



Power Management in an Embedded Multiprocessor Cluster

Matthias Knoth, MIPS Technologies
Embedded World Conference
March 4, 2009

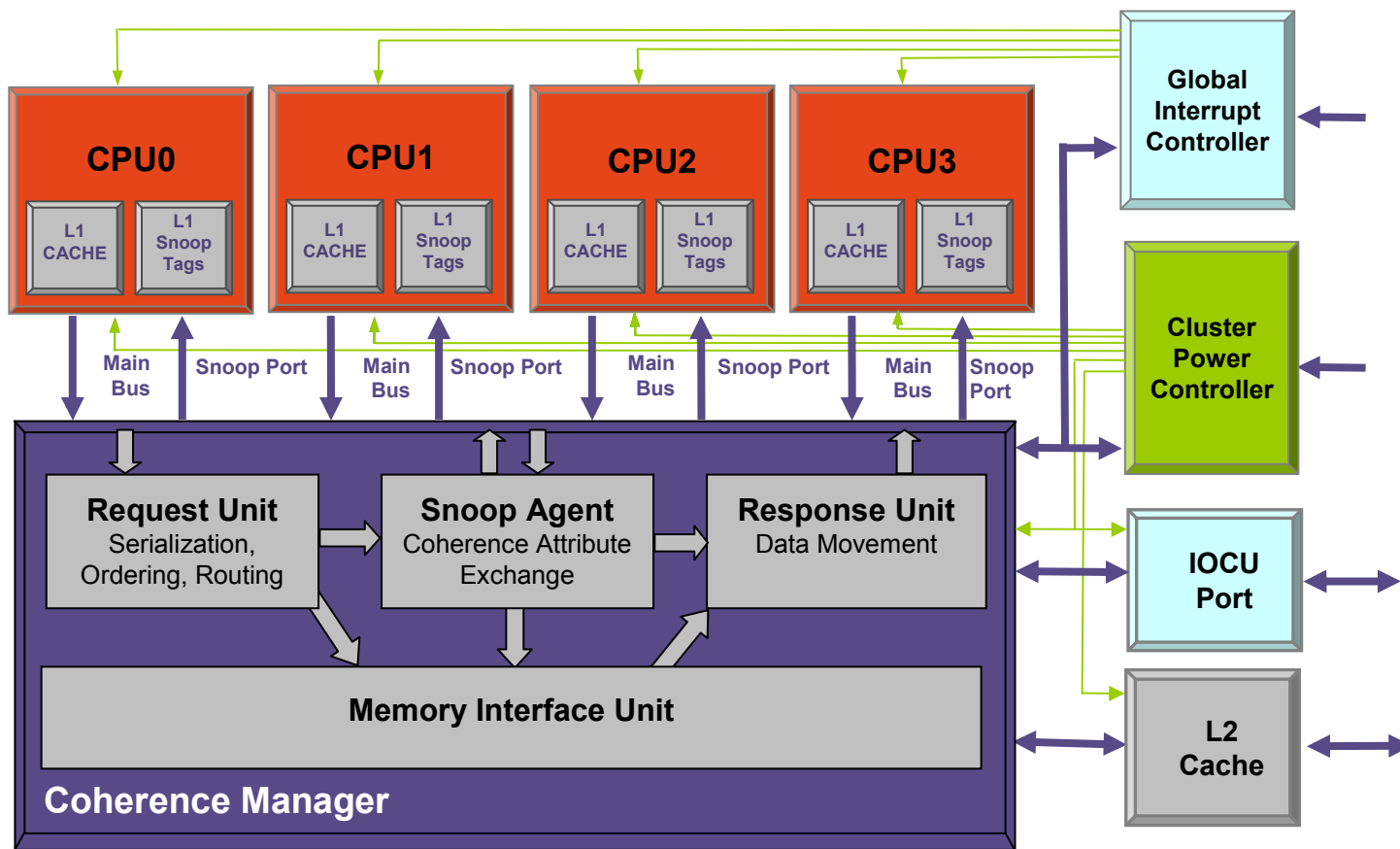
Motivation and Agenda

- ❖ Multiprocessor clusters have migrated into embedded space
- ❖ SoC energy consumption and thermal envelopes are tight
- ❖ Computation platform shares power budgets with rest of SoC
- ❖ Execution parallelism varies between tasks
- ❖ Performance demands often fluctuate temporally
 - Cluster frequency requirements fluctuate
 - Individual processor utilization fluctuates
 - ➔ Great opportunity for power management

