

Kernel driver fam15h_power

Supported chips:

- AMD Family 15h Processors
- AMD Family 16h Processors

Prefix: 'fam15h_power'

Addresses scanned: PCI space

Datasheets:

- BIOS and Kernel Developer's Guide (BKDG) For AMD Family 15h Processors
- BIOS and Kernel Developer's Guide (BKDG) For AMD Family 16h Processors
- AMD64 Architecture Programmer's Manual Volume 2: System Programming

Author: Andreas Herrmann <herrmann.der.user@gmail.com>

Description

1. Processor TDP (Thermal design power)

Given a fixed frequency and voltage, the power consumption of a processor varies based on the workload being executed. Derated power is the power consumed when running a specific application. Thermal design power (TDP) is an example of derated power.

This driver permits reading of registers providing power information of AMD Family 15h and 16h processors via TDP algorithm.

For AMD Family 15h and 16h processors the following power values can be calculated using different processor northbridge function registers:

- BasePwrWatts:
Specifies in watts the maximum amount of power consumed by the processor for NB and logic external to the core.
- ProcessorPwrWatts:
Specifies in watts the maximum amount of power the processor can support.
- CurrPwrWatts:
Specifies in watts the current amount of power being consumed by the processor.

This driver provides ProcessorPwrWatts and CurrPwrWatts:

- power1_crit (ProcessorPwrWatts)
- power1_input (CurrPwrWatts)

On multi-node processors the calculated value is for the entire package and not for a single node. Thus the driver creates sysfs attributes only for internal node0 of a multi-node processor.

2. Accumulated Power Mechanism

This driver also introduces an algorithm that should be used to calculate the average power consumed by a processor during a measurement interval T_m . The feature of accumulated power mechanism is indicated by CPUID Fn8000_0007_EDX[12].

- Tsample:
compute unit power accumulator sample period
- Tref:
the PTSC counter period
- PTSC:
performance timestamp counter
- N:
the ratio of compute unit power accumulator sample period to the PTSC period
- Jmax:
max compute unit accumulated power which is indicated by MaxCpuSwPwrAcc MSR C001007b
- Jx/Jy:
compute unit accumulated power which is indicated by CpuSwPwrAcc MSR C001007a
- Tx/Ty:
the value of performance timestamp counter which is indicated by CU_PTSC MSR C0010280
- PwrCPUave:
CPU average power

- i. Determine the ratio of Tsample to Tref by executing CPUID Fn8000_0007.

$N = \text{value of CPUID Fn8000_0007_ECX[CpuPwrSampleTimeRatio[15:0]]}.$

- ii. Read the full range of the cumulative energy value from the new MSR MaxCpuSwPwrAcc.

$J_{\max} = \text{value returned}.$

- iii. At time x , SW reads CpuSwPwrAcc MSR and samples the PTSC.

$J_x = \text{value read from CpuSwPwrAcc}$ and $T_x = \text{value read from PTSC}.$

- iv. At time y , SW reads CpuSwPwrAcc MSR and samples the PTSC.

$J_y = \text{value read from CpuSwPwrAcc}$ and $T_y = \text{value read from PTSC}.$

- v. Calculate the average power consumption for a compute unit over time period $(y-x)$. Unit of result is uWatt:

```
if (Jy < Jx) // Rollover has occurred
    Jdelta = (Jy + Jmax) - Jx
else
    Jdelta = Jy - Jx
PwrCPUave = N * Jdelta * 1000 / (Ty - Tx)
```

This driver provides PwrCPUave and interval(default is 10 millisecond and maximum is 1 second):

- power1_average (PwrCPUave)
- power1_average_interval (Interval)

The power1_average_interval can be updated at /etc/sensors3.conf file as below:

```
chip.fam15h_power-*
    set power1_average_interval 0.01
```

Then save it with "sensors -s".