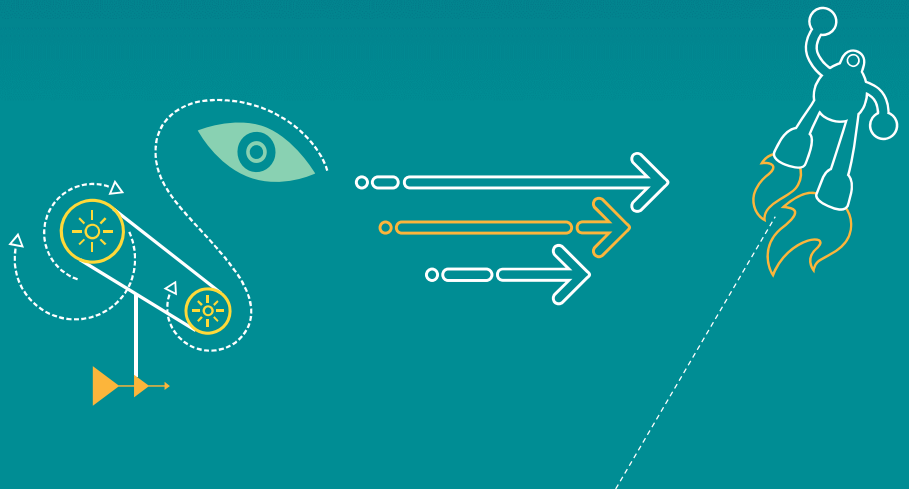
Bus DCVS Overview

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

Qualcomm Technologies, Inc.

80-PB236-1 Rev. C



Confidential and Proprietary – Qualcomm Technologies, Inc.

NO PUBLIC DISCLOSURE PERMITTED: Please report postings of this document on public servers or websites to: DocCtrlAgent@qualcomm.com.

Restricted Distribution: Not to be distributed to anyone who is not an employee of either Qualcomm Technologies, Inc. or its affiliated companies without the express approval of Qualcomm Configuration Management.

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc.

NO PUBLIC DISCLOSURE PERMITTED: Please report postings of this document on public servers or websites to: DocCtrlAgent@qualcomm.com.

Restricted Distribution: Not to be distributed to anyone who is not an employee of either Qualcomm Technologies, Inc. or its affiliated companies without the express approval of Qualcomm Configuration Management

Not to be used, copied, reproduced, or modified in whole or in part, nor its contents revealed in any manner to others without the express written permission of Qualcomm Technologies, Inc.

MSM is a product of Qualcomm Technologies, Inc. Other Qualcomm products referenced herein are products of Qualcomm Technologies, Inc. or its subsidiaries.

Qualcomm and MSM are trademarks of Qualcomm Incorporated, registered in the United States and other countries. Other product and brand names may be trademarks or registered trademarks of their respective owners.

This technical data may be subject to U.S. and international export, re-export, or transfer (“export”) laws. Diversion contrary to U.S. and international law is strictly prohibited.

Qualcomm Technologies, Inc.
5775 Morehouse Drive
San Diego, CA 92121
U.S.A.

© 2017 Qualcomm Technologies, Inc. and/or its affiliated companies. All rights reserved.

Revision History

Revision	Date	Description
A	March 2017	Initial release
B	October 18, 2017	Numerous changes were made to this document; it should be read in its entirety
C	October 20, 2017	Changed the title of the document

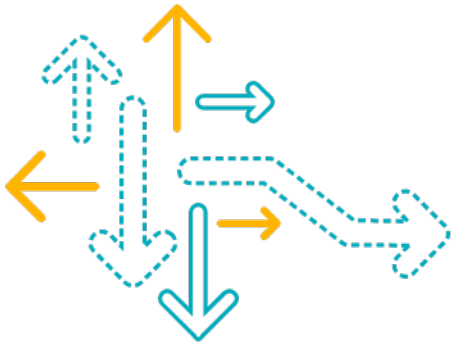
Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

 Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Contents

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION



Introduction

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

 Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS

- Bus dynamic clock and voltage scaling (DCVS) is a bus frequency selection algorithm to select the optimal bus operating point during the variable system workload, using the low-overhead hardware and software algorithm.
- Dynamic control of bus DCVS is necessary to overcome limitations, such as slow response and the inefficient legacy approach that uses static CPU to DDR mapping.
- Final DDR frequency is decided by a maximum of the aggregated votes from different clients.
  - For details, refer to *SDM845 RPM Hardening Overview and Debug* (80-P9301-16).

Qualcomm

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS v3 for SDM845

- The existing bus DCVS governors, (bw_hwmon and memlat) are applied for scaling more devices.
  - Bwmon governor handles the bandwidth bound workload by scaling LLCC and DDR based on the traffic from CPU to LLCC, and LLCC to DDR, respectively.
  - Memory latency governor handles the latency bound workload by scaling L3 and DDR based on IPM(L2$) and IPM (L3$).
- There is not much of a difference in the key scaling framework and algorithm.
  - It is the same algorithm and same bwmon hardware.

Qualcomm

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS v3 for SDM845 (cont.)

MSM8998 vs. SDM845 chipsets:

Chipset	Scaling device	Devfreq device	Devfreq governor	Data monitored
MSM8998	DDR	cpubw	Bw_hwmon	[CPU <-> DDR] traffic
	DDR	memlat-cpu0 memlat-cpu4	Memlat	Instructions per L2$ miss ratio
SDM845	LLCC	cpubw	Bw_hwmon	[CPU <-> LLCC] traffic
	DDR	llccbw	Bw_hwmon	[LLCC <-> DDR] traffic
	DDR	memlat-cpu0 memlat-cpu4	Memlat	Instructions per L3$ miss ratio
	L3	l3-cpu0 l3-cpu4	Memlat	Instructions per L2$ miss ratio

Qualcomm

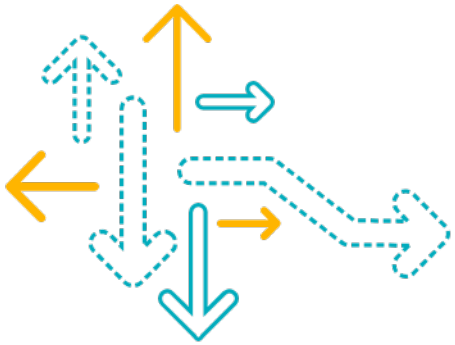
Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS v3