- ❖ Agenda: cluster overview, power management in HW and SW, implementation aspects

MIPS32® 1004K™ Coherent Processing System

Snoop-based memory coherent processing cluster; hosting 2-4 CPUs, each with 2 virtual processor environments and threads

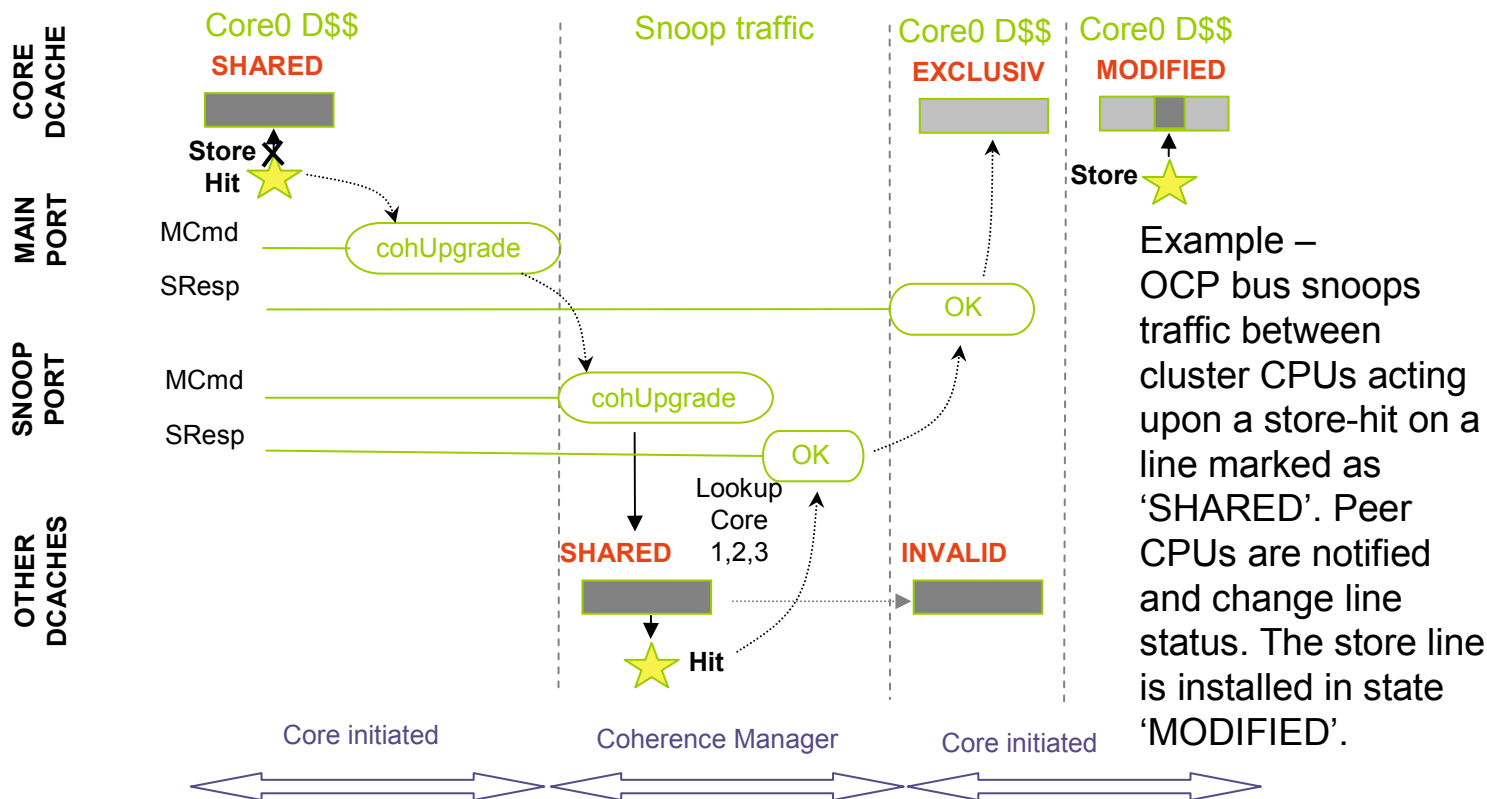


MIPS32® 1004K™ Coherent Cluster - Details

- ❖ Foundry independent, configurable, fully synthesizable IP
- ❖ 800 MHz WC @ 65nm GP TSMC operating frequency
- ❖ 2-4 CPUs multi-threaded with coherent and non-coherent execution, each with 8KB - 64KB L1 cache
- ❖ L2 cache 128KB – 1MB, 4-8 way set associative
- ❖ Snoop-based memory coherence facilitated by central coherence manager
- ❖ Programmable I/O HW coherency unit for data movement
- ❖ Programmable Interrupt Controller for up to 256 interrupts
- ❖ Programmable Power Controller to manage domain states

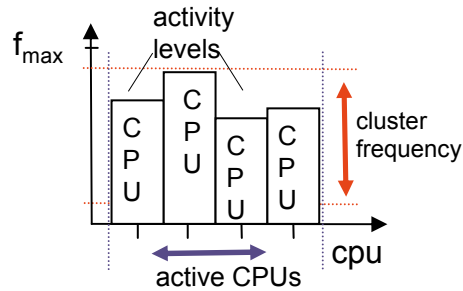
MESI Style, Snoopy Memory Coherence

- Problem: localized L1 caches will contain locally modified data
- L1 holds coherence attributes: Modified, Exclusive, Shared, Invalid
- Duplicate tags permit parallel LD/ST and snoop operation



