Dynamic Energy Model to handle leakage power

Linux Plumbers Conference 2022

Lukasz Luba <lukasz.luba@arm.com>

September, 2022

Agenda

- What is the Energy Model (EM) in Linux kernel
- Relation between Energy Aware Scheduler (EAS) and EM
- Power and temperature relation in recent SoCs
- Runtime adjustable EM
- Other use cases

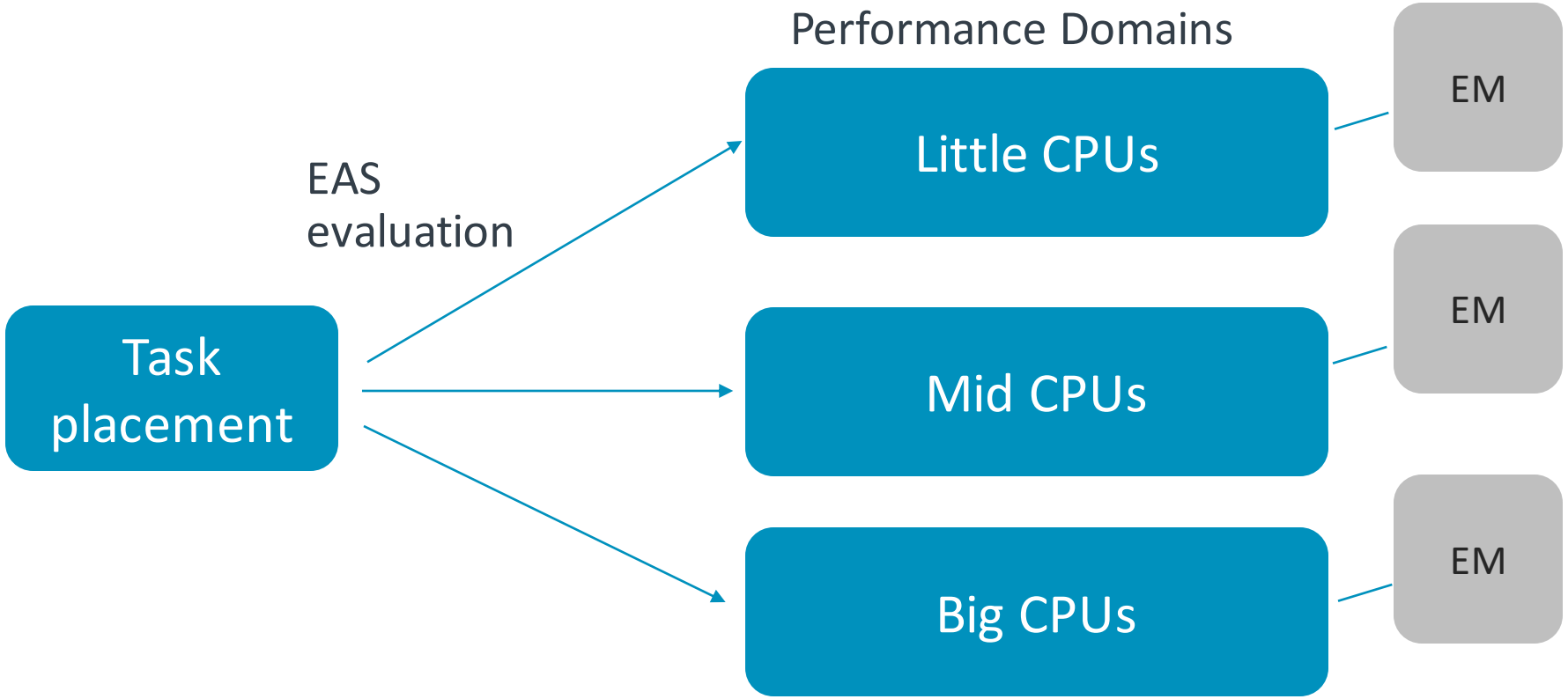
What is the Energy Model (EM) in Linux kernel

- A constant array of frequency, power and cost tuples (setup during boot)
- Contains 'cost' (based on power) for EAS to speed up calculation
- Decision/information source for EAS to make task placement decisions
- There is one EM for each Performance Domain

Frequency [kHz]	Power [uW]	cost
500000	79613	446151
851000	148208	487989
...	...	...
2704000	1801528	1866820
2802000	2158976	2158976

