Power Management

Outline

- **Power Management Mechanisms**
- Server Power Management

Power and Energy

Power (Watts) = Energy (Joules) / Time (sec)

- Power is limited by infrastructure (e.g., power supply)

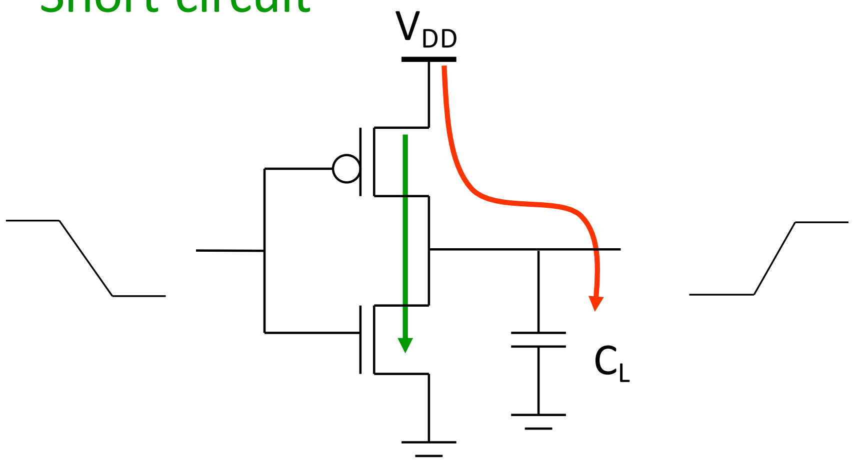
- Energy: what the utilities charge for or battery can store

CMOS Power Consumption

$$P_{total} = P_{dyn} + P_{stat} = P_{tran} + P_{sc} + P_{lkg}$$

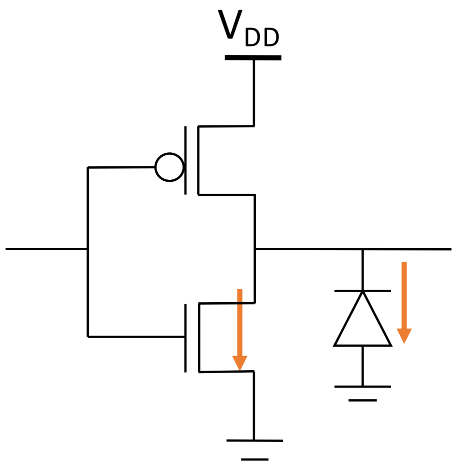
Dynamic Power

- Signal transitions
  - Logic activity
  - Glitches
- Short-circuit



Static Power

- Leakage



Dynamic Power Consumption

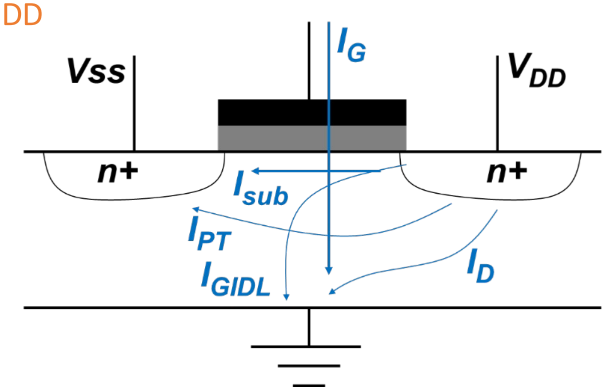
- Dynamic power: $P_{dyn} = \alpha \cdot C_L \cdot f_{clk} \cdot V_{DD}^2$
  - α activity factor (i.e., the probability the given node will change its state from 1 to 0 or vice versa at a given clock tick)
  - C_L total load capacitance
  - f_{clk} clock frequency
  - V_{DD} supply voltage

- Circuit techniques to reduce dynamic power
  - State/bus encoding (↓α)
  - Reduce device size (↓C_L)
  - Pipelining & parallelism (↓ f_{clk}, ↓ V_{DD})

- Run-time techniques to reduce dynamic power
  - Clock-gating (↓α)
  - Dynamic Voltage & Frequency Scaling - DVFS (↓ f_{clk}, ↓ V_{DD})

Leakage Power Consumption

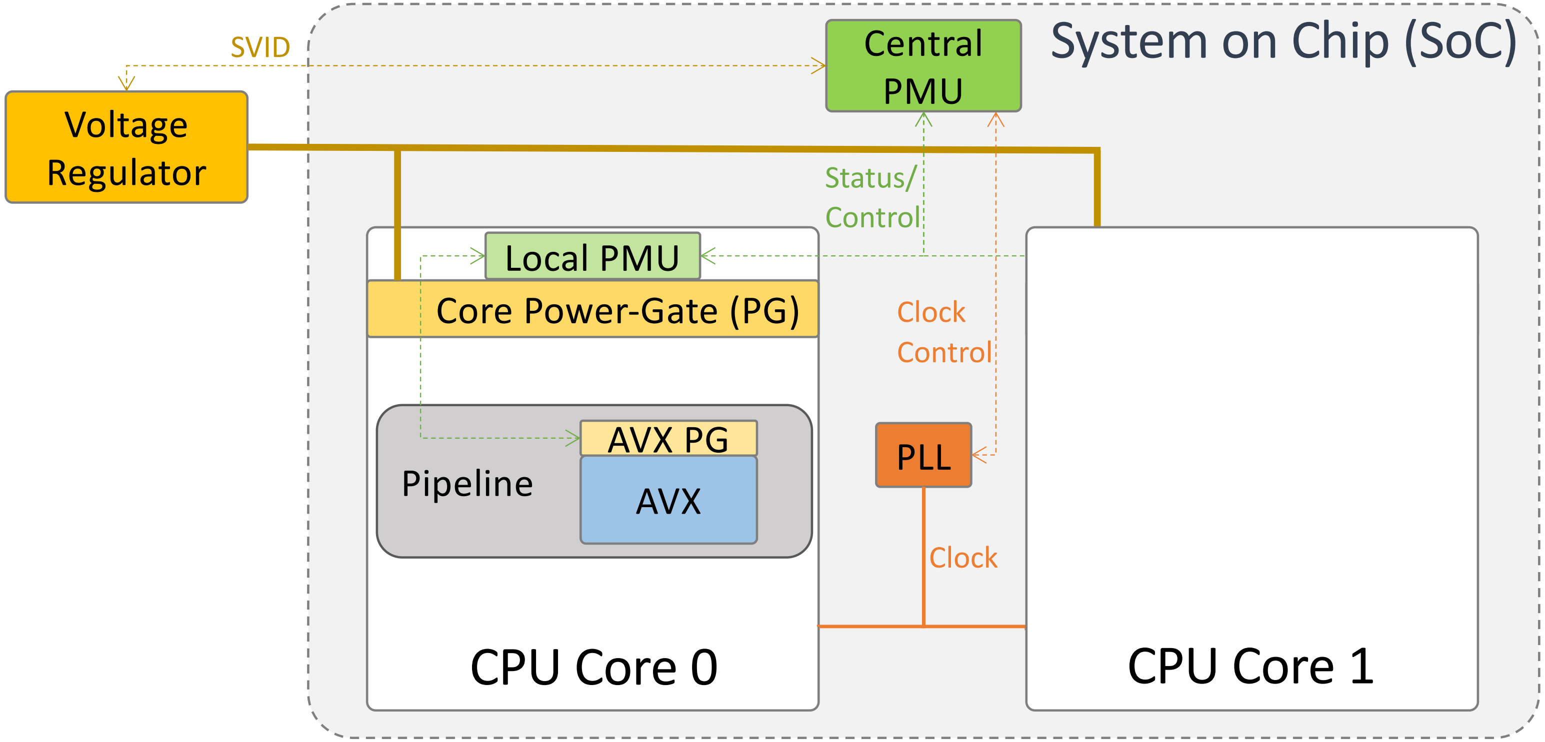
- Static power: $P_{static} = I_{stat} \cdot V_{DD} = (I_{sub} + I_D + I_{GIDL} + I_{PT} + I_G) \cdot V_{DD}$
  - I_{sub} Subthreshold leakage
  - I_D Junction Reverse Bias Current
  - I_{GIDL} Gate Induced Drain Leakage
  - I_{PT} Punch-through Current
  - I_G Gate Tunneling Currents
  - V_{DD} Supply voltage



