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

Now: Parallel/Data-Dense
16:9 @ 7 megapixels
HD video flipcams, phones, webcams (1GB)
3D Internet apps and HD video online, social networking w/HD files
3D Blu-ray HD
Multi-touch, facial/gesture/voice recognition + mouse & keyboard
All day computing (8+ Hours)

Immersive and interactive performance

- Samuel Naffziger, VLSI Symposium 2011, Kyoto
Power/Performance Trade-Off

Processor Power-Frequency/Voltage

Performance/Watt maximized at the lowest voltage operating point

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
# APU Power Management

## Hardware Policy
- **DDR PHY**
  - Self-Refresh
  - PreChrg. Pwr-Dn
  - PLL Off

- **IO PHY**
  - Link width
  - Link suspend
  - PLLs Off

- **CPU**
  - DVFS
  - Deep Pdn
  - PLLs off

- **GPU**
  - DVFS
  - T-put scale
  - Power Dn

- **NB**
  - DVFS
  - Power Dn

## AC/DC Policy
- **Speed, physical interfacechg**

- **Link rate change**

- **Max PState**

- **Max PState**

- **Max PState**

## Boost
- **DVFS**

- **T-put Scale**

- **DVFS**

## OS
- **P-States**

- **Halt**

## Driver
- **Power Dn**

- **Power Dn**

- **Set PStates**

---

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

---

11 | Practical Power Gating and Dynamic Voltage/Frequency Scaling Issues | August 17, 2011 | Stephen Kosonocky
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

Diagram:
- **P-states**: Core executing at some frequency
  - C1/C2: Core halted and frequency ramped down
  - C3/C4: Core halted, clocks ramped to even lower frequency, maintain caches, maintain core
  - C5: Core halted, clocks ramped to even lower frequency, flush caches, maintain core
  - C6: Core off, flush caches, core state saved to DRAM
**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
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](image-url)
TYPICAL PACKAGE LAYER ROUTING LAYERS

- **VDD**
- **VDDIO**
- **VDDA**
- **VDDR**
- **VDDNB**

**M1** – PWR (top metal layer)  
**M2** – routing layer 1

**Package capacitors**  
Space limited for capacitors

**Routing congestion**
**CORE VOLTAGE POWER DELIVERY IN PACKAGE**

- 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?

25ns trace of DGEMM benchmark

VDD

Normalized Current

Time

X86 core Supply simulation w/ package board model

Increased local decoupling cap could help
PACKAGE CAPS FOR SUPPLY DECOUPLING

Most effective

Low impedance path to the cap due to closely spaced planes

Marginally effective

High impedance path to the cap due to few long vias

Hard to find places for additional components
LARGE ON-DIE DECOUPLING CAPACITOR FOR REDUCTION OF SUPPLY NOISE

Trench Capacitor

Shift resonant frequency from 80MHz -> 7MHz


100fF/um² ~ 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

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

CMOS Inductor vs. Small Package size components

- Embedded inductors need 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

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 \((V_{in} - V_{out})\)
  \[
  \text{Eff} = \frac{\text{Power Out}}{\text{Power In}} = \sim \frac{V_{out}}{V_{in}}
  \]
- 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\textsubscript{in}) range.

Required gain-bandwidth product challenging high for processor loads:

- \( \text{BW} \sim 1\text{GHz} \) 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

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

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
DVFS SYSTEM IMPLEMENTED WITH DUAL VOLTAGES

- Fully integrated 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
    - $V_{ddHigh}$ or $V_{ddLow}$
    - Disconnected for leakage reduction
  - Each power gate comprises 48 individually-controllable parallel transistors

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

\[
V_{\text{LOGIC}} = V_{dd} - V_{\text{IRPG}} \\
V_{\text{IRPG}} = R_{ON} \times I_{MAX}
\]
During sleep mode

- **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 $I_L$

\[
P_{SLEEP} = V_{dd} \times I_{SLEEP}
\]

\[
\frac{dI_{CUTOFF}}{dt} = \left( \frac{I_{SLEEP} - I_L}{\Delta t_{CUTOFF}} \right)
\]
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


>20x leakage savings
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

1% VDD loss ~ 1% frequency loss

Target IR drop: $V_{IRPG} \sim 11\text{mV} @ 1.1\text{V}$
**POWER GATE CELL I-V CHARACTERISTICS**

- **FET operates in the linear region**
- **Can be approximated as a resistor**
- **Target region of operation**

![Graph showing the relationship between VDS and IDS with annotations for FET operation and target region.](image-url)
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

\[
V_{IRPG} = 2 \times \text{rows} \times I_{ROW} \times R_{PG}
\]

\[
V_{IR_{WC}} = V_{IRPG} + \sum V_{IRW}
\]
IR DROP EXAMPLE (UNIFORM POWER DENSITY)

IR Drop Along Virtual VSS
Perpendicular to Standard Cell Rows

Max current density = 1.0A/mm²
75.57ohm footer
EM EXAMPLE (UNIFORM POWER DENSITY)

EM Margin Along Virtual VSS
Perpendicular to Standard Cell Rows

Max current density = 1.0A/mm²

Thick metal will provide more EM margin

1X Metal EM Limit
1.3X Metal EM Limit

Standard Cell Rows Between Footers

EM Limit
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 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

STRAIGHT EDGE & ZIG/ZAG EDGE PACKAGE PLANE BOUNDARIES

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
In-rush current can cause supply bounce on neighboring IP

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
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

Large in-rush setting
Transient droop limited by CC6 exit

Small in-rush setting
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:

- 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