Relation between Energy Aware Scheduler (EAS) and EM

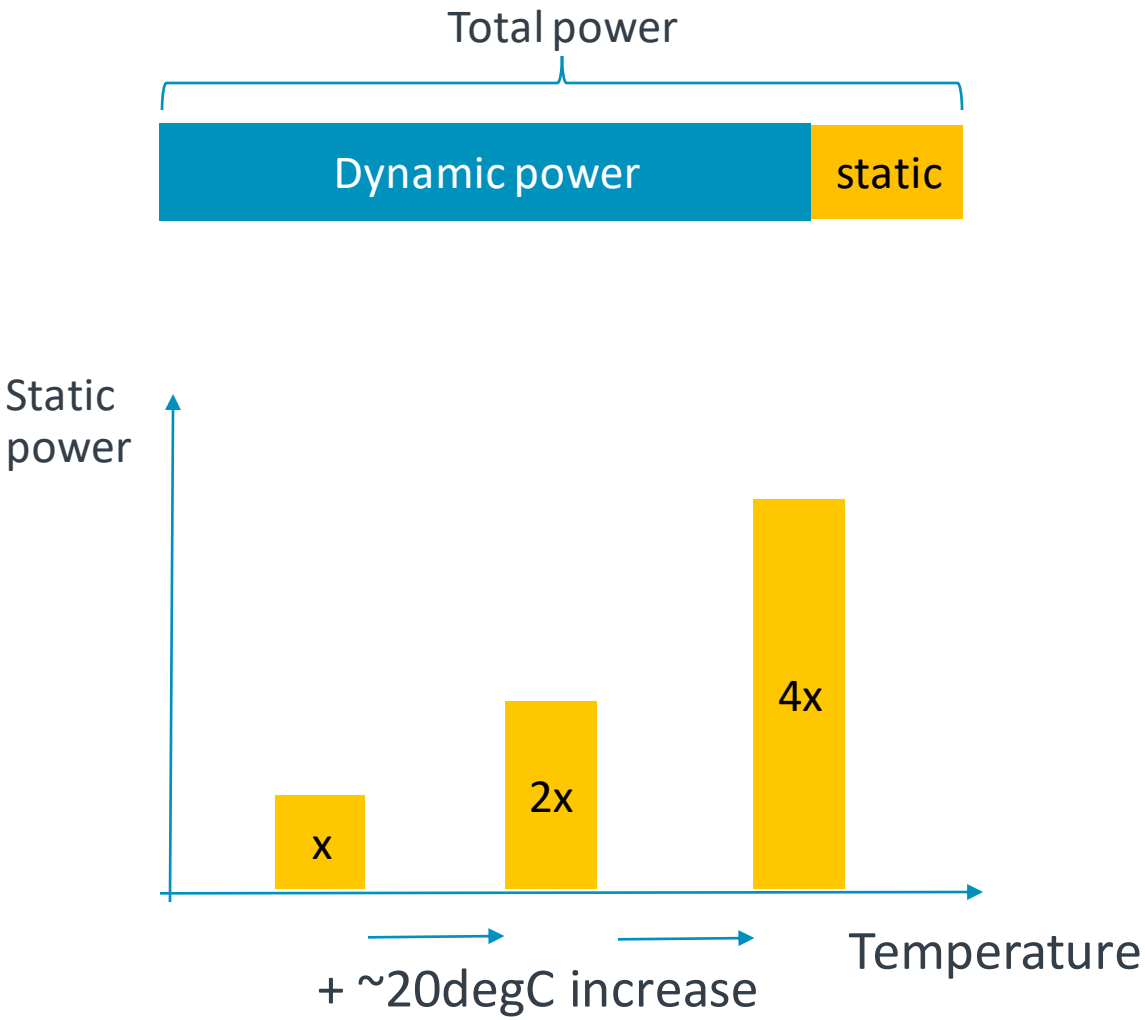
- EAS tries to minimize energy by looking at all possible Performance Domains (Big CPUs, Medium (Mid) CPUs, Little CPUs) when selecting the CPU for a woken-up task
- EAS uses EM information for comparisons



