/out CC6
1 core max freq

Large in-rush setting

Small in-rush setting

Transient droop limited by CC6 exit

Transient droop limited by # cores running max frequency

STATE SAVE AND RESTORE

- Regulate virtual supply during sleep mode to low voltage
  - Moderate leakage reduction possible
  - Globally applies to all retention elements in gated region
  - Example:
    - K. Kumagai et al., "A Novel Powering-down Scheme for Low Vt CMOS Circuits," Symposium on VLSI Circuits, 1998 (Virtual Rail CMOS).
    - L. Clark, S. Demmons, "Standby Power Management for a 0.18µm Microprocessor," ISLPED 2002 (Intel XScale processor).
- Retention flops with always-on latch cell
  - Many variants possible
  - Some with explicit control or automatic restore
  - Good practice to avoid adding power domain crossing in functional path of flop
  - Can increase timing uncertainty with virtual supply bounce
  - Reduces requirements of always-on supply distribution
- Write-out critical state to memory using serial or parallel paths
  - Can utilize scan mechanism to avoid a dedicate bus
  - Low area overhead
  - Can be a large time overhead depending on total size of state restore memory

CONCLUSION

CONCLUSION

- DVFS provides a nice capability for optimization of CPU/GPU/APU performance per application
  - As we add more cores and IP to the SOC, it's will be difficult to continue to provide unique voltage rails with external regulators
  - Integrated regulation has it's own challenges, more details will be covered by the other presenters
- Power gating can mitigate the need for additional voltage rails
  - Maximum frequency impact is minimal
  - Integration by embedding in IP or surrounding in a ring each have their own issues
  - Opportunity is limited by energy breakeven, in-rush noise, and state restore overhead
  - In-rush current control is necessary to prevent frequency impacts