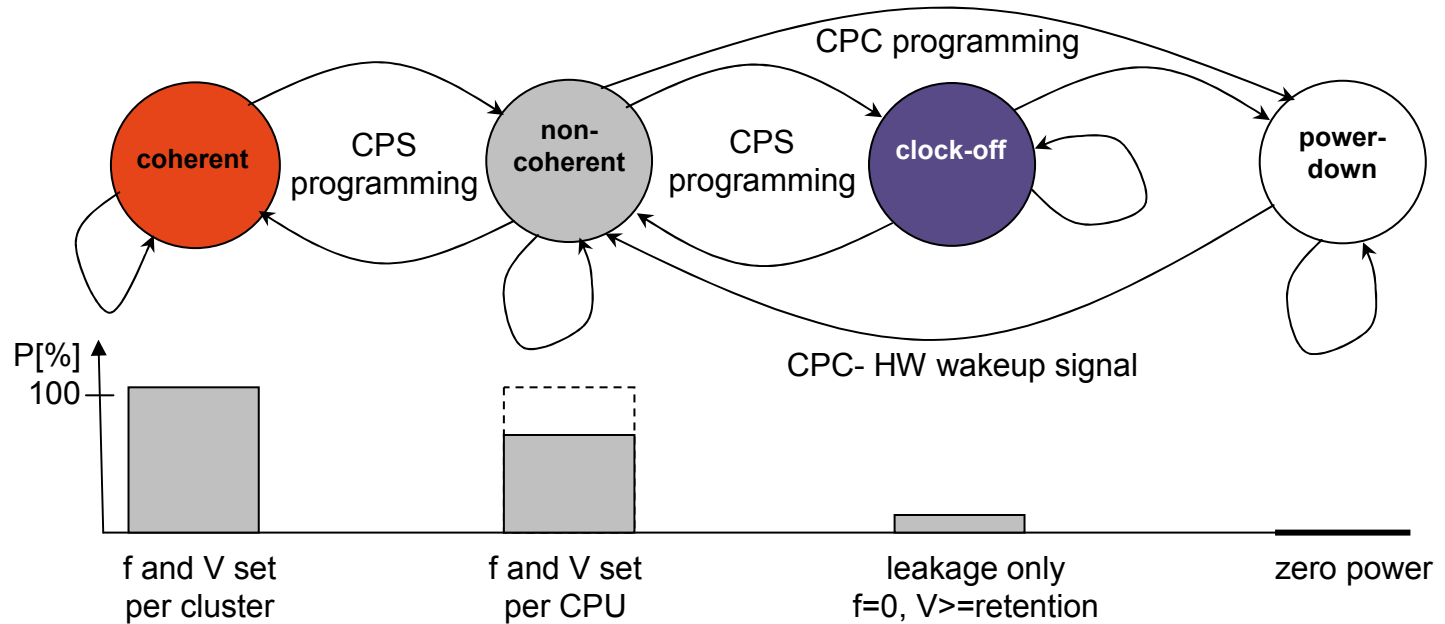
Cluster Power Management Hardware

- Power Controls: active CPUs, cluster frequency, cluster voltage



- Cluster frequency change on demand
- Power-down of unused CPUs
- Cluster voltage follows frequency requirements

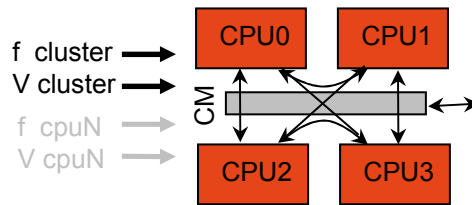
- Processor states determine power regime



Cluster and CPU Power Modes

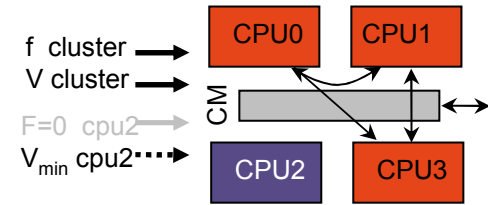
- CPUs in coherent or non-coherent operation – controlled by the coherence manager
- Coherent CPUs operate at same speed to support snooping
- Non-coherent CPU's frequency can be varied independently
- Clock-Off mode retains CPU state and removes dynamic power
- Power-Down mode removes leakage power