arm

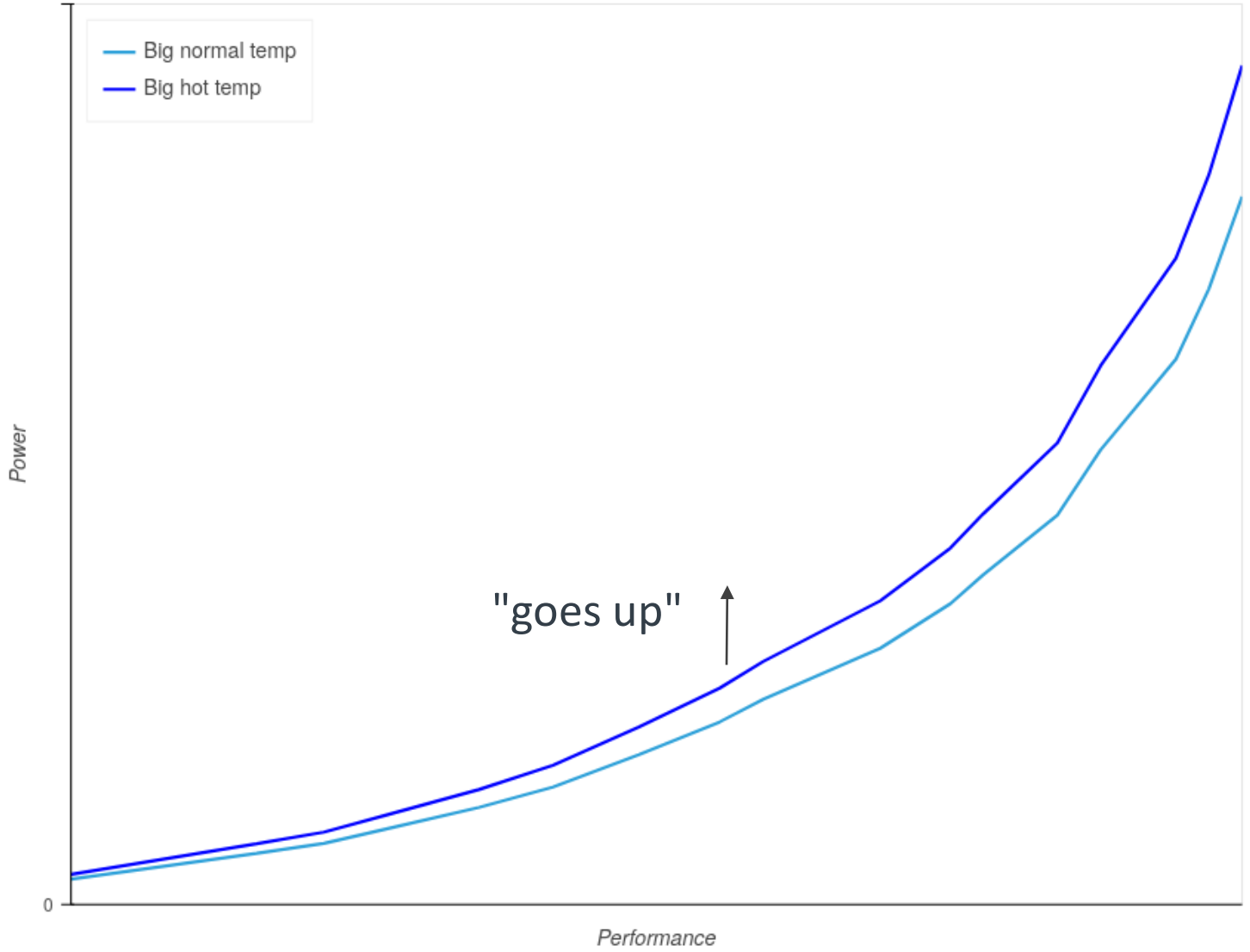
Power and temperature relation in recent SoCs

- Total power: dynamic + static power
- The static power (leakage) increases with temperature
- SoC has different types of CPUs, which are built with different goals: performance or power efficiency
- Performance cores (Big CPU) leak more
- Performance cores are more affected by the temperature increase than efficient cores (Mid CPU)