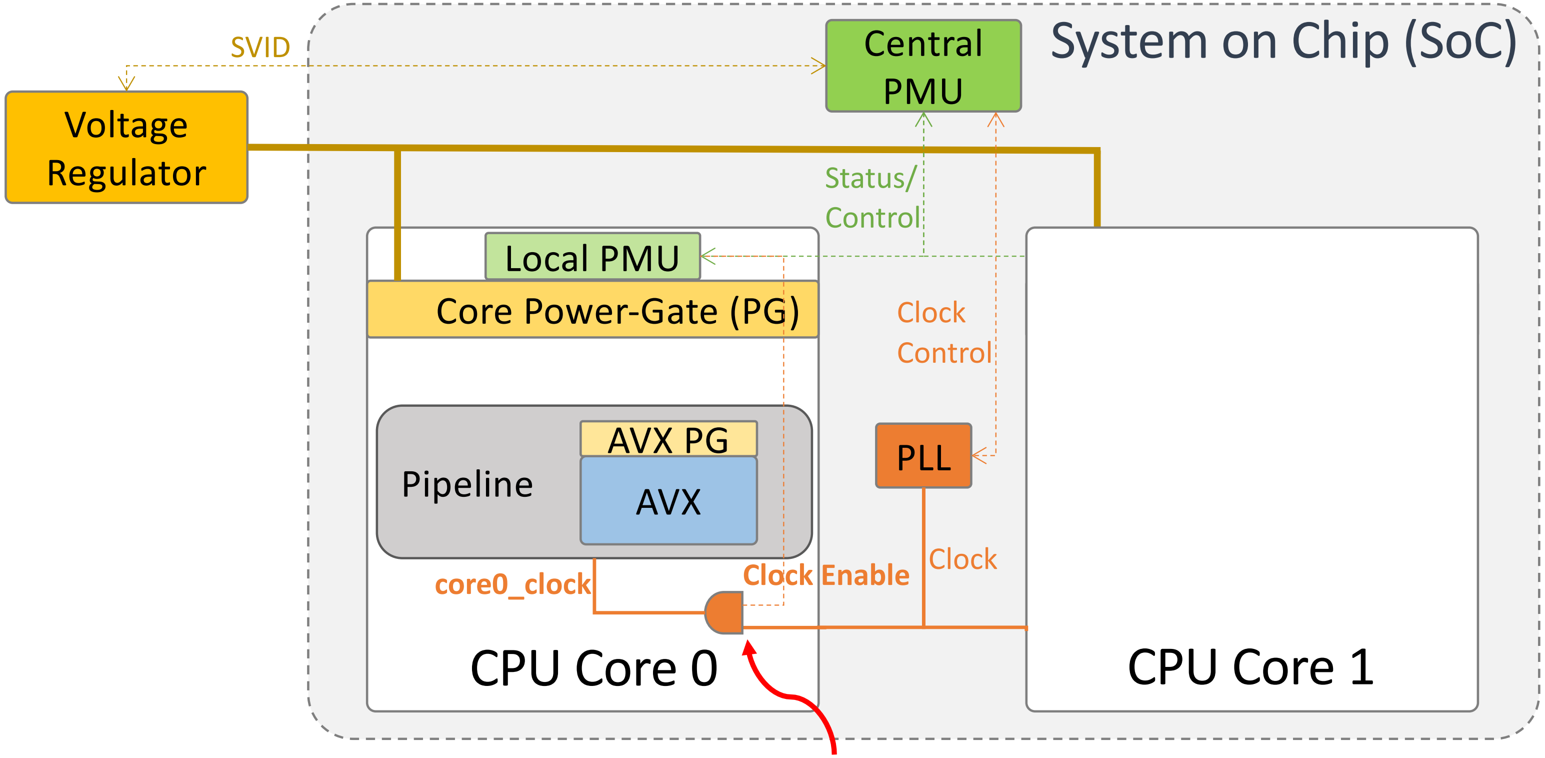
- Circuit techniques to reduce leakage power
  - Increase V_t: use Multiple-threshold (V_t) devices (↓ I_{stat})
    - Use low V_t devices (have high leakage) only in critical circuits

- Run-time techniques to reduce leakage power
  - Reduce idle circuit's voltage to retention (↓ V_{DD})
  - Power-gate idle circuit (↓ V_{DD})
  - Dynamic Voltage & Frequency Scaling - DVFS (↓ V_{DD})

PM Architecture Overview

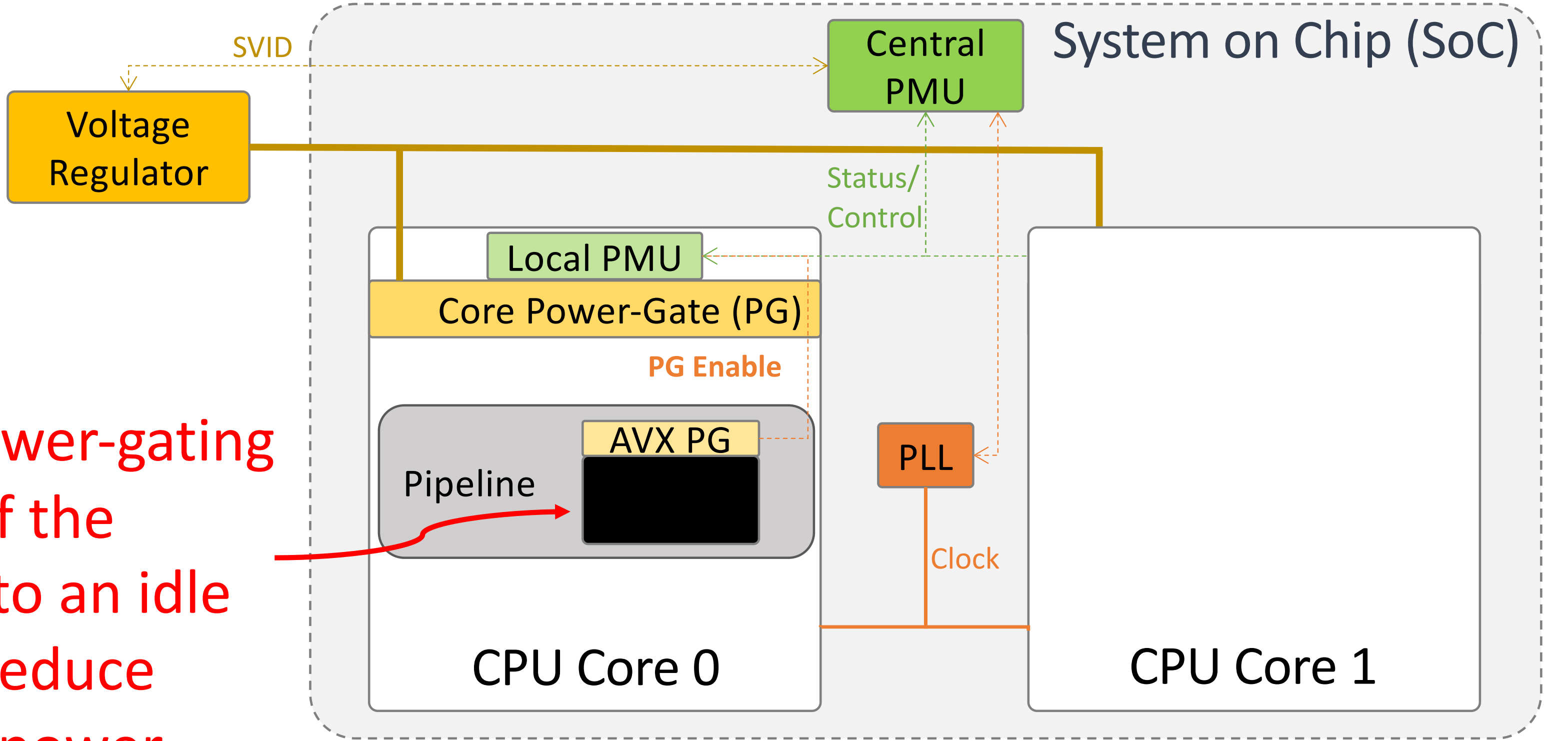


Clock-gating



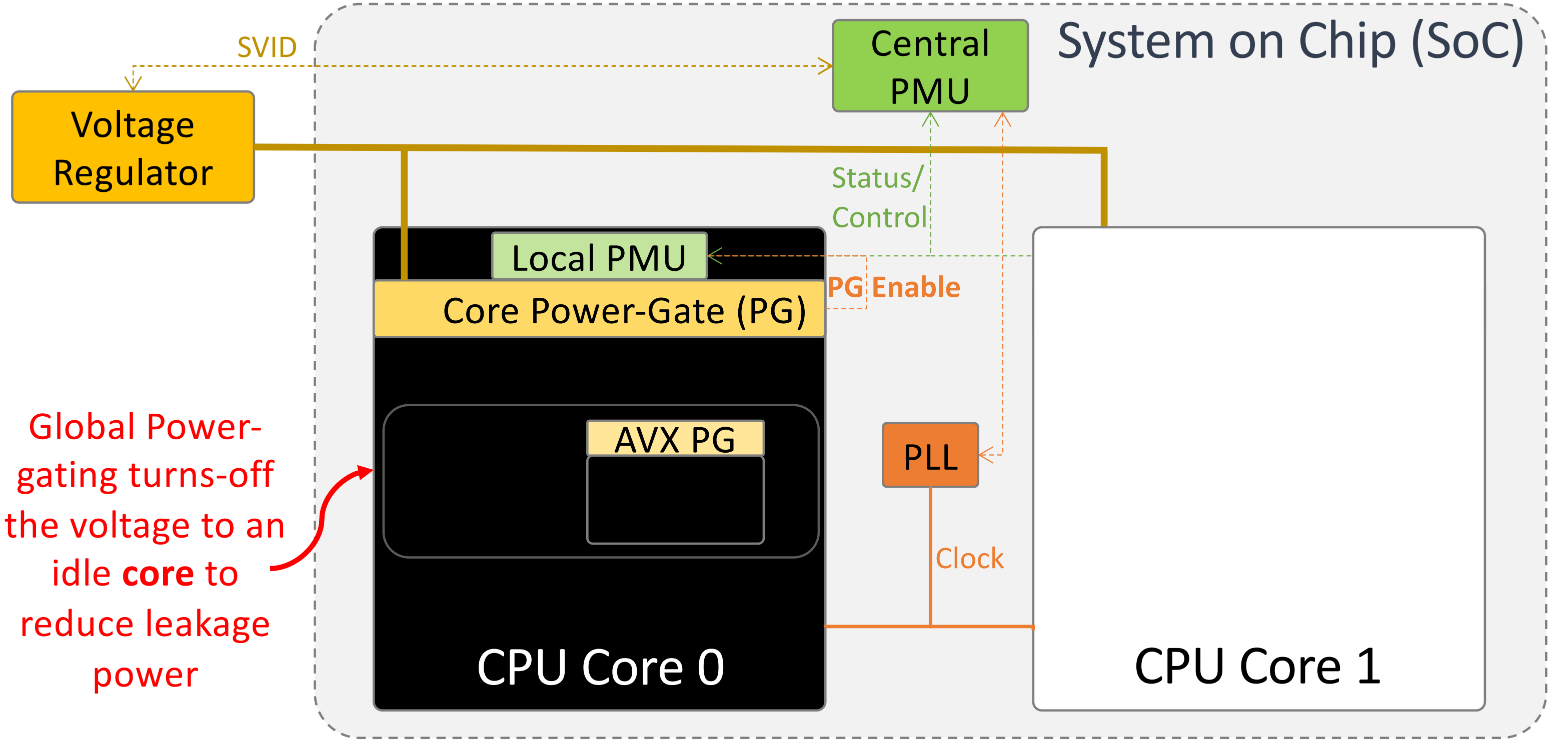
Clock-gating stops the clock to an idle core/unit to reduce dynamic power

