Propagating thermal constraints to the scheduler

Punit Agrawal <punit.agrawal@arm.com>

CPU based cooling mechanisms

- Rely on controlling compute capacity to control power dissipation
- Control parameters
  - frequency
  - active core count
  - utilisation
- Work independent of the scheduler
  - essential to close the loop to maximize performance

ARM

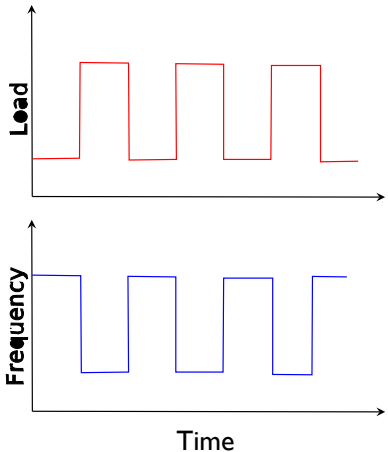
Building blocks

- Energy Aware Scheduling (EAS)
  - introduces core capacity and corresponding power consumption
- Power allocator
  - controls power consumption to control temperature
  - allocates power to requesting cooling devices
  - cooling devices use power model to translate power to performance and vice versa

Frequency control feedback

- Multiple DVFS domains are common in modern SoCs...
  - per-core
  - cluster of cores
  - multi-socket
- Heterogenous setups such as big.LITTLE
- Feedback the current maximum frequency / capacity
  - Due to capacity awareness, scheduler can now react to changes in DVFS domain capacity
  - On big.LITTLE, capacity inversion handled correctly as well
- Testing a prototype that builds on top of capacity awareness introduced by EAS

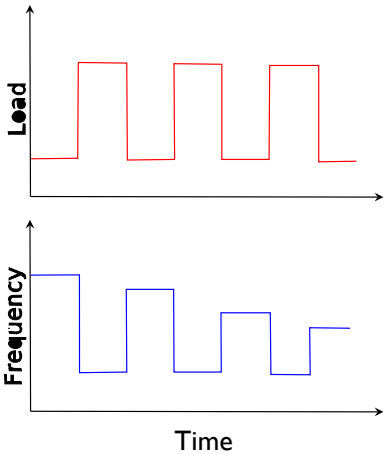
Single DVFS domain - I

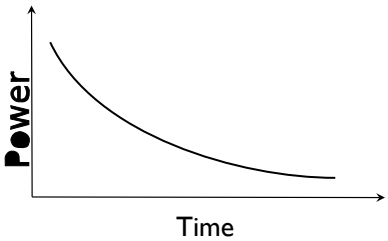


- For a power allocation, the behaviour today
  - translation to performance is modulated by load
  - load varying from 100% to N*100%
  - frequency cap inversely follows load
- Response at thermal sub-system granularity

ARM

Single DVFS domain - II

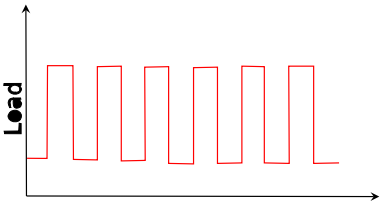


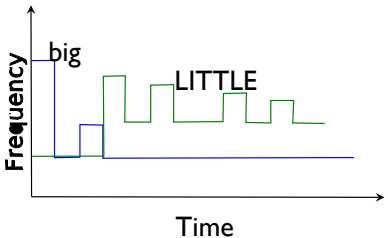


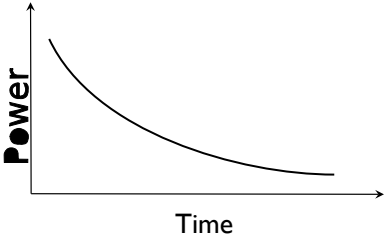
- Peak performance reducing with power allocation
  - continues load modulated translation to performance
- Response at thermal sub-system granularity

ARM

Multiple DVFS domains



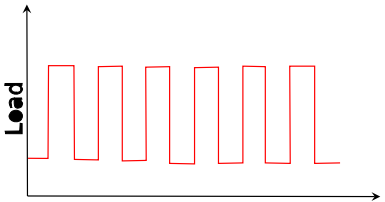


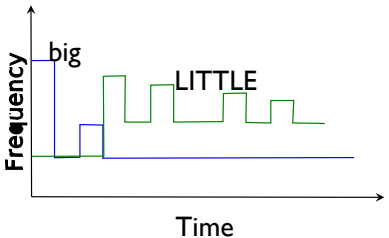


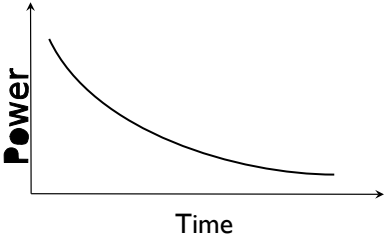
- Cooling device span all CPUs DVFS domain
  - heterogeneous uArch such as big.LITTLE
  - different peak frequencies

ARM

Multiple DVFS domains







- Cooling device span all CPUs DVFS domain
  - heterogeneous uArch such as big.LITTLE
  - different peak frequencies
- Three regions of behaviour
  - big frequency varying out of phase with load
  - LITTLE frequency varying as load
  - LITTLE frequency varying out of phase with load

ARM

Going further

- Respond at scheduler granularity
  - will hook into SCHED_DVFS

ARM

Going further

- Respond at scheduler granularity
  - will hook into SCHED_DVFS
- Beyond frequency control
  - idling cores in conjunction with scheduler

Thank You