Power and temperature relation in recent SoCs

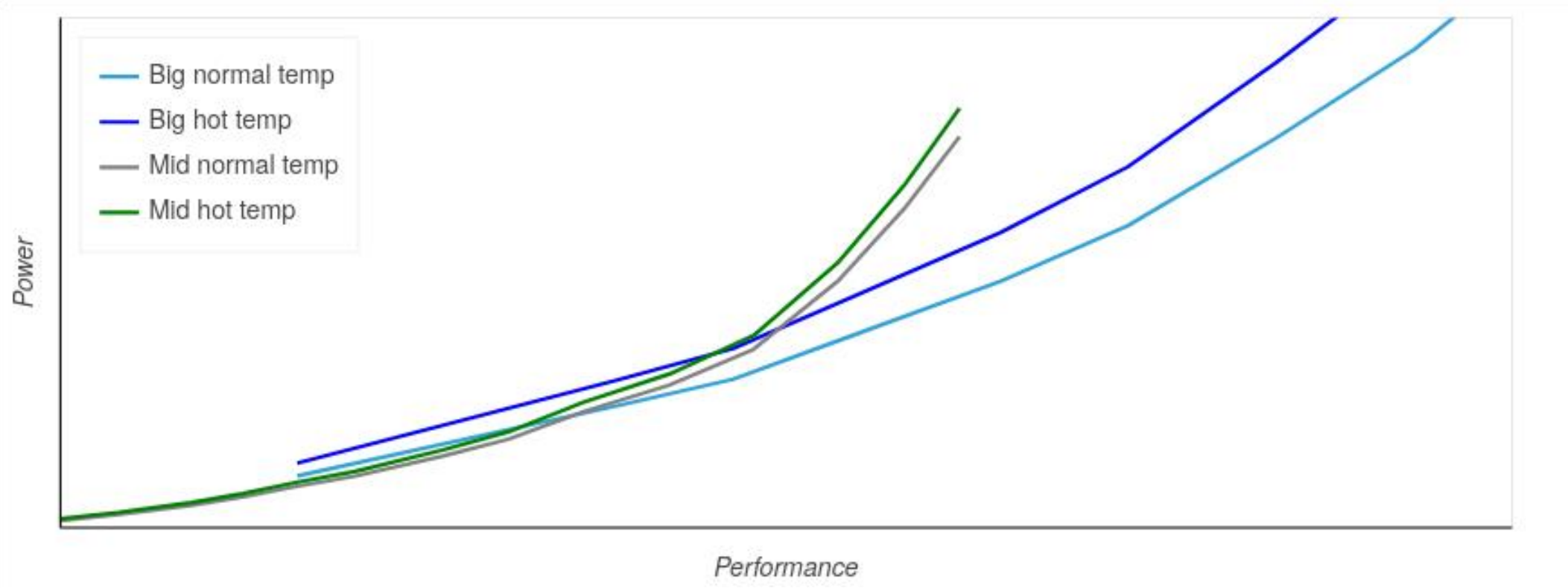
- Temperature of the SoC can be increased by adjacent devices (e.g. GPU)
  - e.g. increases the CPUs temperature by +20..30degC more than their normal temperature at the same frequency
- Power vs. Performance curve of a CPU can "go up"
  - Big CPU's curve would go up a bit more



arm

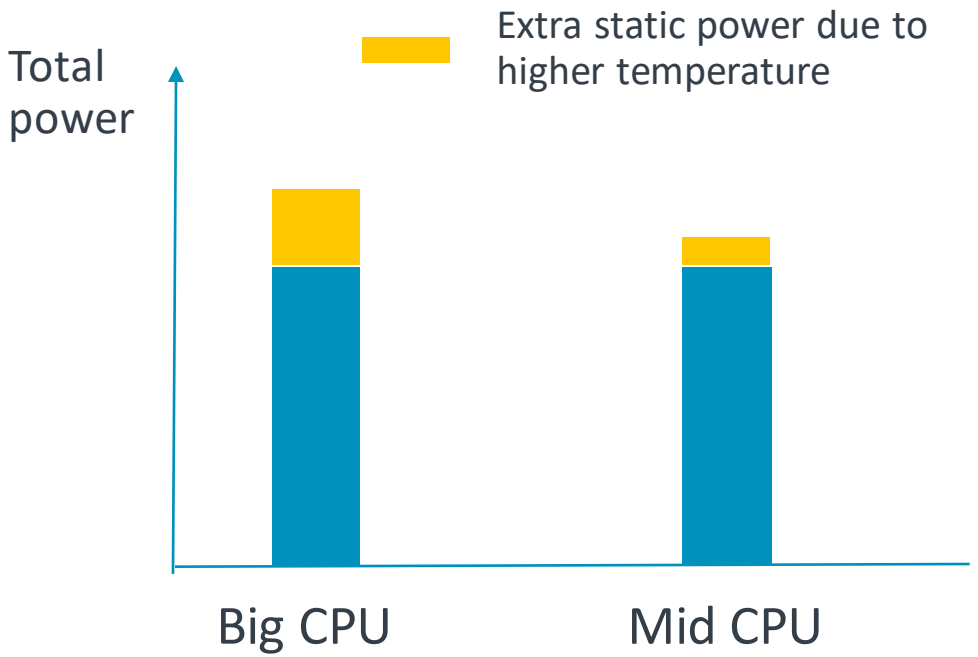
Power and temperature relation in recent SoCs