Local Power-gating



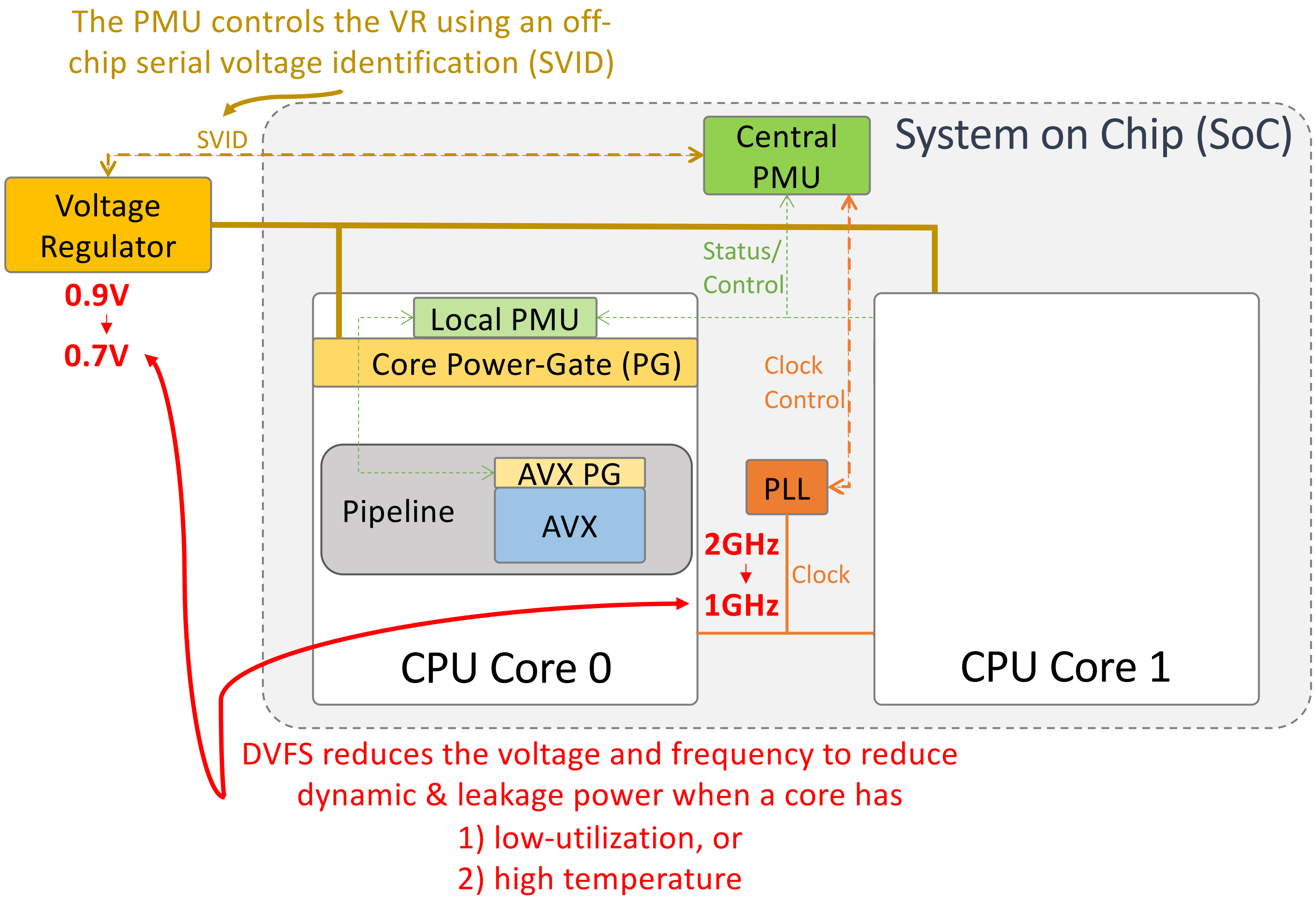
Local power-gating turns-off the voltage to an idle **unit** to reduce leakage power

Global Power-gating

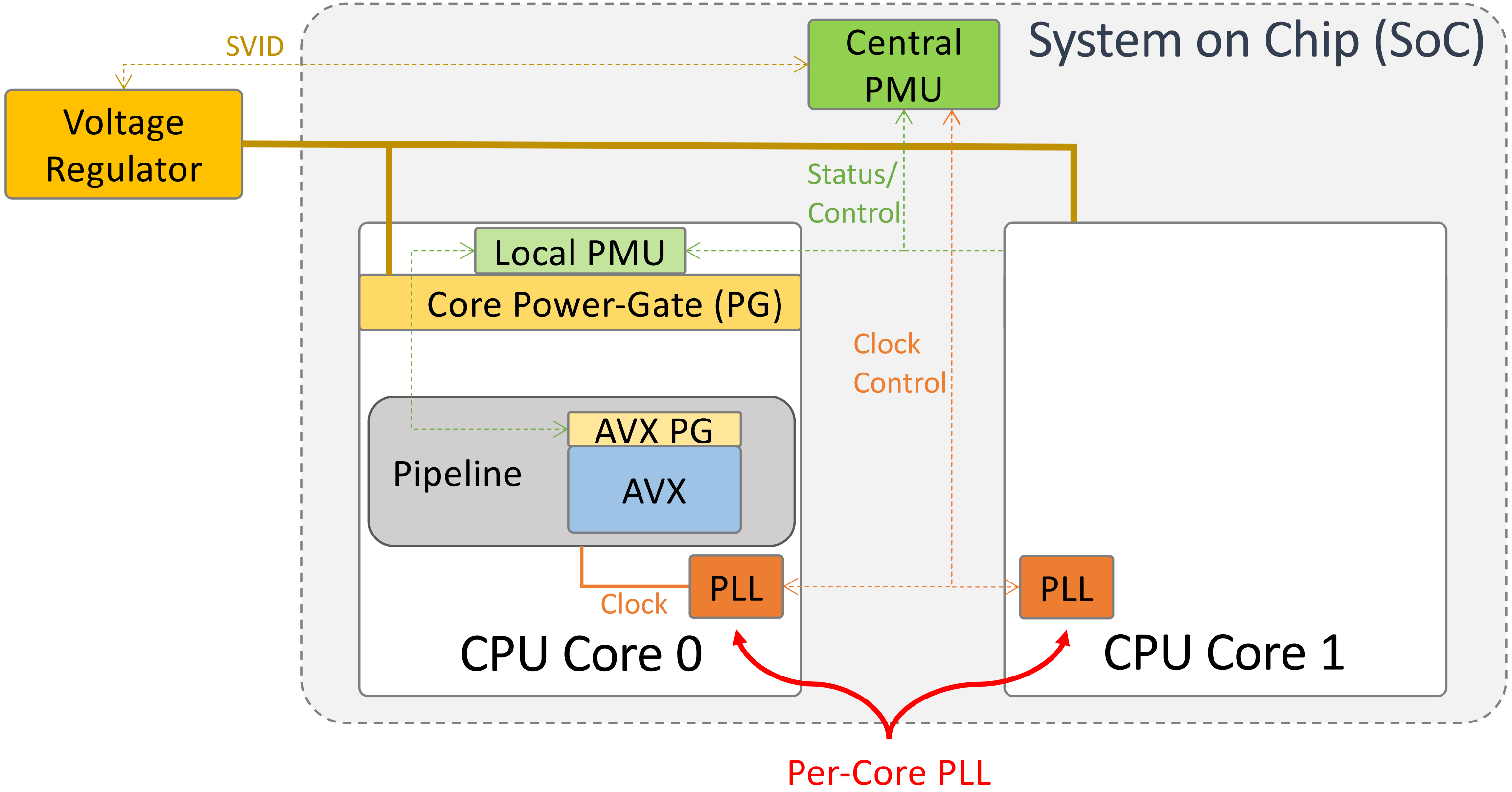


Global Power-gating turns-off the voltage to an idle **core** to reduce leakage power