- Aggregated bandwidth (AB) – The average bandwidth of data transfer a bus master expects to do with the double data rate (DDR).
- Instantaneous bandwidth (IB) – An indication of the amount of DDR latency a bus master is willing to tolerate.
- Final DDR frequency = Max (sum (AB_1..AB_n), Max (IB_1..IB_n))

Qualcomm
2018-07-23 23:41:57 PD
songpeng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION



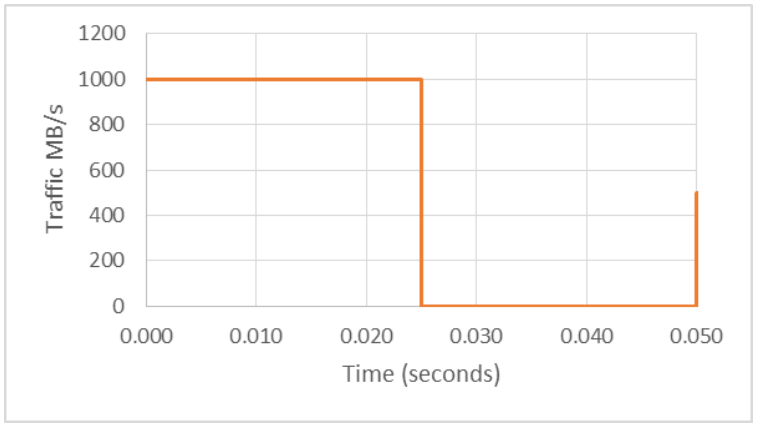
Bus DCVS v1 and v2 Overview

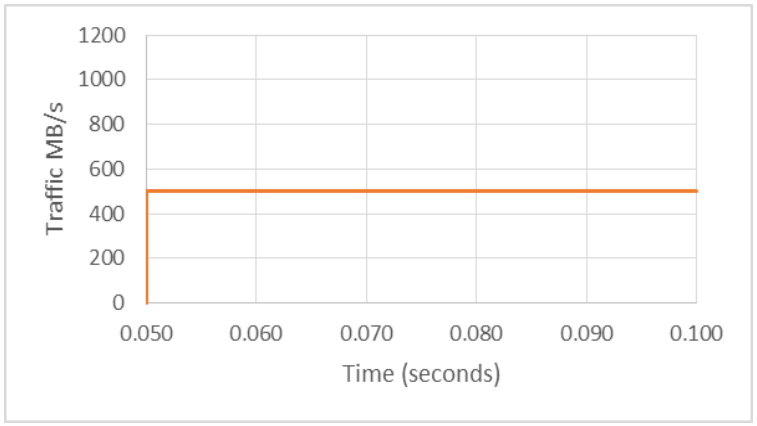
Qualcomm
2018-07-23 23:41:5 CST
songpeng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS v1

- Bus DCVS v1 samples traffic every 50 ms.
  - It uses interrupts to reduce the sampling time in proportion to load increase.
  - However, when the load increases by 4x, it still takes 12.5 msec to react, which is still relatively slow.
- A large sampling window also gives a very low resolution picture of the traffic and smooths out the peaks and valleys.
- Since the CPU and bus master traffic can be very bursty, these tiny bursts of peak traffic affect the performance of the CPU and bus master.
- Not having a traffic resolution to spot these peaks can result in poor AB/IB choice, as the following two graphs cannot be distinguished from one another.





Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS v2

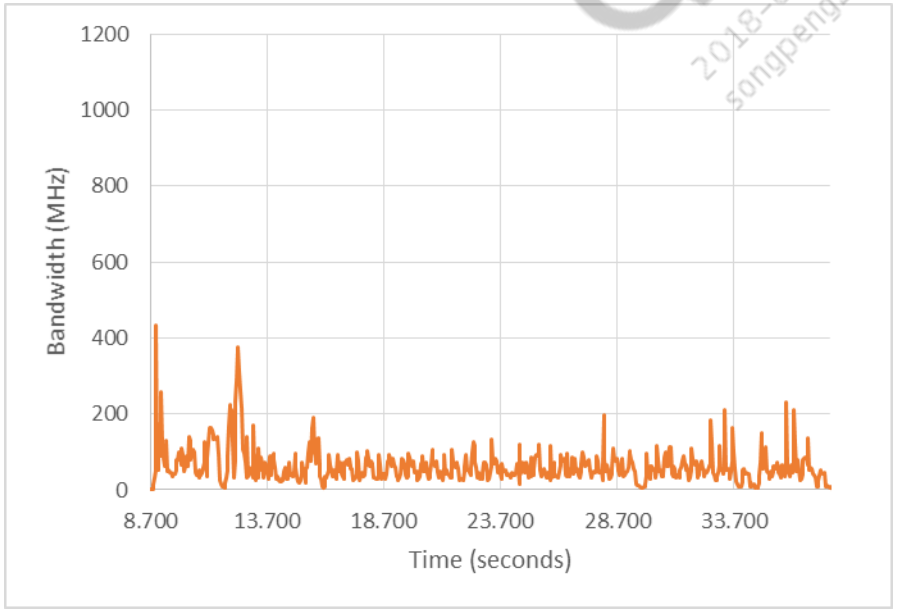
- All improvements in bus DCVS v2 are built on having a high resolution view of the traffic.
  - To achieve this, bus DCVS v2 uses a 4 msec short sample window and a 50 msec decision window.
  - This allows a better picture of the peaks and valleys and also allows distinguishing between a steady load and sporadic load.
- The higher resolution sampling allows for more advanced improvements to the algorithm; it is implemented completely in the software without a noticeable overhead.
  - Sampling is done in the interrupt context to reduce the task switch latency.
  - Performs intelligent stretching of the 4 msec window for CPU bound use cases to reduce overhead.
  - Automatically stops 4 msec sampling when the bus master goes idle.

