Introduction to RAPL

-By Suman M(2014SIY7524)
Radhika D (2014SIY7530)

Power Measurement

- Power Measurement tools:
- Use hardware performance counters to measure energy and power values but are not accurate enough.
- Use external device which will measure the current being supplied to CPU but these are not granular enough.
- Turbo decisions are driven by models, which by nature tend to be conservative.

RAPL: Running Average Power Limit

- Intel introduced RAPL in Sandy Bridge microarchitecture ,
  - CORE series (i3, i5, i7),
  - Celeron ,Pentium,
  - Xeon E3 and E5.
- Power measurement is based on TDP (thermal design power) which is a "round up" average of power measurements of processor intensive benchmarks. Thus gives better and safe cut off.

RAPL

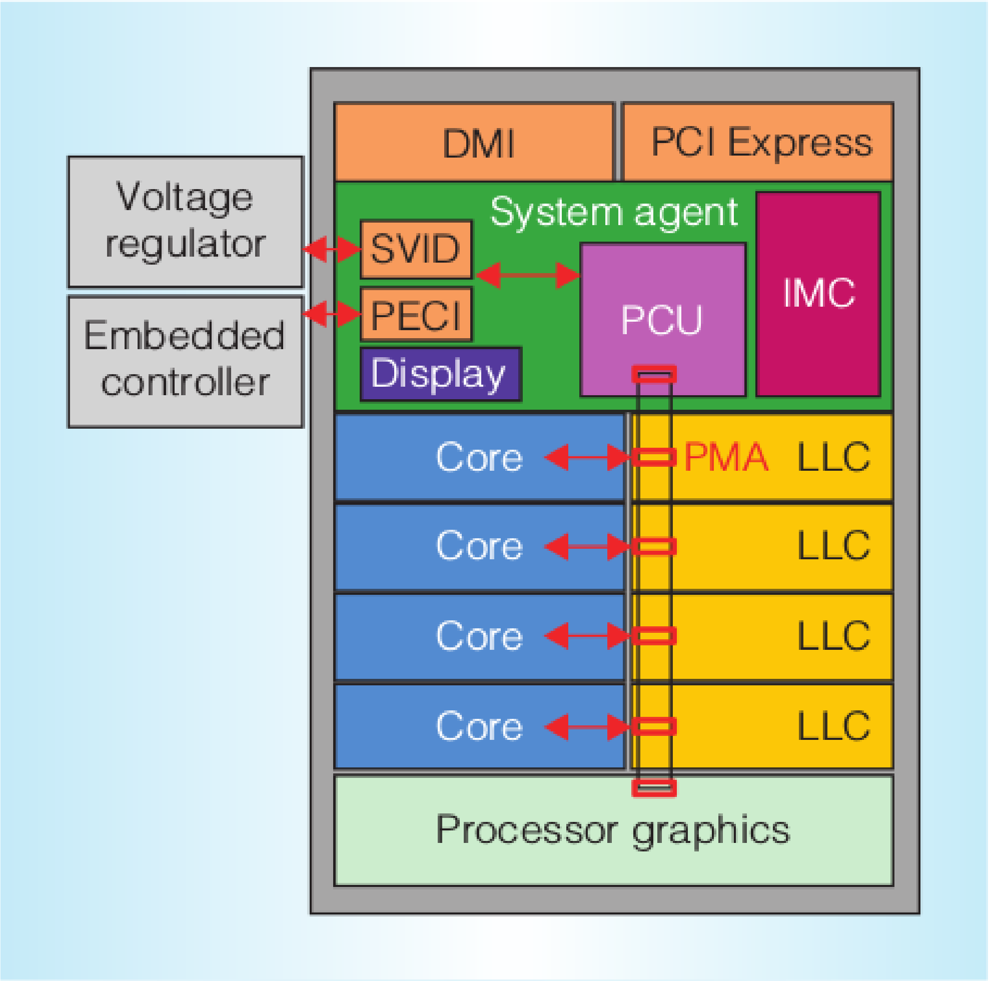

Figure 2. Sandy Bridge's power-management architecture. Sandy Bridge block diagram

RAPL: PCU (Package Control Unit)

- On chip logic and embedded controller running power mgmt firmware
- Communicates internally with cores, rings and SA
- Monitors physical conditions
  - Voltage, temperature, power consumption
- Control Power states
  - CPU and PG voltage and freq
  - Controls voltage regulators, DDR and system
- External power mgmt interface
  - External inputs
    - Accepts external
    - System pwr mgmt requests and limits
    - Power and temperature readings
  - MSR, MMIO and PECI system bus