- Snoop traffic between CPUs required
- Frequencies in lockstep



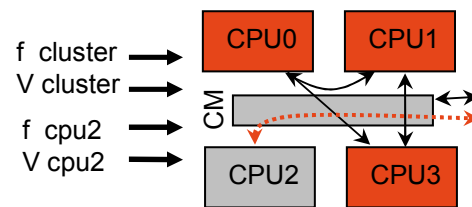
Coherent Mode

- CPU is inactive
- CPU state frozen
- Leakage current



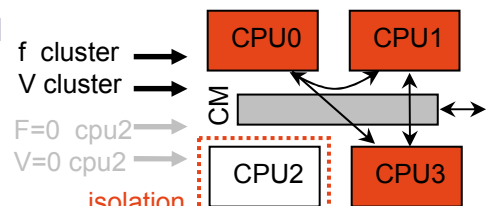
Clock-Off Mode

- CPU acts independently
- Frequency can be modified



Non-Coherent Mode

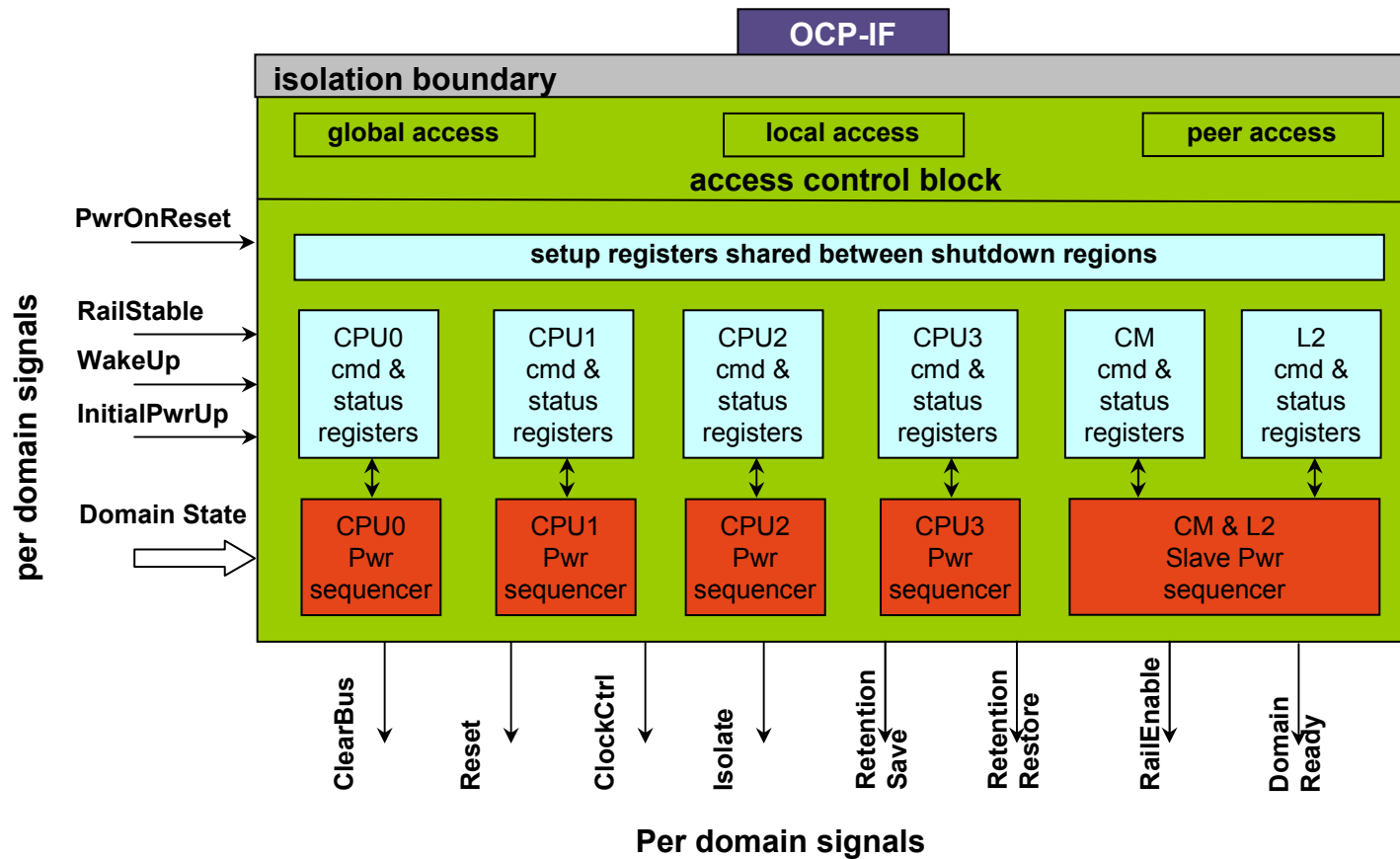
- CPU isolated
- Power-off
- No leakage
- State retention if required



Power-Down Mode

The Cluster Power Controller

- Provides cluster wide power gating, clock and reset control
- Programmable platform peripheral; accessible through cluster members
- Wake-up scheduling through external events or peer CPU access