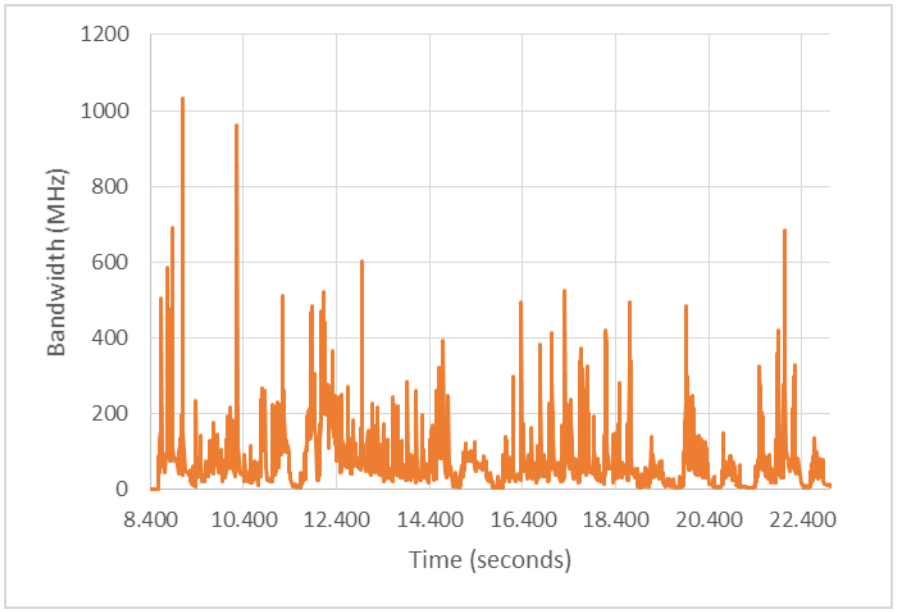
Qualcomm 2018-07-23 21:57:01 sivapeng2@qualcomm.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS v1 vs. v2 – Resolution Comparison

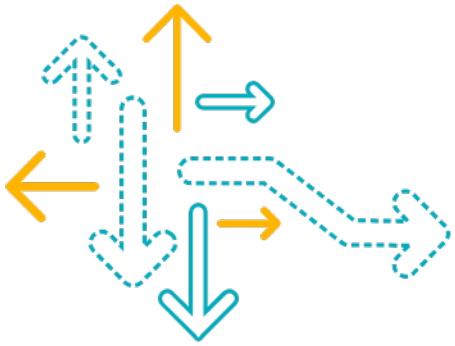
- The following charts are from the browser section of the PCMark benchmark.
- The measured bandwidth is plotted in terms of MHz (Mbps divided by bus width) for easier visualization.
- v1 (left) does not measure anything past 400 MHz.
- However, v2 (right) captures peaks all the way up to 1000 MHz.
- The peaks in the middle sections are completely averaged out in v1.





Qualcomm
2018-07-23 ... songpeng2@huawei.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION



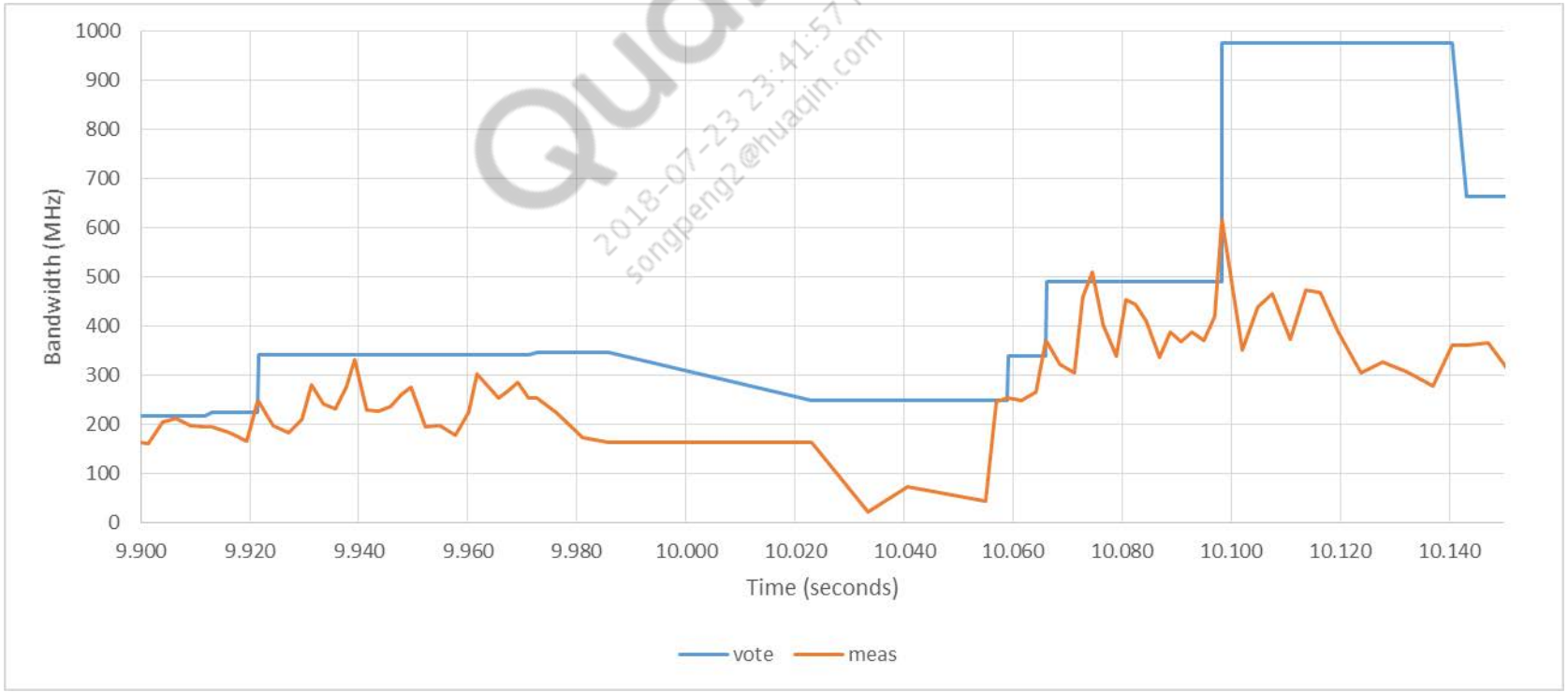
Bus DCVS v2 Features

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

 Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Quick Reaction Time, Stay Ahead of Traffic

- Since the samples are taken every 4 msec and counter IRQs are still used, the maximum reaction time is 4 msec.
  - When the load increases by 4x, the reaction time is cut down to 1 msec.
- The algorithm also overvotes (up_scale) when traffic increases to stay ahead of the traffic increase.