RAPL: PCU (Package Control Unit)

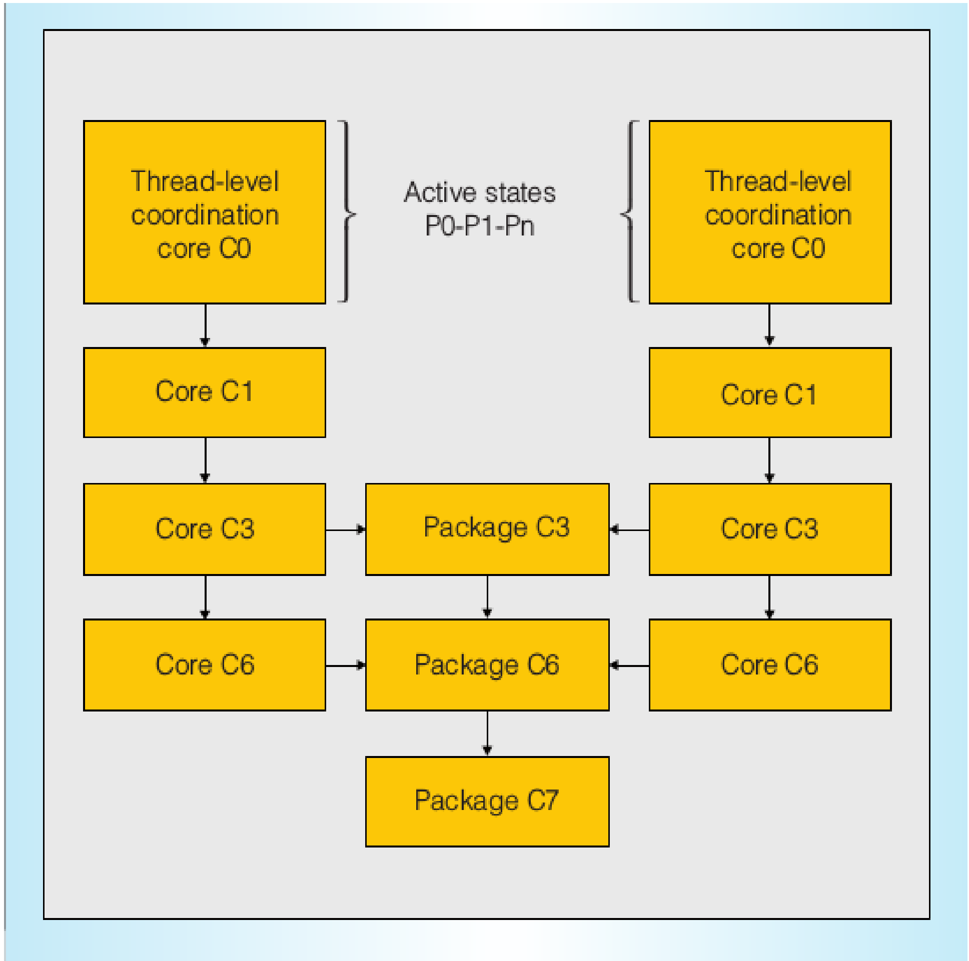

Figure 8. Sandy Bridge package C-state coordination.

- Sandy bridge introduced new PCU managed C-states
- Deeper C-states offers more power savings at the cost of longer latency enter and exit states
- OS controls each core individually
- Where as PCU coordinates between the cores and threads

RAPL: Running Average Power Limit

- P-states (power states): a voltage/frequency pair
  - P1 is guaranteed frequency
  - P0 max possible frequency
  - Pn is energy efficient state