Power Management - Software Aspects

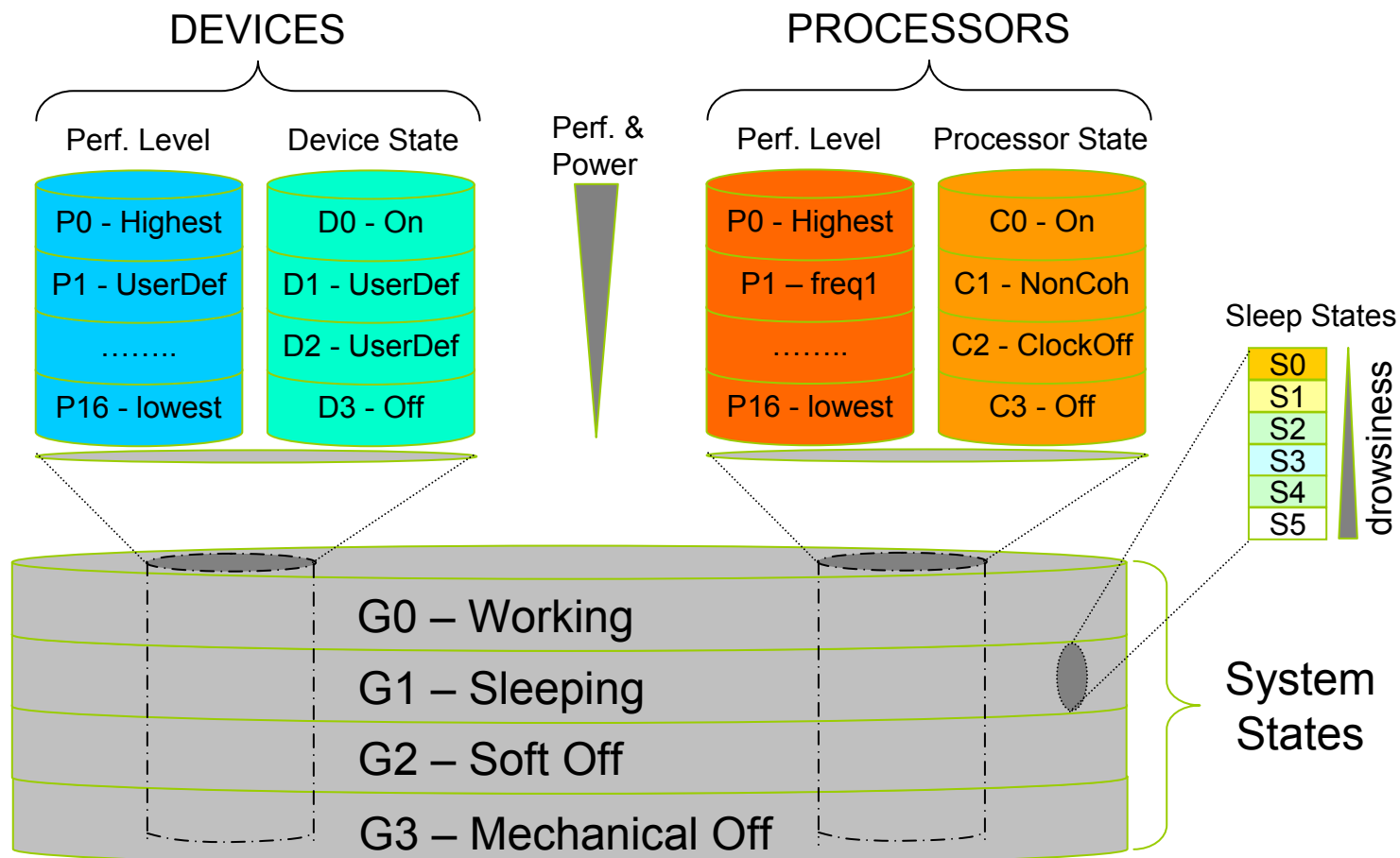
- ❖ **Power-aware OS responds to immediate demand and collects statistics to predict performance requirements**
- ❖ **Performance levels and processor state are control targets**
- ❖ **OS assigns threads and moves them between CPUs**

- ❖ **Performance levels:**
 - Primary process variable – determined by performance statistics
 - Controls cluster frequency and HW thread allocation
- ❖ **Processor states:**
 - Secondary process variable – derived from performance level
 - States: Coherent, Non-Coherent, Clock-Off, Powered-Down

- ❖ **Unused CPUs will be shut down by the OS**
- ❖ **Idle or sleeping CPUs can power-down and wake-up autonomously – programmable via cluster power controller**
- ❖ **OS modulates cluster frequency according to demand**