DVFS

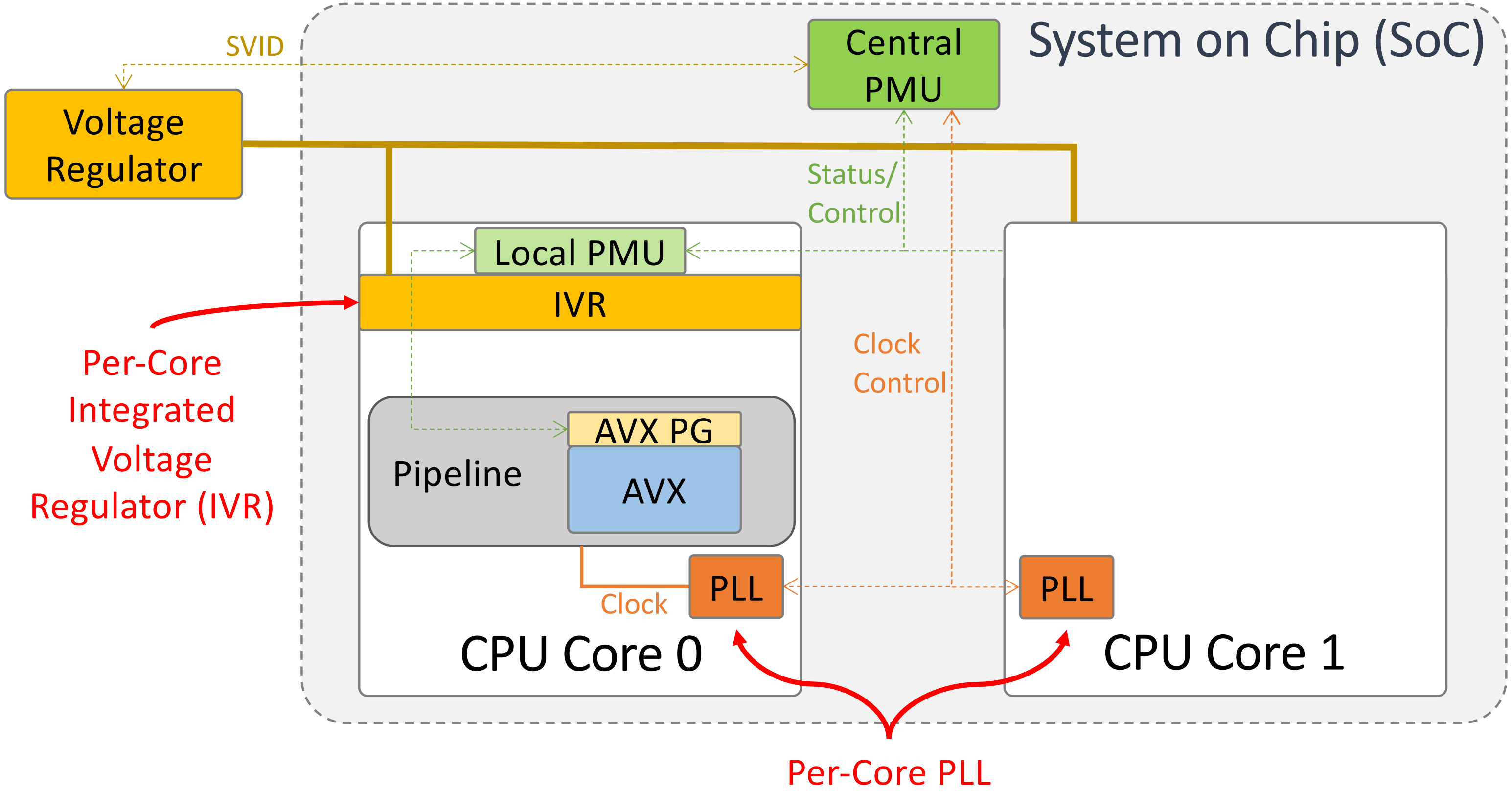

The PMU controls the VR using an off-chip serial voltage identification (SVID)
SVID
Central PMU
System on Chip (SoC)
Voltage Regulator
0.9V
0.7V
Status/ Control
Local PMU
Core Power-Gate (PG)
Clock Control
AVX PG
Pipeline
AVX
PLL
2GHz
1GHz
Clock
CPU Core 0
CPU Core 1
DVFS reduces the voltage and frequency to reduce dynamic & leakage power when a core has
1) low-utilization, or
2) high temperature

Advanced PM Architecture (I)



Per-Core PLL enable *per-cores Dynamic Frequency Scaling (DFS)*

Advanced PM Architecture (II)



Per-Core IVR and PLL enable *per-core DVFS*

More Advanced PM Features

- There are more advanced PM features:
  - Power budget management
  - Computational sprinting (e.g., Turbo)
  - Maximum current limit protection
  - Maximum voltage limit protection
  - Voltage emergency prevention & avoidance
  - Adaptive voltage scaling
  - Reliability degradation mitigation
  - System level idle power-states
  - System level DVFS
  - Race to halt
  - Hardware duty cycling
  - …

Outline

- Power Management Mechanisms
- **Server Power Management**

Server Power Management

- Core Idle States
- Package Idle States
- DRAM Idle States
- IO Link States

Core Idle Power State – Core C-states

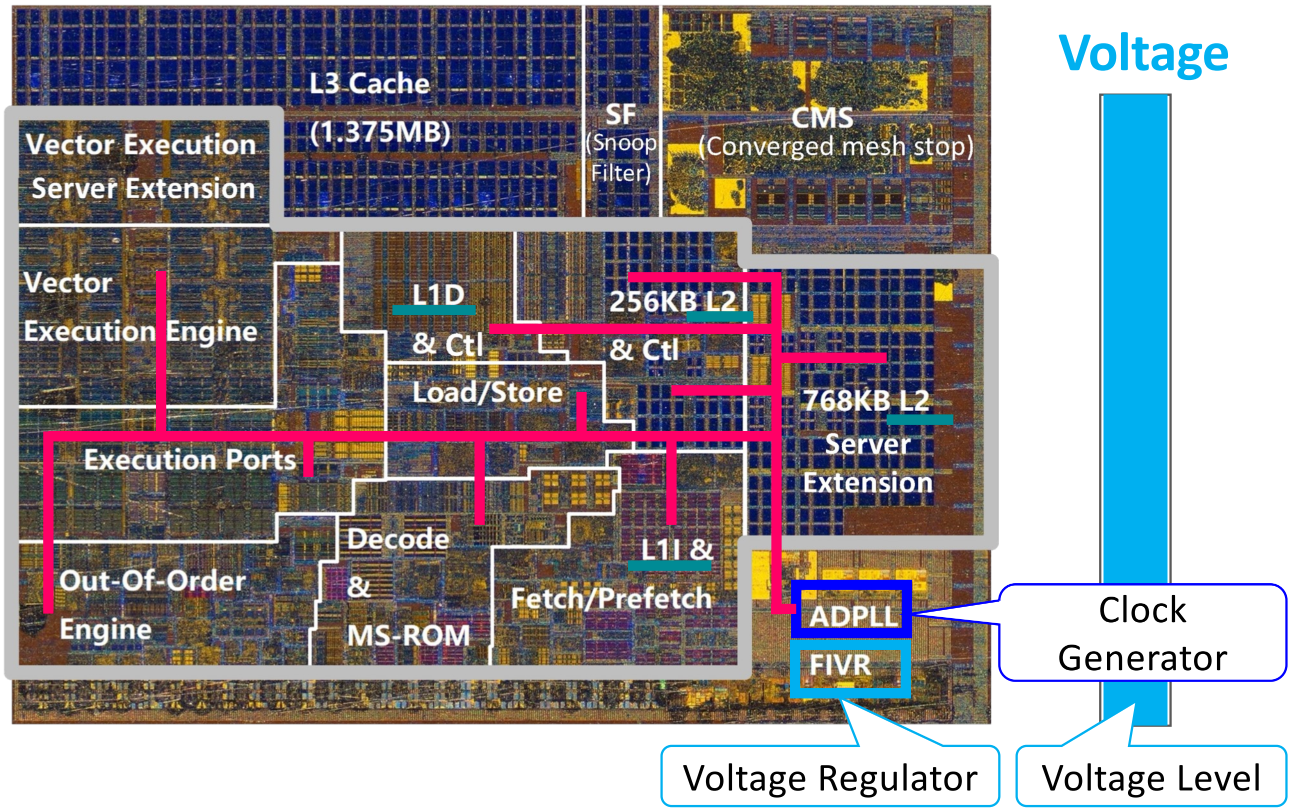
- Core C-states are power saving states enable the core to reduce its power consumption during idle periods

- Intel's Skylake architecture offers four main Core C-state:

Core State	Sleep Level	Power per core	Transition Latency
C0	Active	4W	--
C1	Shallow	1.4W	2µs
C1E	Medium	0.9W	10µs
C6	Deep	0.1W	133µs

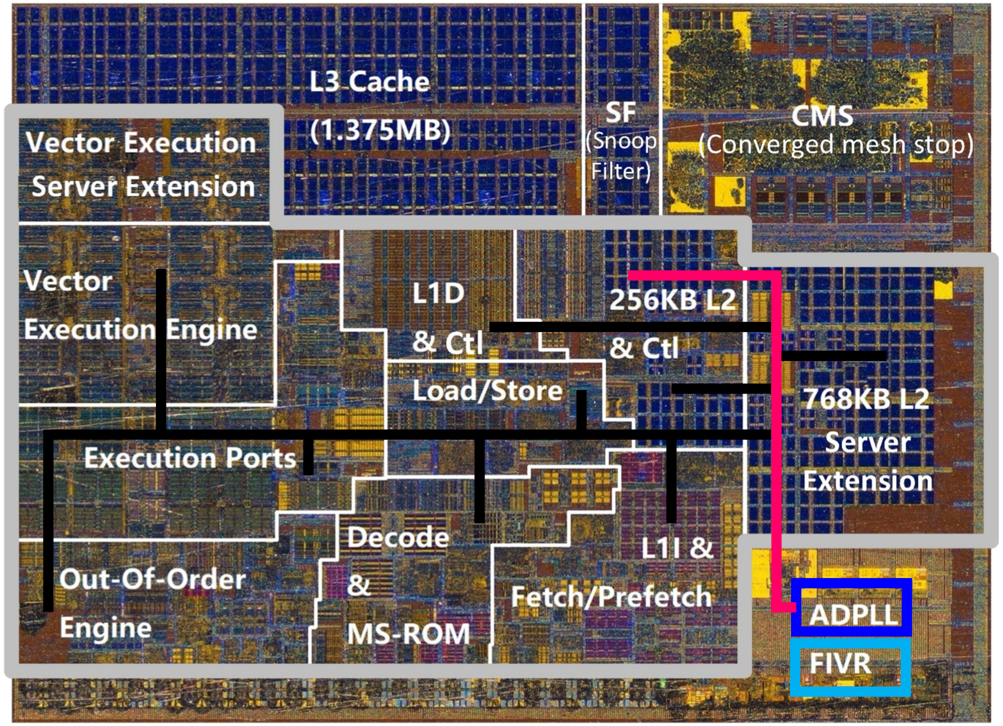
Transition Latency: Time to switch from an active to an idle state (and back)

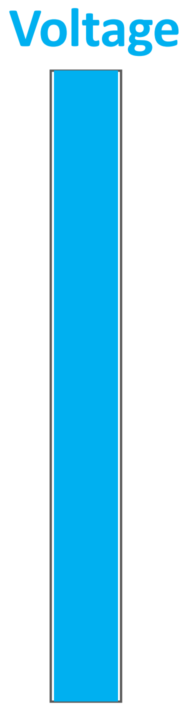
C0 (Active) Core C-state



C-State	Clocks	ADPLL	L1/L2 Cache	Voltage	Context
C0	Running	On	Coherent	Nominal	Maintained