RAPL

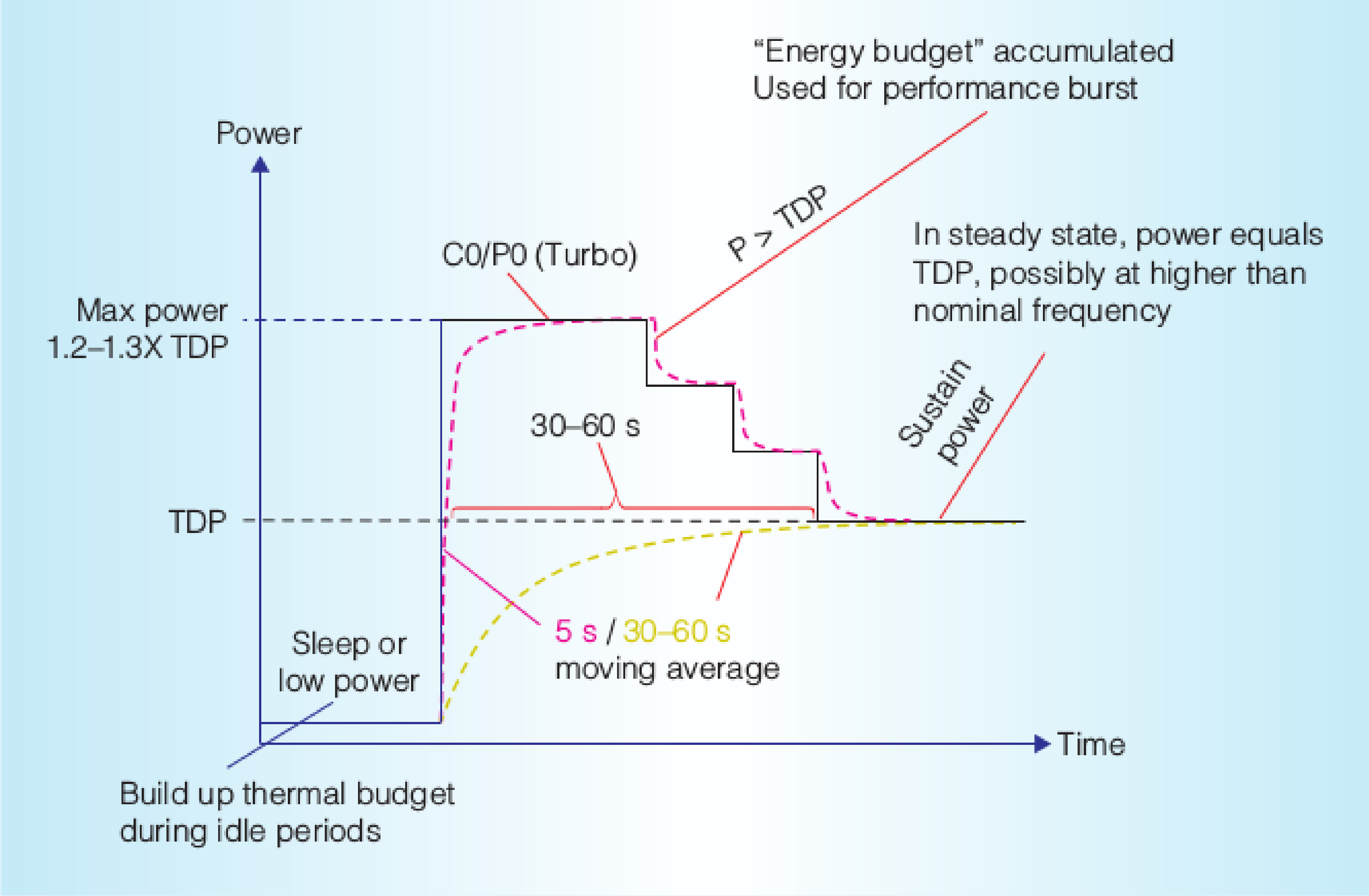

Figure 4. Dynamic behavior of the Intel Turbo Boost. After a period of low power consumption, the CPU and graphics can burst to very high power and performance for 30 to 60 seconds, delivering a responsive user experience. After this period, the power stabilizes back to the rated TDP.

RAPL

- RAPL Domains:
  - ENERGY_STATUS : for power monitoring
  - POWER_LIMIT and TIME_WINDOW : for controlling power
  - PERF_STATUS : for monitoring the performance impact of the power limit
  - RAPL_INFO : contains information on measurement units, the minimum and maximum power supported by the domain
  - For each of Package, PP0 (core device), PP1(uncore device) and DRAM.

RAPL

- Different tools available to measure power that use RAPL counters
  - Turbostat
  - PowerTop

- To read/write to MSR:
  - rdmsr [options] regno
  - wrmsr [options] regno value...
  - msrtool [-hvqrkl] [-c cpu] [-m system] [-t target ...] [-i addr=hi[:]lo] | [-s file] | [-d [:]file] | addr...

THANK YOU