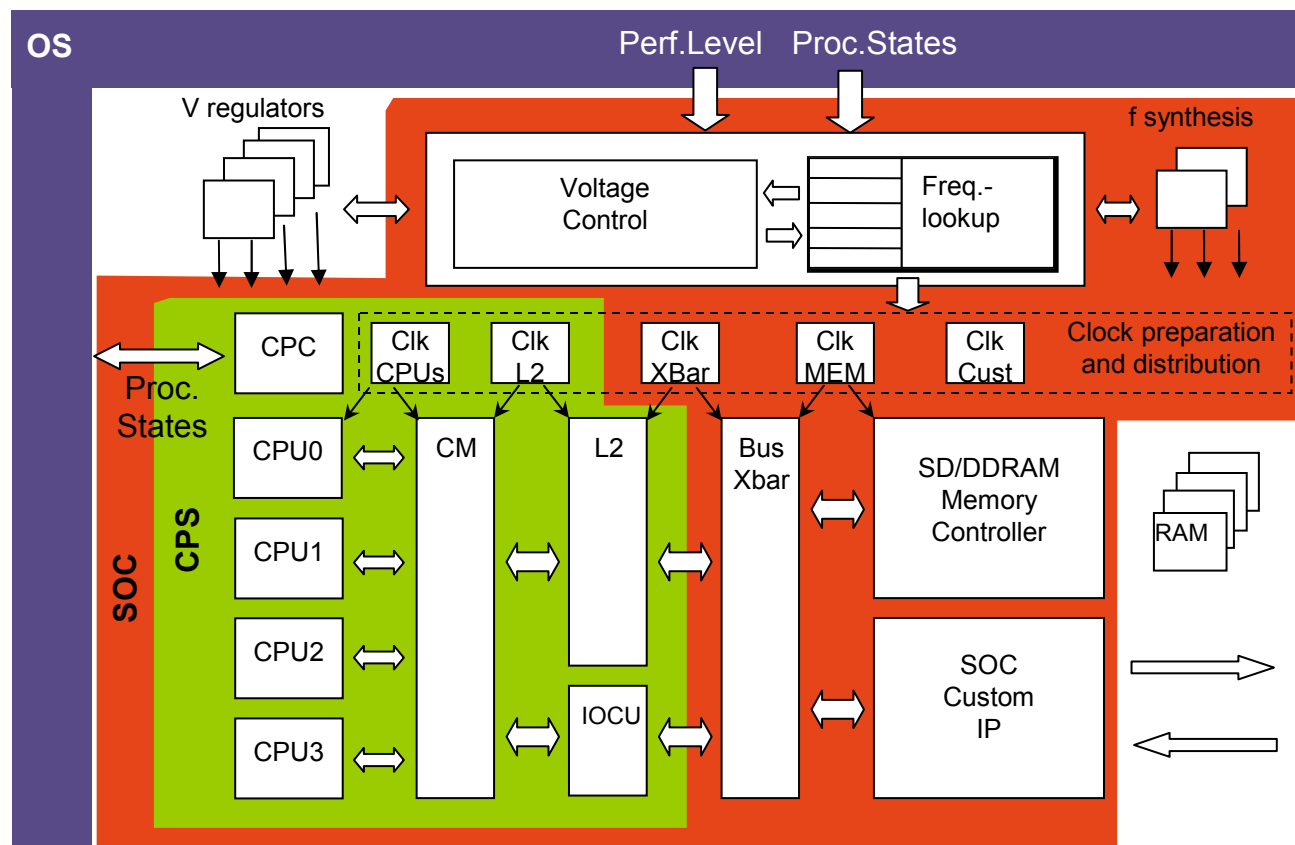
Example: ACPI Power Management

- Defines 4 system, 4 device, 4 processor states
- Defines up to 16 performance levels for each class