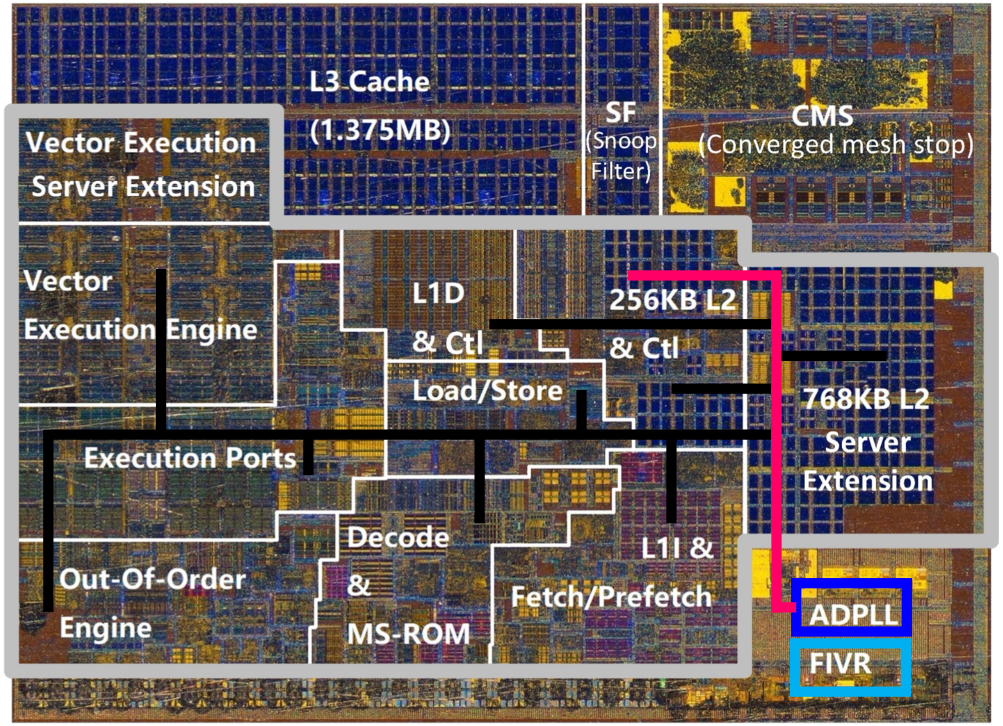
C1 (Shallow) Core C-states

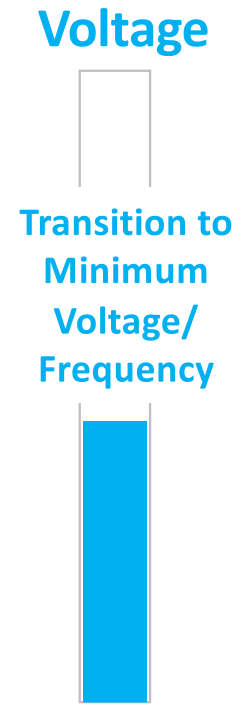




C-State	Clocks	ADPLL	L1/L2 Cache	Voltage	Context
C1	Most Stopped	On	Coherent	Nominal	Maintained

C1E (Medium) Core C-state

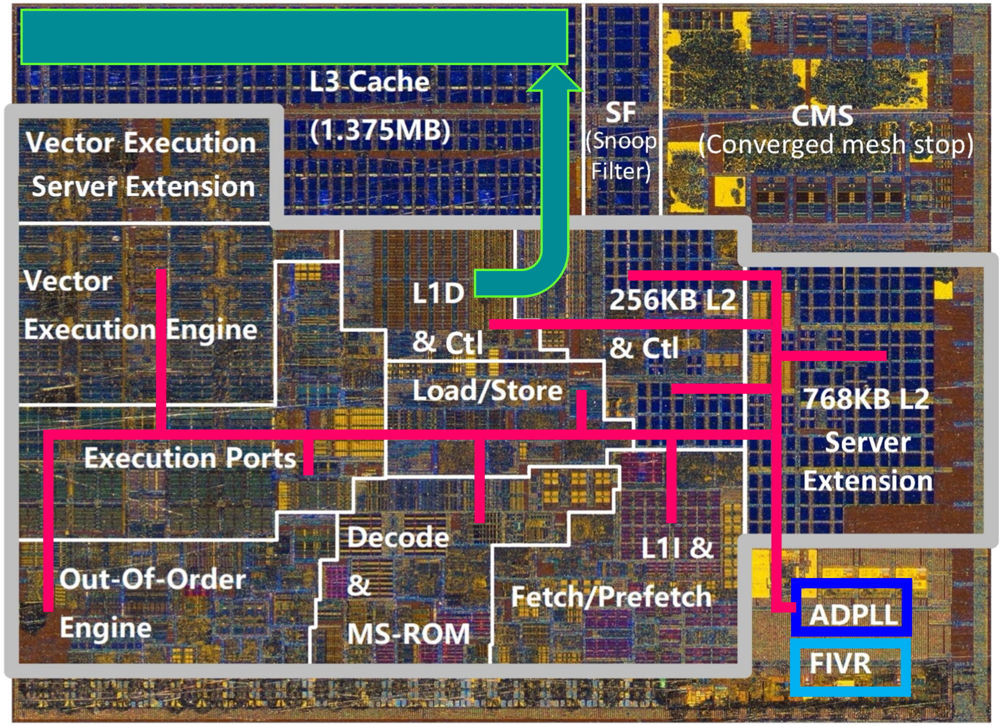




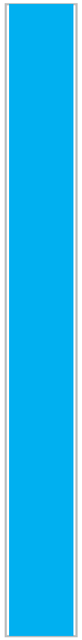
C-State	Clocks	ADPLL	L1/L2 Cache	Voltage	Context
C1E	Most Stopped	On	Coherent	Min V/F	Maintained

C6 (Deep) Core C-state

Flush L1/L2 Caches



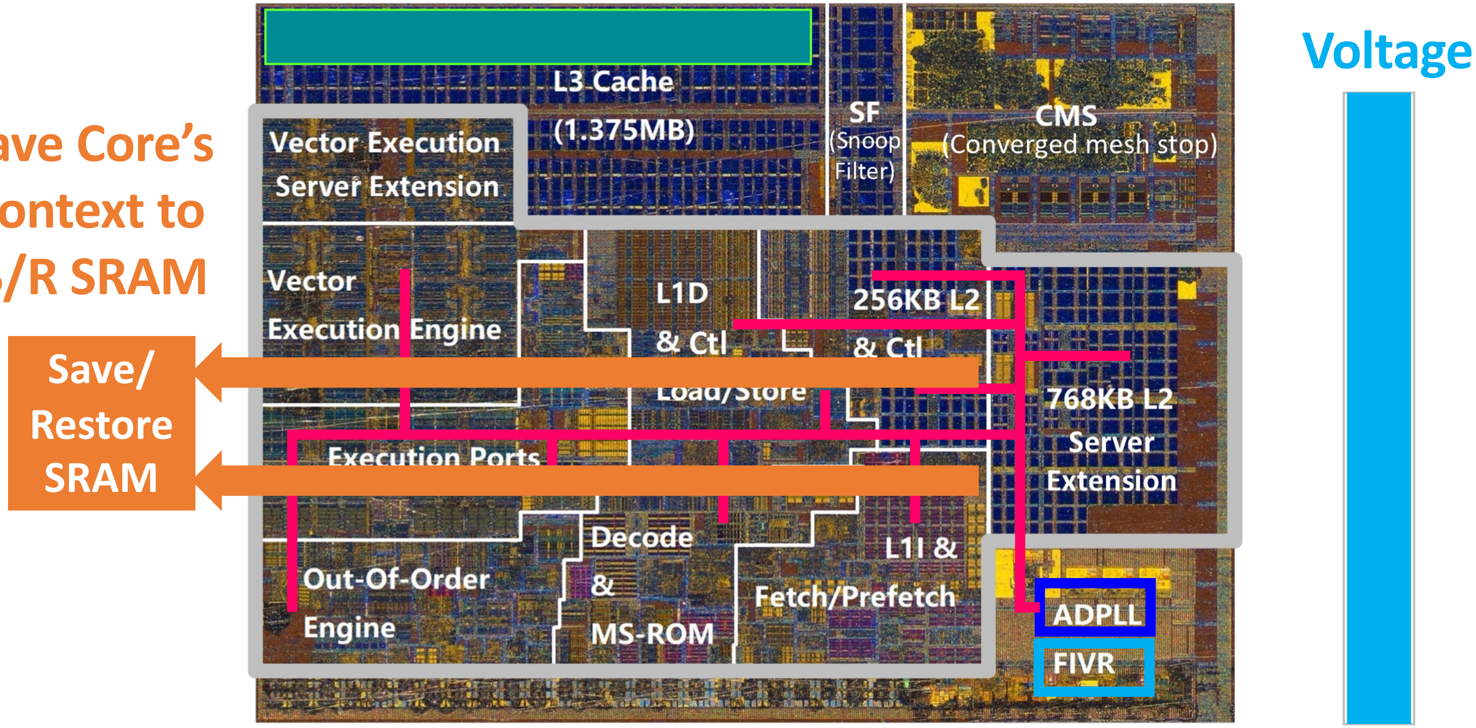
Voltage



C-State	Clocks	ADPLL	L1/L2 Cache	Voltage	Context
C6	Running	On	Flushed	Nominal	Maintained