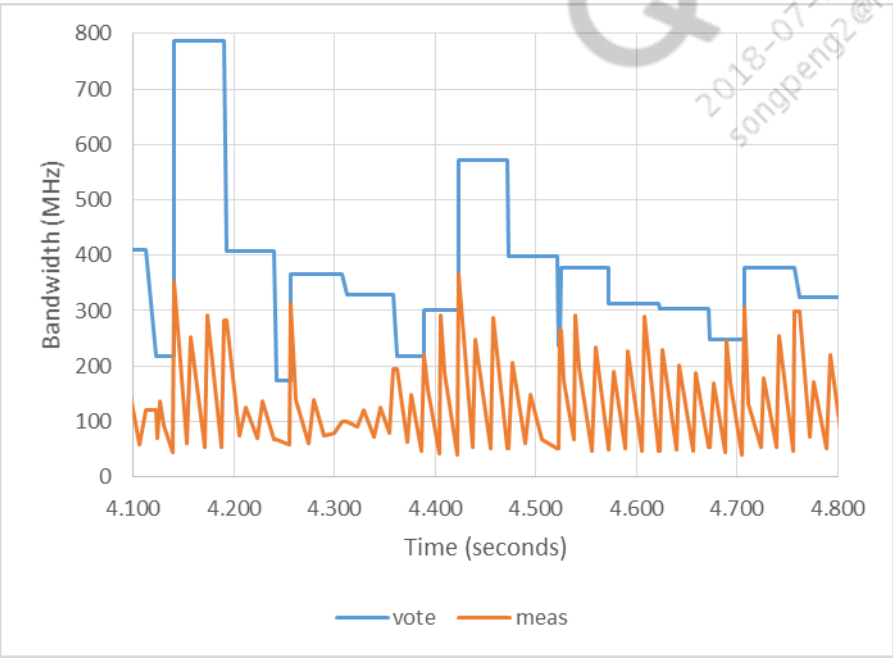


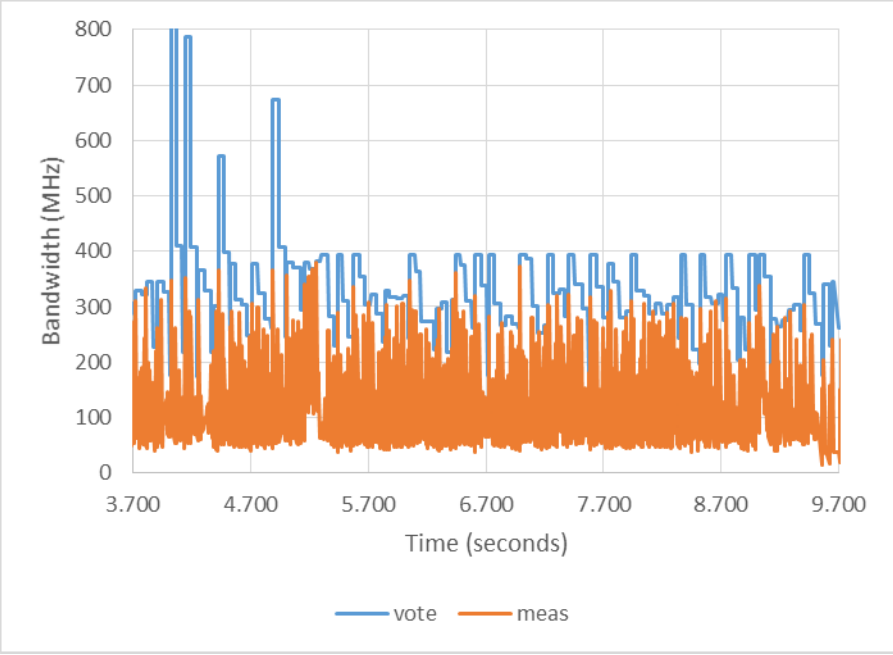
Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Smart Overvote Based On History

- Overvoting can be incorrect (traffic does not increase) and can push the overall DDR residency in higher frequency regions from 0% to x% for low-power use cases.
- To be smart about overvoting, the algorithm looks at the historic peaks in past hist_memory decision windows.
  - If the measured bandwidth is less than the historic peak, the over-voting value is limited to the historic peak.



