New challenges in using LLC sched domain on wakeup path
AMD ZEN

Much smaller LLC SD compared to Intel.

- Each CCX has 4/8 cores with its own L3 cache.
- Each CCX is a NUMA node and LLC sched domain from kernel’s perspective.
- Pulling issue from the previous discussion is more severe.
- Less space to spread out loads during the wakeup.
ARM

No LLC SD
• No hyperthreading -> No SMT sched domain
• Some don’t expose L3 cache or SLC to the kernel -> No MC sched domain
• If kernel doesn’t see SMT and MC, it doesn’t have an LLC SD

How does it affect CFS wakeup?
• wake_wide() ALWAYS return 1, meaning no wake affine at all
• select_idle_sibling() cannot spread out tasks within a SD
• As a consequence, CFS wakeup always place a task back to the CPU the wakee was previous on no matter what
Should we continue to use LLC SD the way we are using in the wakeup path?

- Assuming we can fix the pulling issue, LLC SD on AMD chips may still be too small to work with. Cannot move wakee to idle CPUs outside two CCXes
- For some ARM chips, always place back to previous CPU. No smart placement at all. Can we do better?