C6 (Deep) Core C-state

Save Core's Context to S/R SRAM

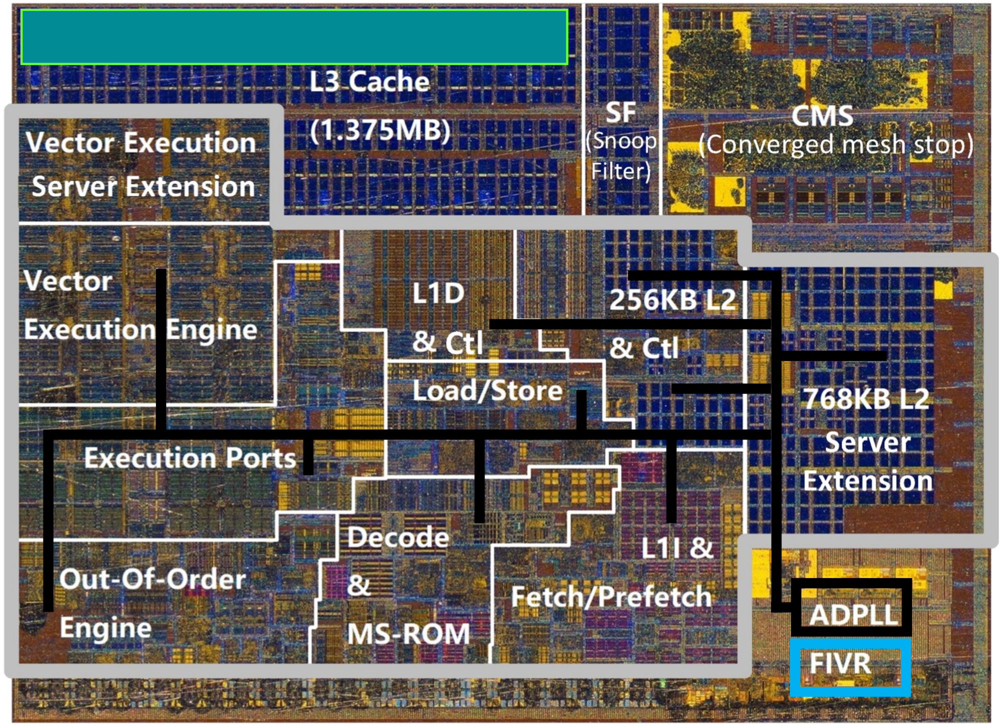


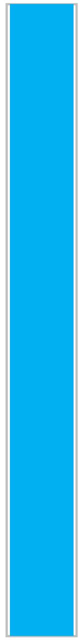
C-State	Clocks	ADPLL	L1/L2 Cache	Voltage	Context
C6	Running	On	Flushed	Nominal	S/R SRAM

C6 (Deep) Core C-state

Turn-off the **clocks** and **PLL**

Save/ Restore SRAM



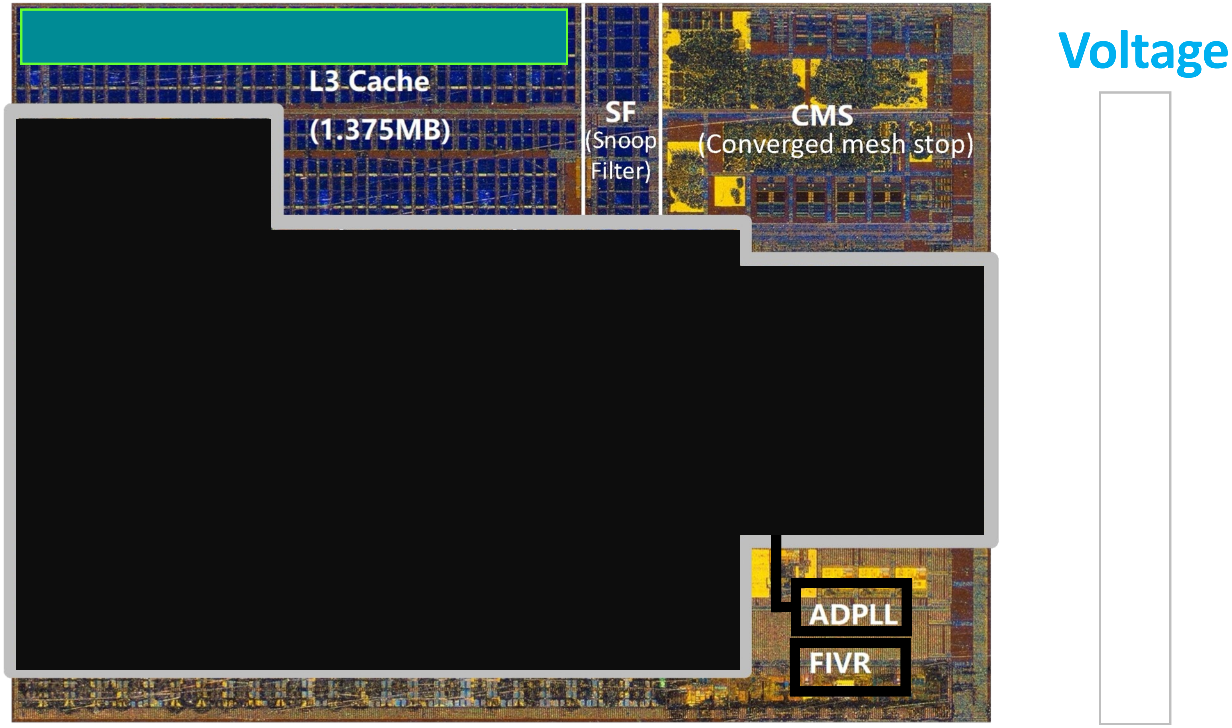


C-State	Clocks	ADPLL	L1/L2 Cache	Voltage	Context
C6	Stopped	off	Flushed	Nominal	S/R SRAM

C6 (Deep) Core C-state

Turn-off the voltage

Save/ Restore SRAM



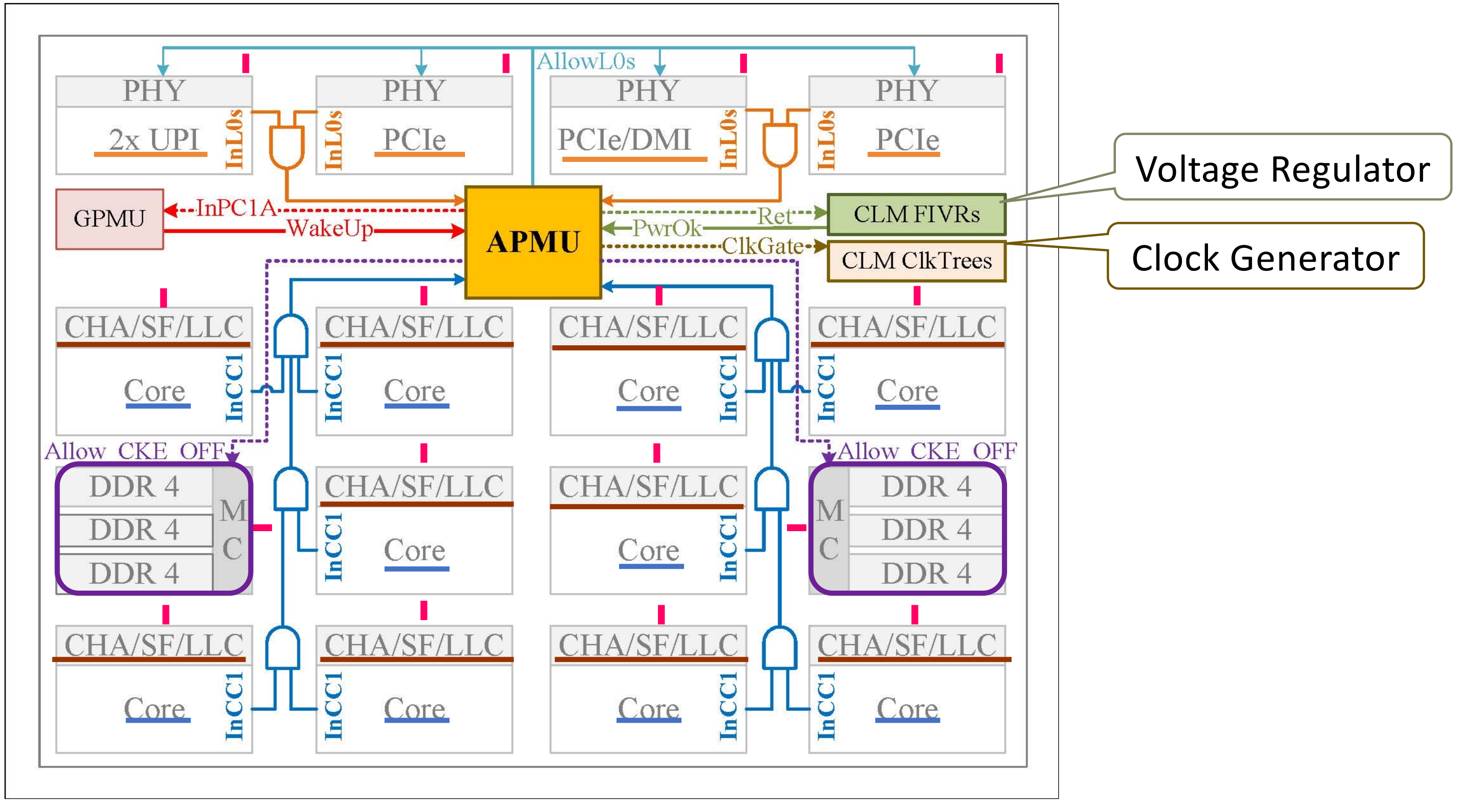
C-State	Clocks	ADPLL	L1/L2 Cache	Voltage	Context
C6	Stopped	off	Flushed	Shut-off	S/R SRAM

Package C-states

- Package C-states are power saving states that enable the uncore and DRAM to reduce their power consumption during idle periods
- For a system to enter Package C-states, the cores and IO links should be idle

- Intel's Skylake architecture offers three Package C-states:
  - PC0 - Active
  - PC2 - Intermediate (non-architectural)
  - PC6 - Deep