- Big CPU and Mid CPU curves position is different, therefore EAS decision should also be different



Power and temperature relation in recent SoCs

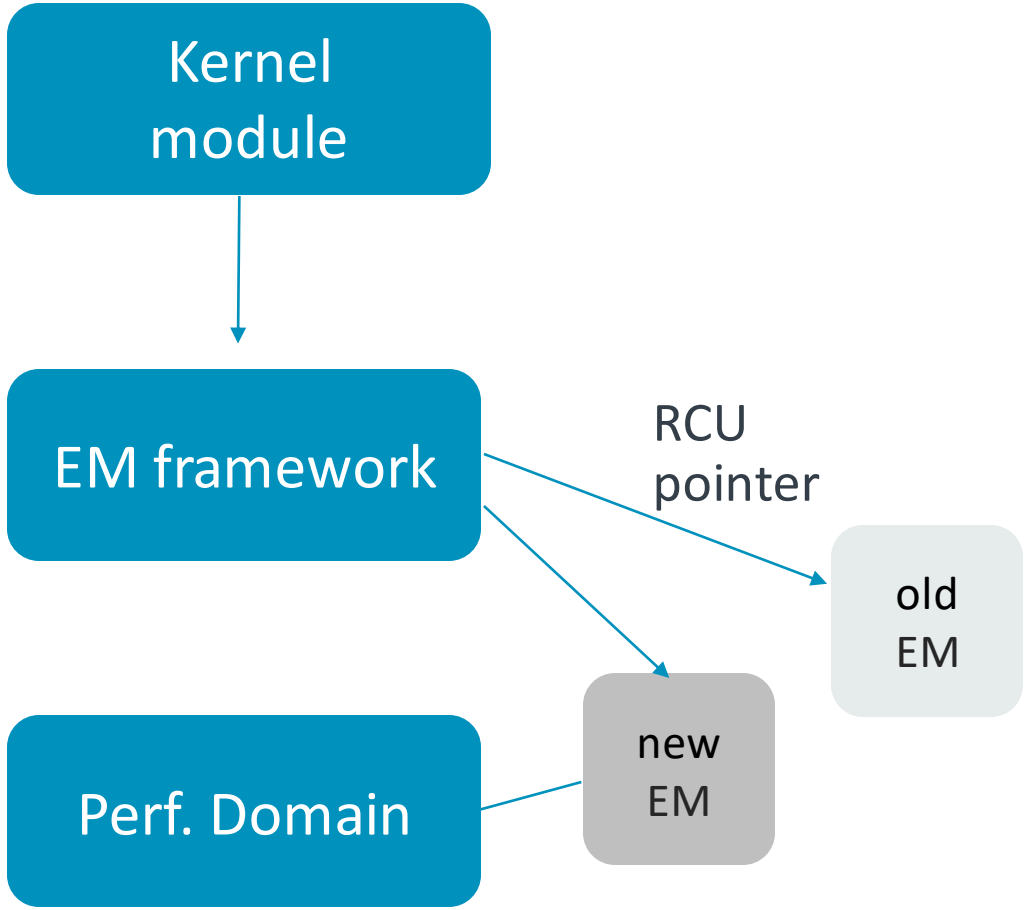
- Example power plots from a real phone (2021)
- CPU's temperature +20degC vs. normal due to GPU heat
- Big CPU Power increase
  - +15 … 18.5%
- Mid CPU Power increase
  - +5 … 8%



arm

Runtime adjustable EM

- Runtime EM change requested by a kernel module
  - No sysfs interface for a user-space
  - No thermal framework changes
- EM main data structure is allocated by the EM framework
  - EM is the memory owner (task scheduler requirement)
- New 'power' values are populated by the caller (kernel module)
- New EM data is used by the EAS during task placement (after RCU re-assignment for the pointer)



Other use cases

- allow to provide (after boot) the total power values for each frequency
  not limited to any formula or DT data
- allow to provide power values proper for a given SoC manufactured with different **binning** and read from FW or kernel module
- change EM at runtime for a specific workload on screen, which is utilizing HW resources differently (Gaming, video recording, web browsing)

arm

Thank You
Danke
Gracias
Grazie
谢谢
ありがとう
Asante
Merci
감사합니다
धन्यवाद
Kiitos
شكرًا
ধন্যবাদ
תודה

Links

- https://android-review.googlesource.com/c/kernel/common/+/1906500/1