Contribution ID: **24** Type: **not specified**

Fine grain frequency control with kernel governors

Wednesday, September 14, 2022 6:10 PM (10 minutes)

We introduced AMD P-State kernel CPUFreq driver [1] early of this year that is using ACPI CPPC based fine grain frequency control instead of legacy ACPI P-States, and it is merged into kernel 5.17 [2]. The AMD P-State will be used on most of the Zen2/Zen3 and future AMD processors.

There are two types of hardware implementations: "full MSR solution"and "shared memory solution". "full MSR solution"provides the architected MSR set of CPPC level registers to manage the performance hints which is the fast way to control frequency updates. However, "share memory solution"is that CPUs only support a mailbox model for the CPPC registers in the system memory, we have to map the system memory which shared with CPPC and use kernel RCU locks to manage synchronization with the way in kernel ACPI CPPC libraries.

The initial driver is developed on "full MSR solution"processors and can get better performance per watt scaling in some CPU benchmarks. However, we face the performance drops [3] which compared with legacy ACPI CPUFreq driver in "shared memory solution"processors. The traditional kernel governors such as ondemand, schedutil, and etc. might not be fully suitable for the fine grain frequency control, because there were 166~255 performance states in AMD P-State that compared with only 3 ACPI P-States. CPU CFS scheduler governor might face more frequently performance change in AMD P-State.

In following days, we will support more features include Energy-Performance Preference which is the balance between performance and energy and Preferred Core which is to have a best performance single core/thread in one package. We want to discuss how to refine the CPU scheduler or kernel governors to allow the platform to specify an order of preference that processes should be scheduled on the different cores.

In this session, we would like to have a discussion to how to improve the kernel governors which have better performance per watt scaling on fine grain frequency control, how to leverage the new Energy-Performance Preference and Preferred Core features to improve the Linux kernel performance and power efficiency.

For details of AMD P-State, please see [4].

References:

[1] https://lore.kernel.org/lkml/20211224010508.110159-1-ray.huang@amd.com/
[2] https://www.phoronix.com/scan.php?page=news_item&px=AMD-P-State-Linux-5.17
[3] https://lore.kernel.org/linux-pm/a0e932477e9b826c0781dda1d0d2953e57f904cc.camel@suse.cz/
[4] https://www.kernel.org/doc/html/latest/admin-guide/pm/amd-pstate.html

Primary author: HUANG, Ray (AMD)

Presenter: HUANG, Ray (AMD)

Session Classification: Power Management and Thermal Control MC