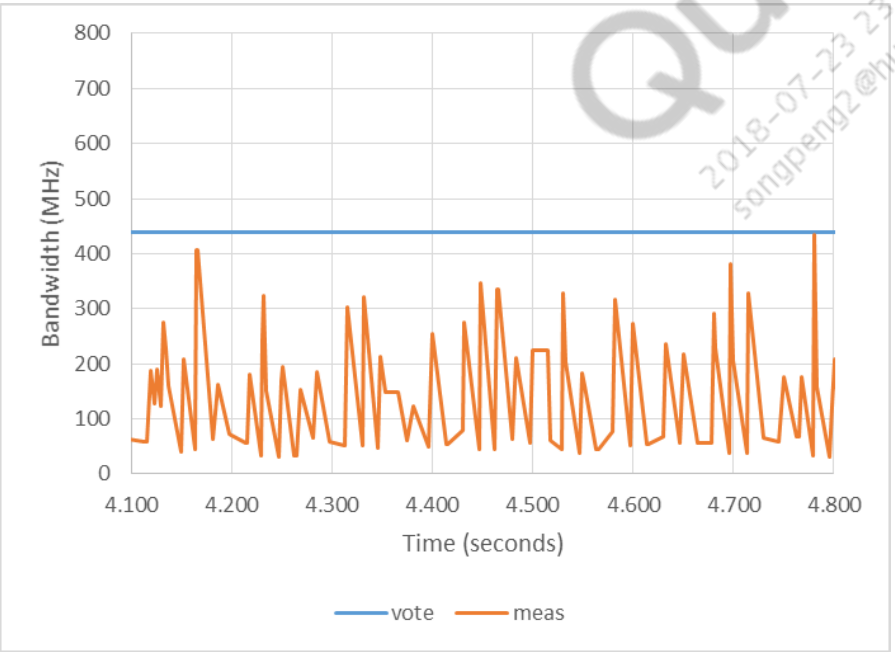


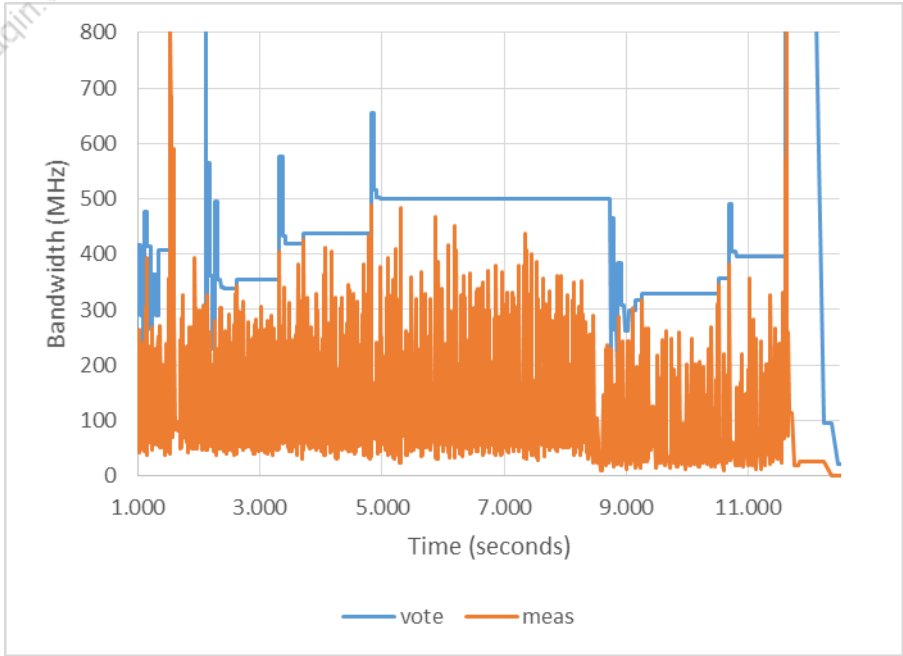
Qualcomm
2018-07-23 23:41:3 PDT
songpeng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Pattern Detection and Hysteresis

- The algorithm detects patterns in the traffic, predicts future bandwidth increases, and keeps the bandwidth vote (DDR frequency) high in anticipation of the increase.
  - This is done by looking for hyst_trigger_count repeating peaks of similar height within hyst_length decision windows. If it does, it then kicks in hysteresis and keeps it enabled as long as the peaks keep repeating.



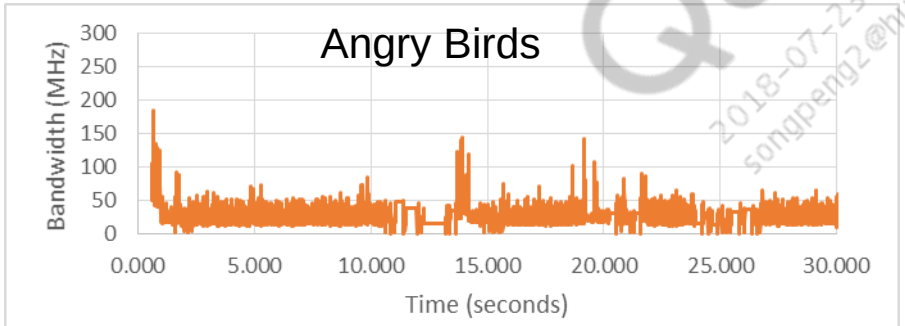


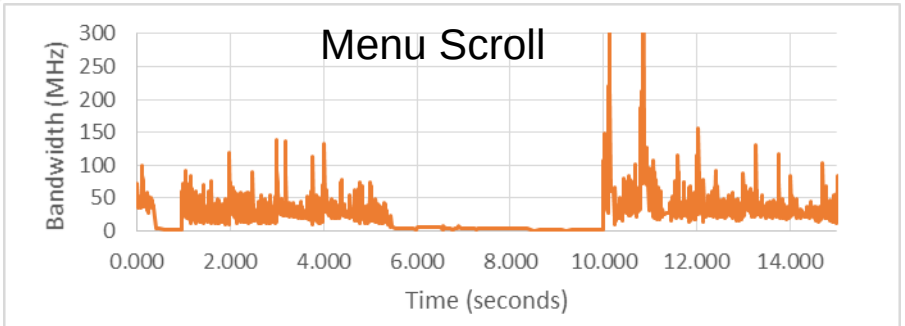
Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

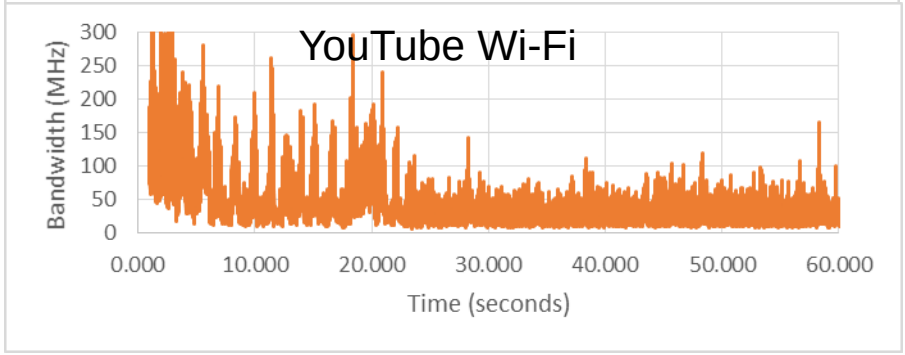
Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

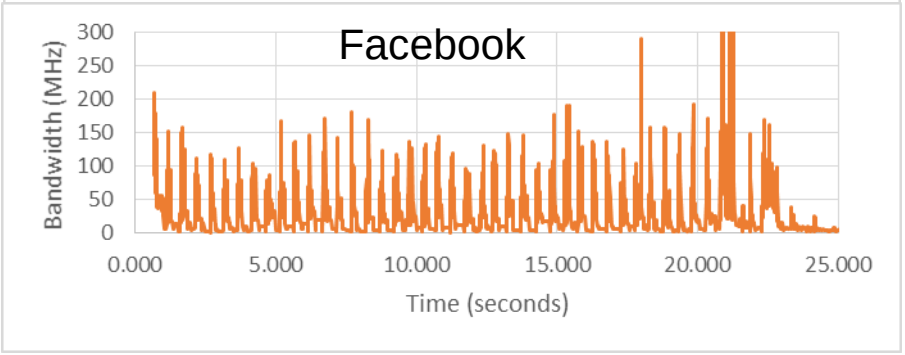
Detect Low-Power Mode, Being Conservative

- A lot of power-sensitive use cases have very little traffic from the CPU to DDR.
  - In these cases, the aggressiveness of the io_percent tunable used in the formula, IB = AB * 100 / io_percent, can be reduced.
- If the measured bandwidth is below low_power_ceil_mbps for low_power_delay decision windows, then the algorithm uses the less aggressively tuned low_power_io_percent for determining IB.