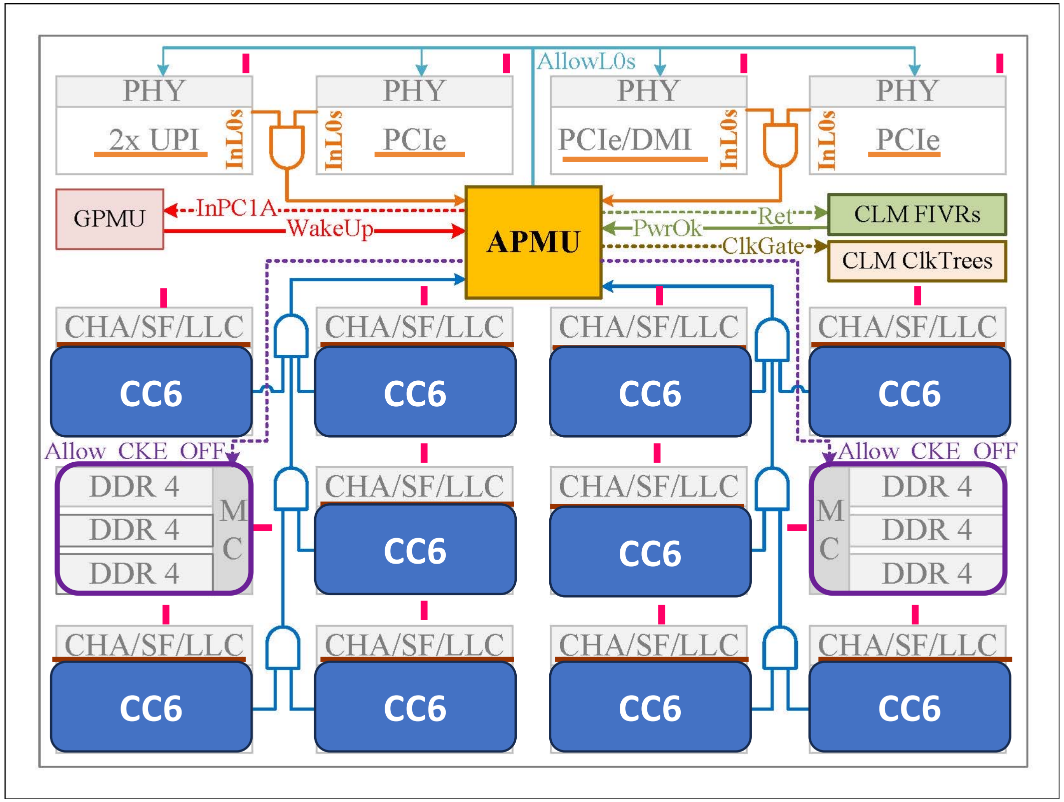
PC0 (Active) Package C-state



PC-state	CC-state	Clock	PLL	Voltage	LLC	IO	DRAM
PC0	C0-C6	On	On	Nom	Coherent	On	On