Performance Level driven V/f Control

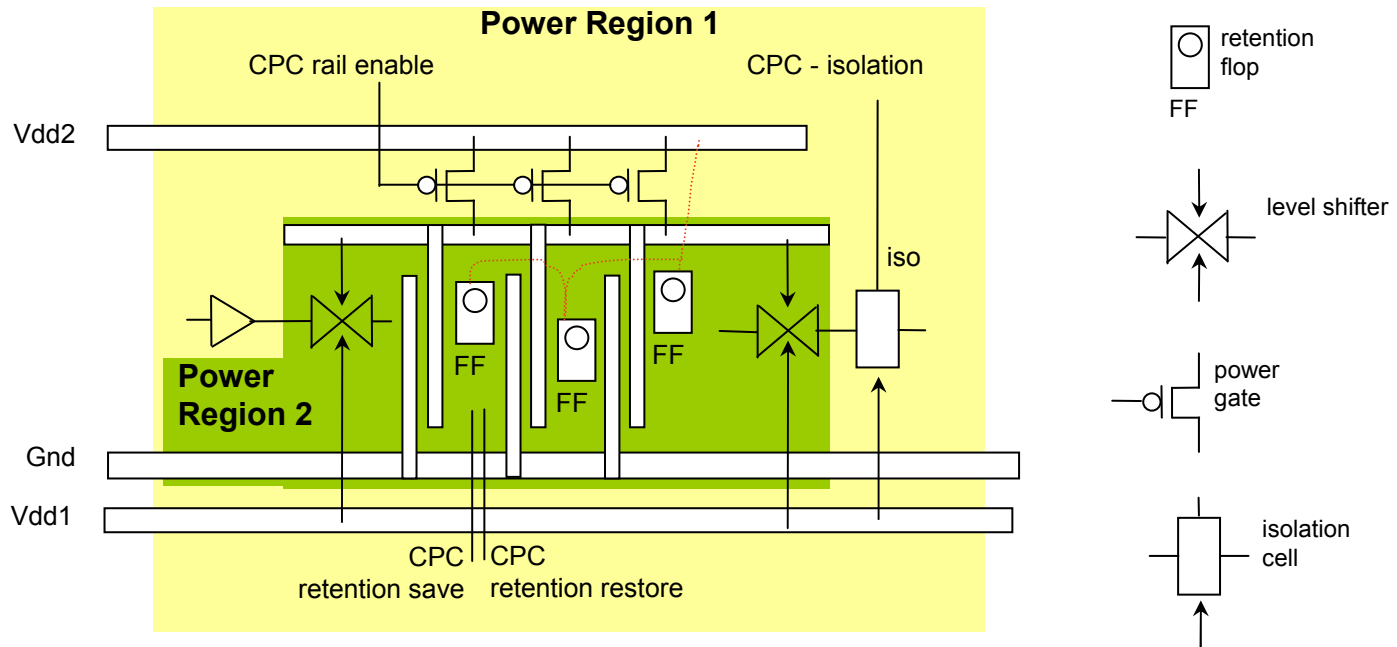
- Power-aware OS sets desired performance level and processor state
- CPC and V/F-control execute on target settings



OS performance level driven frequency and voltage control in an SoC

Silicon Implementation Aspects

- Power gating silicon methodology has become mainstream
- Supported by silicon library vendors and EDA tool vendors
 - Power gate cells disconnect power domain supply or ground rails
 - Isolation cells ensure proper drive from un-powered to powered domains
 - Level shifting cells translate logic levels between domains
 - Retention flops retain state for power-down domains



Power management cell application in a power gated domain

End



MIPS, MIPS32, MIPS64, MIPS-Based, MIPS-Verified, and the MIPS Technologies Logo are trademarks of MIPS Technologies, Inc. and registered in the U.S. Patent and Trademark Office. MIPS, MIPS32, MIPS64, MIPS-Based, MIPS Logo, MIPS Technologies Logo, CorExtend, Pro Series, M4K, 4KE, 4KEc, 24K, 24KE, 34K, 74K, 1004K and FS2 are trademarks or registered trademarks of MIPS Technologies, Inc. in the United States and other countries.