Qualcomm
2018-07-23 songpeng2@huaqin.com

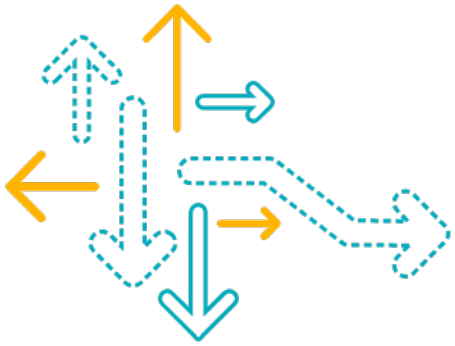
Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Quick Dropping Vote When Traffic Stops

Since the decision window is 50 msec and there is a quick reaction time for traffic increases, the algorithm supports quickly dropping the bandwidth vote if the measured traffic is below the down_thres% of the current bandwidth vote for down_count short sample windows.

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

 Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION



Bus DCVS v2 Tunables

Qualcomm
2018-07-23 23:41:5 PDT
songpeng2@huaqin.com

 Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Bus DCVS Tunables Summary (SDM835 and SDM845)

Category	Tunable	Default value	Description
Bandwidth measurement and AB/IB estimation	sample_ms	4	Short sampling window
	polling_interval	50	Period of decision window
	io_percent	34	▪ Value used to compute IB from AB ▪ Percentage of time CPU spent accessing DDR (I/O)
	bw_step	190	▪ Bandwidth vote's are rounded up to multiples of bw_step ▪ Smaller value results in more frequent bandwidth changes ▪ Higher value results in less frequent bandwidth vote updates, but could also result in unnecessarily higher bandwidth vote caused by rounding up
Overvoting	up_thres/down_thres	10/0	▪ Bandwidth threshold values are to detect abrupt bandwidth changes ▪ For example, if the bandwidth measured during each short sample window increases multiple times by up_thres percentage of the maximum bandwidth seen during previous decision window, the current decision window is terminated early to issue a new bandwidth vote
	up_scale	250	When the decision window terminates early, up_scale scaling factor is used to overestimate the required bandwidth in proportion to the increase in traffic
	hist_memory	20	Based on the maximum traffic bandwidth seen in the most recent hist_memory decision windows, the over-voting value is limited to the historic peak if the measured bandwidth < historic peak
Pattern detection and hysteresis	hyst_trigger_count	3	By looking for hyst_trigger_count repeating peaks of similar height within hyst_length decision windows, it then kicks in hysteresis and keeps it enabled as long as the peaks keep repeating
	hyst_length	10	
	idle_mbps	1600	If the max_mbps of a decision window is below idle_mbps, then the algorithm suppresses Hysteresis mode for that decision window (only for the current window)
Low-power mode	Low_power_ceil_mbps	0	▪ When low_power_delay consecutive decision windows have a max_mbps value <low_power_ceil_mbps>, the less aggressively tuned low_power_io_percent is used for IB calculation ▪ Set ceil_mbps to 0 to disable feature ▪ *QTI recommends to disable this feature without strong data
	Low_power_io_percent	34	
	Low_power_delay	20	

*Tunables are subject to change

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Power Impact of Tuning Parameter (SDM835)

- By increasing io_percent parameter, bimc runs at lower frequency, leading to savings in DDR power.
- However, note the regression on the CPU power after increasing the io_percent to 70.
  - Therefore, parameter values should be chosen with discretion.

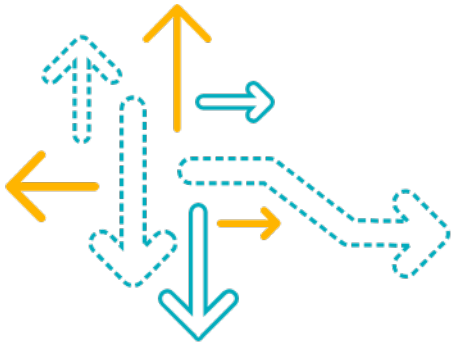
Power impact of different io_percent values (io_percent@34 – io_percent@70)

Usecase	CX+MX+CPU+ EBI+GFX	CX	MX	Total CPU	EBI	GFX
Playing Subwaysurf	5.96 mA	5.12 mA	0.70 mA	-4.23 mA	4.93 mA	0.06 mA

DDR residency

DDR Freq.	Residency(%)	
	IO_percent @34	IO_percent @70
100	0.00%	0.00%
150	0.00%	0.00%
200	0.06%	0.09%
300	0.09%	68.05%
412	0.08%	17.71%
547	17.26%	13.10%
681	51.52%	0.92%
768	12.00%	0.14%
1017	16.45%	0.00%
1296	2.48%	0.00%
1555	0.06%	0.00%
1804	0.00%	0.00%