PC6 (Deep) Package C-state

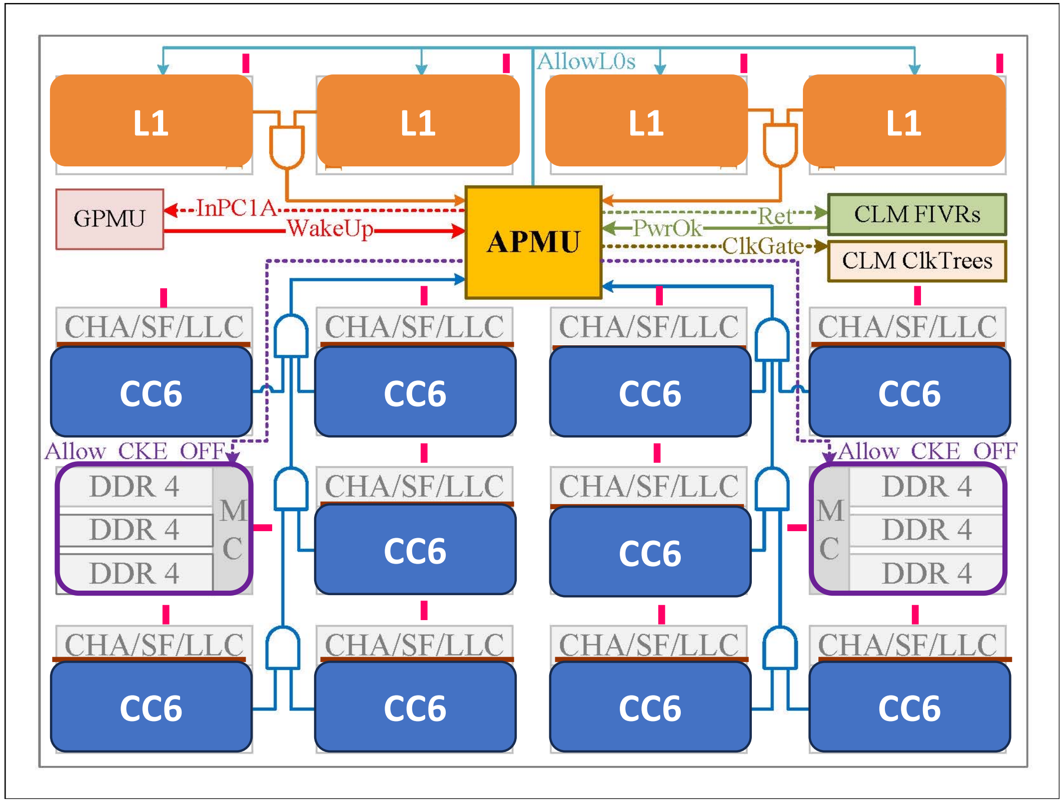
All cores in CC6



PC-state	CC-state	Clock	PLL	Voltage	LLC	IO	DRAM
PC6	C6	On	On	Nom	Coherent	On	On

PC6 (Deep) Package C-state

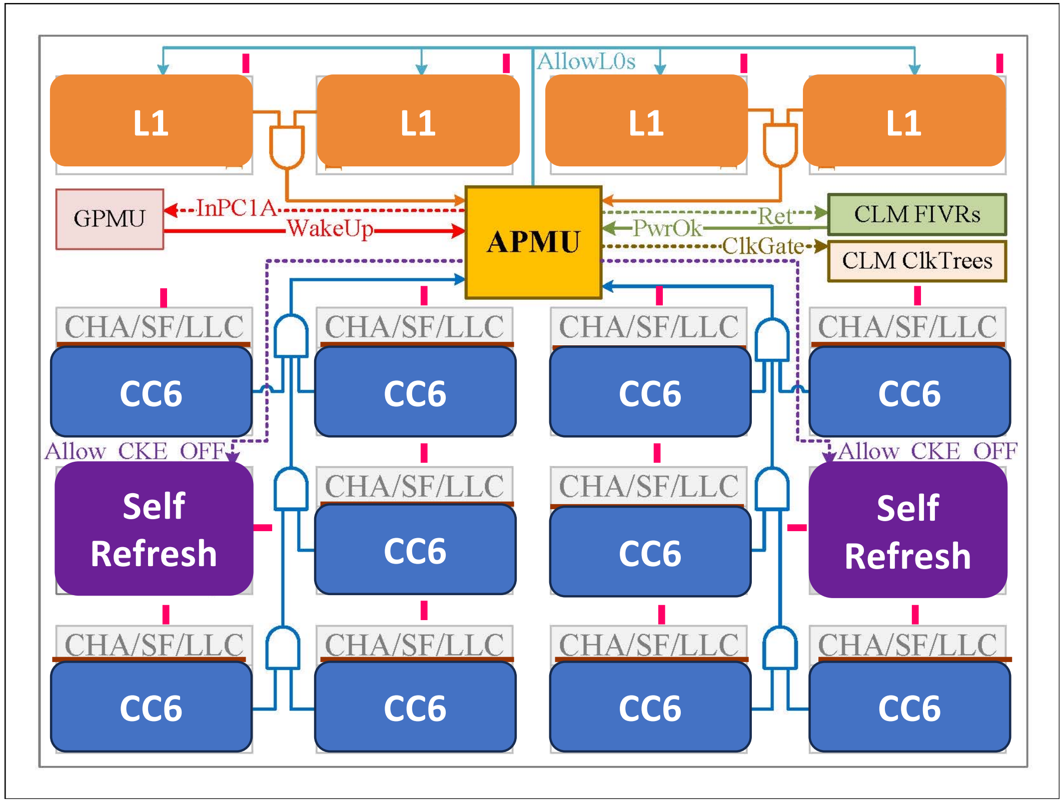
IOs in L1 state



PC-state	CC-state	Clock	PLL	Voltage	LLC	IO	DRAM
PC6	C6	On	On	Nom	Coherent	L1	On

PC6 (Deep) Package C-state

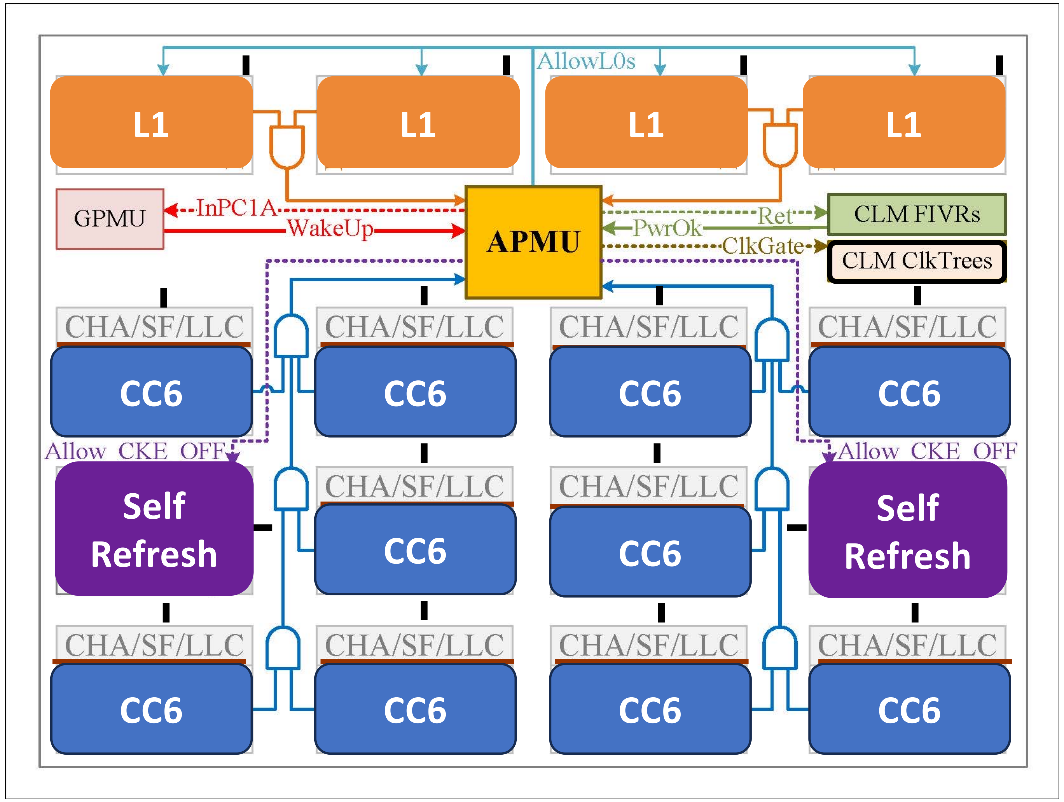
Dram in Self Refresh



PC-state	CC-state	Clock	PLL	Voltage	LLC	IO	DRAM
PC6	C6	On	On	Nom	Coherent	L1	SR

PC6 (Deep) Package C-state

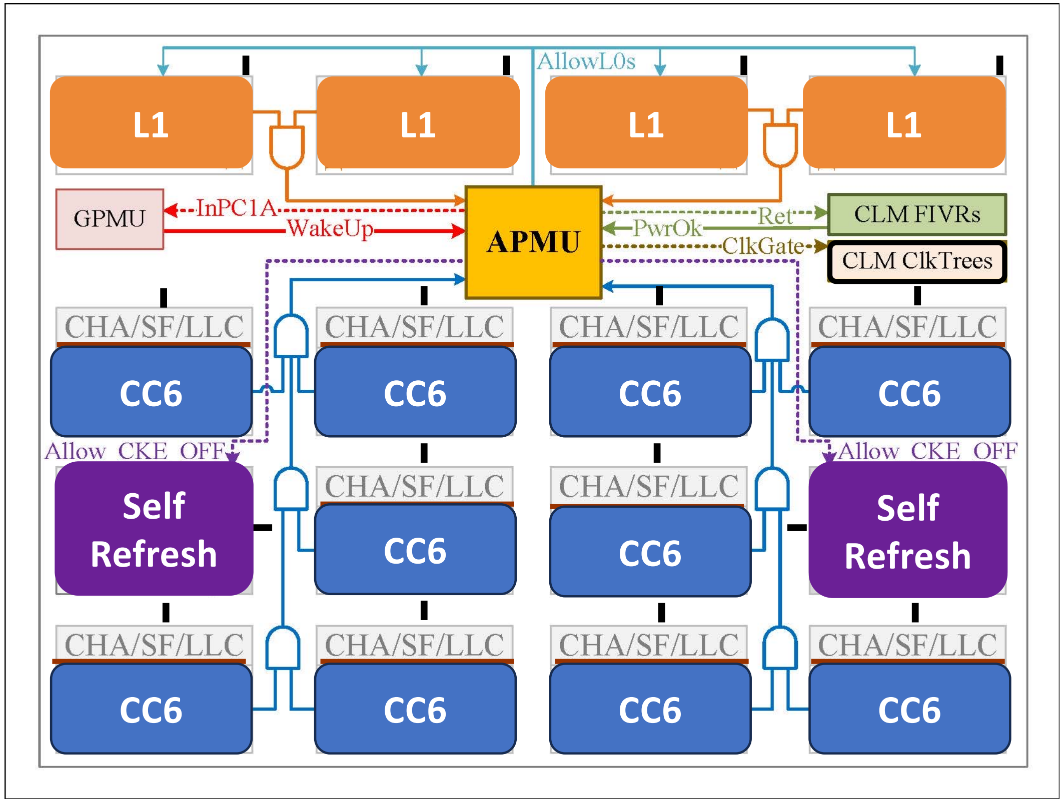
Turn-off the clocks and PLL



PC-state	CC-state	Clock	PLL	Voltage	LLC	IO	DRAM
PC6	C6	Stopped	Off	Nom	Coherent	L1	SR

PC6 (Deep) Package C-state

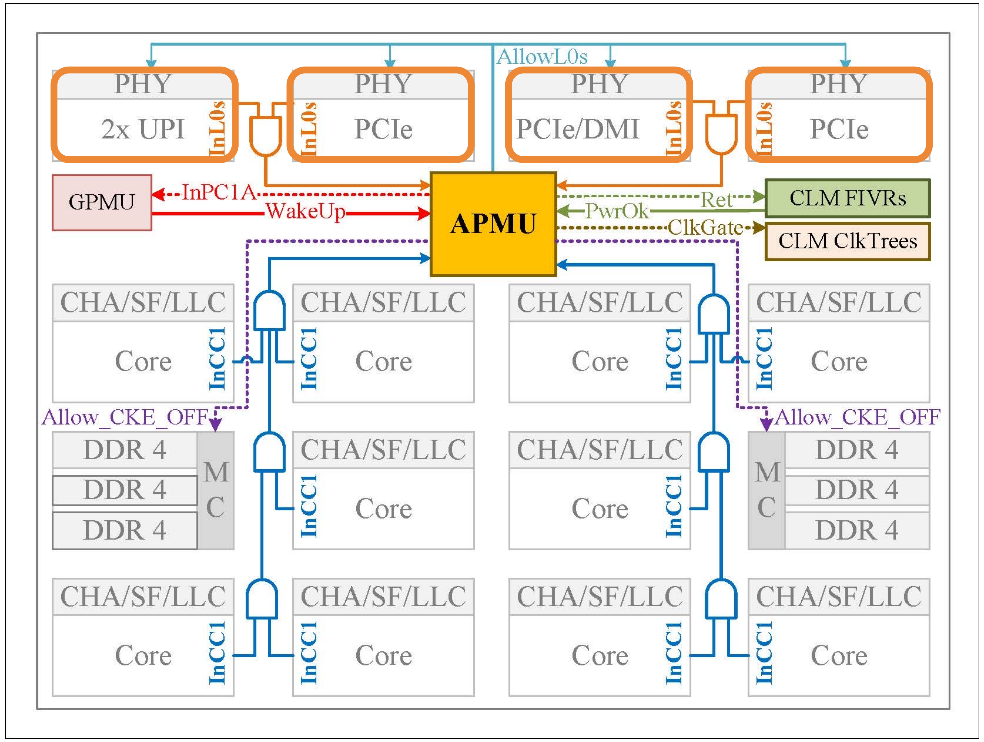
Reduce CLM voltage to retention



PC-state	CC-state	Clock	PLL	Voltage	LLC	IO	DRAM
PC6	C6	Stopped	Off	Ret	Coherent	L1	SR

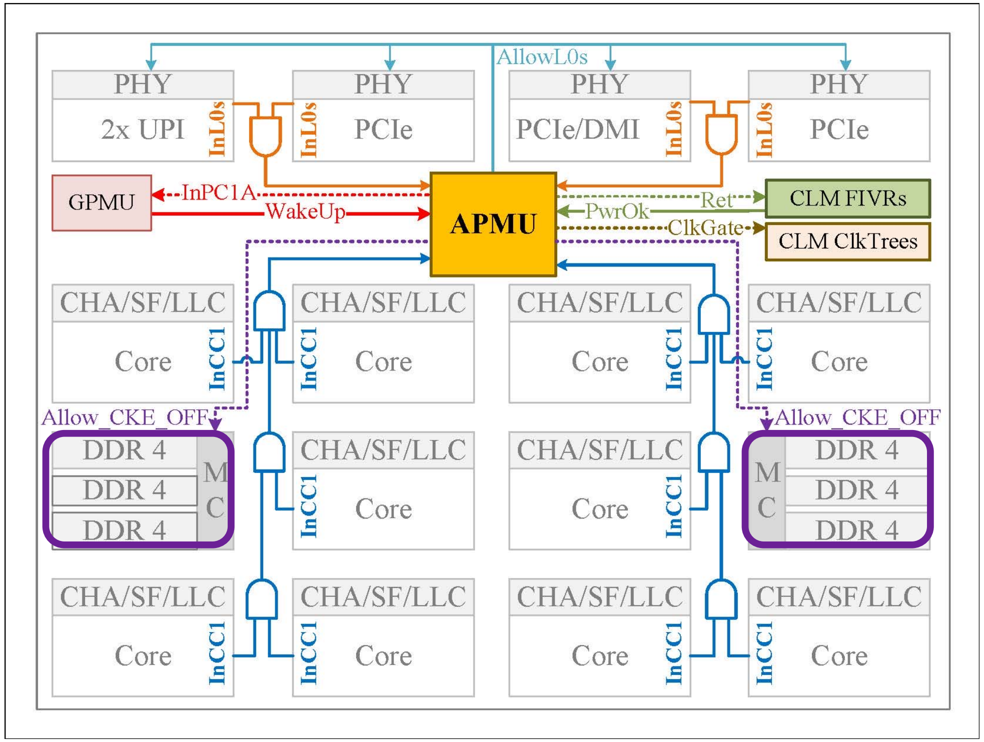
IO Link States

- Link Power States: Power saving states that enable the IO to reduce its power consumption during idle periods.
  - L0: Active
  - L0p: Partial Active (<10ns, 25%)
  - L0s: Standby (<64ns, 50%)
  - L1: Link Down (us)



DRAM Idle States

- CKE (Clock Enable)-OFF mode:
  - Normally, the MC (Memory Controller) sends clock signal to DRAM
  - When MC turns off the CKE signal the DRAM can enter either the Active Power Down mode or the Pre-charged Power Down (10-30ns, >50%).
- Self-refresh:
  - Normally, the MC sends refresh commands to DRAM
  - When DRAM enters self-refresh, DRAM is responsible to issue the refresh commands as a result the interface between the SoC and DRAM can be turned-off (several us).



Power Management Flow