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION



Bus DCVS v3

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

 Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Overview

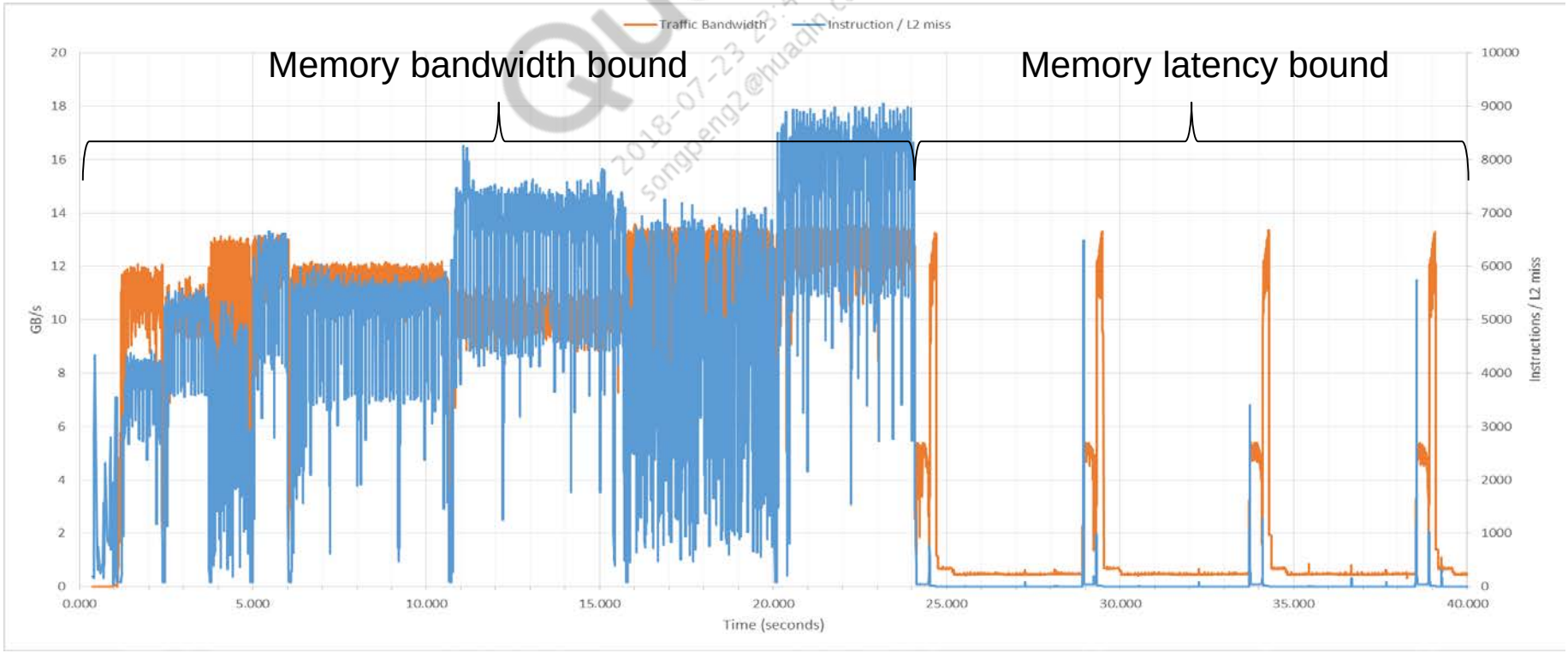
- BWMON v2
  - Existing DCVS v2, bw_hwmon governor, is enhanced by offloading the traffic tracking feature to hardware, BWMON v2.
  - Hardware samples traffic periodically on every micro sampling windows. (default is 4 msec).
  - BWMON v2 triggers less interrupts than BWMON v1 does because it triggers interrupts only when the bandwidth needs to be changed immediately.
  - All other scaling algorithms stays same as in DCVS v2.
- Memory latency governor
  - Designed to handle memory latency bound workloads more effectively.
  - Detects memory latency bound workload and scales DDR/Bus frequency.
  - Scales bus and DDR frequency to operating point that brings good balance between performance and power.

Qualcomm

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Memory Latency Governor – Detect Memory Latency-Bound Workloads

- The following figure shows that the memory latency bound workloads do not generate large traffic.
  - Hence, the existing algorithm does not handle them effectively.
- Low instructions per L2 miss ratio, IPM, is observed during memory latency bound workloads.
- Memory latency governor uses PMU in each CPU to calculate number of IPMs.
  - If it is less than the “ratio_ceil” threshold, the workload is classified as memory latency bound workload.



Qualcomm
2018-07-23 22:41:57 PDT
songpeng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Memory Latency Governor – Decide Bandwidth

- Memlat governor scales bimc to a higher frequency when the CPU runs at higher frequencies.
- Memlat governor uses PMU to calculates CPU frequency.
- The maximum frequency of cores with an IPM less than ratio_ceil is used to look up a mapping table of core frequency to a required bandwidth vote.

  Examples (ratio_ceil is 400 for all)

Case 1:

Core	IPM	Freq. (Mhz)
Core0	270	120
Core2	300	310

Core	IPM	Freq (Mhz)
Core1	200	640
Core3	70	700

The IPM of all cores is less than ratio_ceil. Core3's frequency is the largest. Memlat governor votes **4173** as IB.

Case 2:

Core	IPM	Freq. (Mhz)
Core0	430	120
Core2	532	310

Core	IPM	Freq (Mhz)
Core1	800	30
Core3	649	700

Memlat governor does **NOT** vote any because the IPMs of all cores are greater than ratio_ceil.

Case 3:

Core	IPM	Freq. (Mhz)
Core0	460	1100
Core2	100	310

Core	IPM	Freq (Mhz)
Core1	210	320
Core3	349	920

The IPM of cores 1, 2 and 3 are less than ratio_ceil. Core3's frequency is largest among them. Memlat governor votes **7759** as IB.

CPU frequency to IB mapping table

Core frequency (Hz)	IB
300000	1525
480000	3143
900000	4173
1017000	7759
1296000	9887

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Memory Latency Governor – Impact on Power (SDM835)

The memory latency governor may increase DDR Power, but faster DDR frequency reduces the CPU time taken in processing memory latency bound workloads and may reduce overall system power.

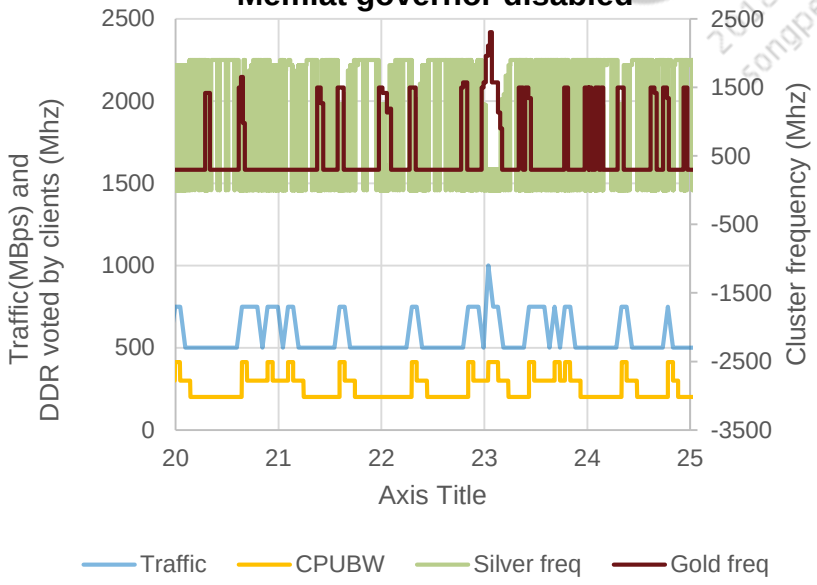
Power impact of Memlat governor
(Memlat governor disabled - Memlat governor enabled)

UseCase	CX + MX + CPU + EBI + GFX	CX	MX	CPU	EBI	GFX
Scrolling contacts	-2.08	-1.94	-0.41	1.77	-0.79	-0.71
Typing message on SMS	-1.59	-1.74	-0.05	0.88	-0.49	-0.19
SuwaySurf	8.16	-12.78	-2.8	30.32	-5.29	-1.29

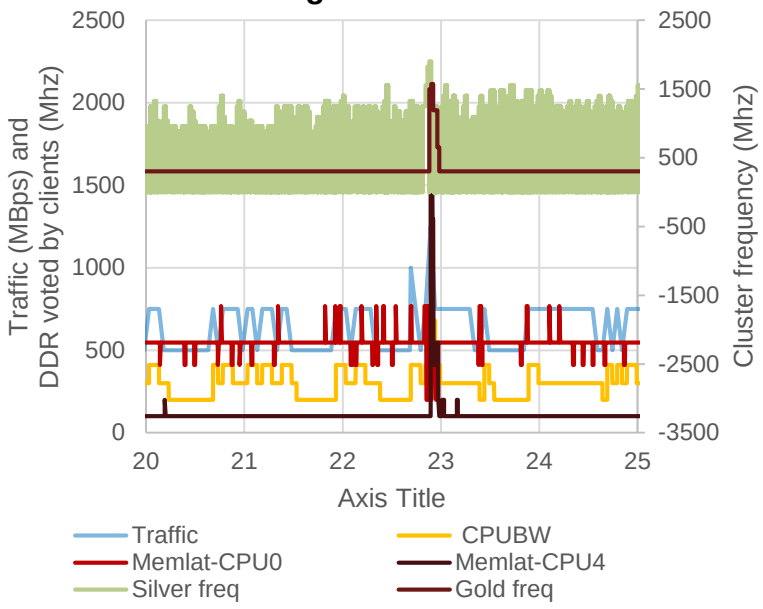
Regression on CX, EBI with memlat

Gain on CPU power with memlat

Memlat governor disabled



Memlat governor enabled



Qualcomm
2018-07-23 23:41:57 PDT
songpeng@qti.qualcomm.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Memory Latency Governor – Tunables Summary (SDM835)

Name	Description	Default value	Description
ratio_ceil	▪ Memlat governor compares IPM against ratio_ceil ▪ If it is less than ratio_ceil memory latency, governor classifies workload as memory latency bound and scales DDR and bus frequency	400	Write to /sys/class/devfreq/*qcom,memlat-cpu*/mem_latency/ratio_ceil
polling_interval	Polling_interval for monitoring performance counters and voting DDR and bus frequency (unit: msec)	10	Period of decision window
qcom,core-dev-table	▪ A mapping table of core frequency to a required bandwidth vote at the given core frequency. ▪ Each entry consists of one pair of CPU frequency and IB. ▪ This table is hard coded in a device tree file (tunables are subject to change)	**Memlat-cpu0** < 300000 1525 > < 499200 3143 > < 1113600 4173 > < 1881600 5859 > **Memlat-cpu4** < 300000 1525 > < 480000 3143 > < 900000 4173 > < 1017000 7759 > < 1296000 9887 > < 1555000 11863 > < 1804000 13763 >	Modify kernel device tree entry “qcom,core-dev-table”

Qualcomm 2018-07-23 23:41:57 FDT shaoheng2@huaqin.com

Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

Memory Latency Governor – Tunables Summary (SDM845)

Name	Description	Default value	Description
ratio_ceil	▪ Memlat governor compares IPM against ratio_ceil. ▪ If it is less than ratio_ceil memory latency governor classifies workload as memory latency bound and scales DDR and bus frequency	400	Write to /sys/class/devfreq/*qcom,memlat-cpu*/mem_latency/ratio_ceil
polling_interval	Polling_interval for monitoring performance counters and voting DDR and bus frequency (unit: msec)	10	Period of decision window
qcom,core-dev-table	▪ A mapping table of core frequency to a required bandwidth vote at the given core frequency ▪ Each entry consists of one pair of CPU frequency and IB ▪ This table is hard coded in a device tree file (tunables are subject to change)	**Memlat-cpu0** < 300000 762 > < 748800 1720 > < 1132800 2086 > < 1440000 2929 > < 1593600 3879 > **Memlat-cpu4** < 300000 762 > < 499200 1720> < 806400 2086 > < 1036800 2929 > < 1190400 3879 > < 1574400 4943 > < 1728000 5931 > < 1958400 6881 > **L3-cpu0** < 300000 300000000 > < 748800 576000000 > < 979200 652800000 > < 1209600 806400000 > < 1516800 883200000 > < 1593600 960000000 > < 1708800 1305600000 > **L3-cpu4** < 300000 300000000 > < 1036800 576000000 > < 1190400 806400000 > < 1574400 883200000 > < 1804800 960000000 > < 1958400 1305600000 >	Modify kernel device tree entry “qcom,core-dev-table”

Qualcomm 2018-07-23 23:[illegible] dongpeng2@[illegible]

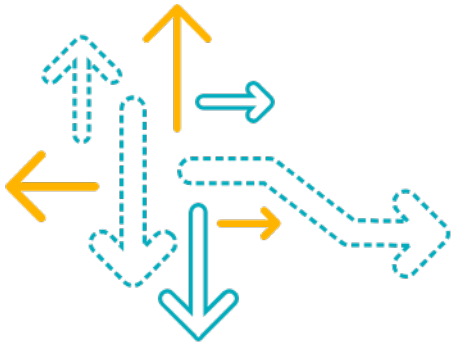
Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION

References

Title	Number
Qualcomm Technologies, Inc.	
SDM845 RPM Hardening Overview and Debug	80-P9301-16

Acronym or term	Definition
AB	Aggregated bandwidth
DCVS	Dynamic clock and voltage scaling
DDR	Double data rate
IB	Instantaneous bandwidth

Qualcomm 2018-05-23 23:41:57 PDT



Questions?

https://createpoint.qti.qualcomm.com

Qualcomm
2018-07-23 23:41:57 PDT
songpeng2@huaqin.com

 Confidential and Proprietary – Qualcomm Technologies, Inc. | MAY CONTAIN U.S. AND INTERNATIONAL EXPORT CONTROLLED INFORMATION