Testing the Latest PCIe 5.0 Power Excursion and PCIe 6.0 L0p Power States

Kyle McRobert

Hardware Engineering Manager

Objectives

- Brief introduction to Power Excursions
- How to measure Power Excursions
- Brief introduction to L0p
- How to measure power usage of L0p and how much power you could save

Power Excursion Introduction

Power Excursion – A temporary condition in which power exceeds maximum sustained power.

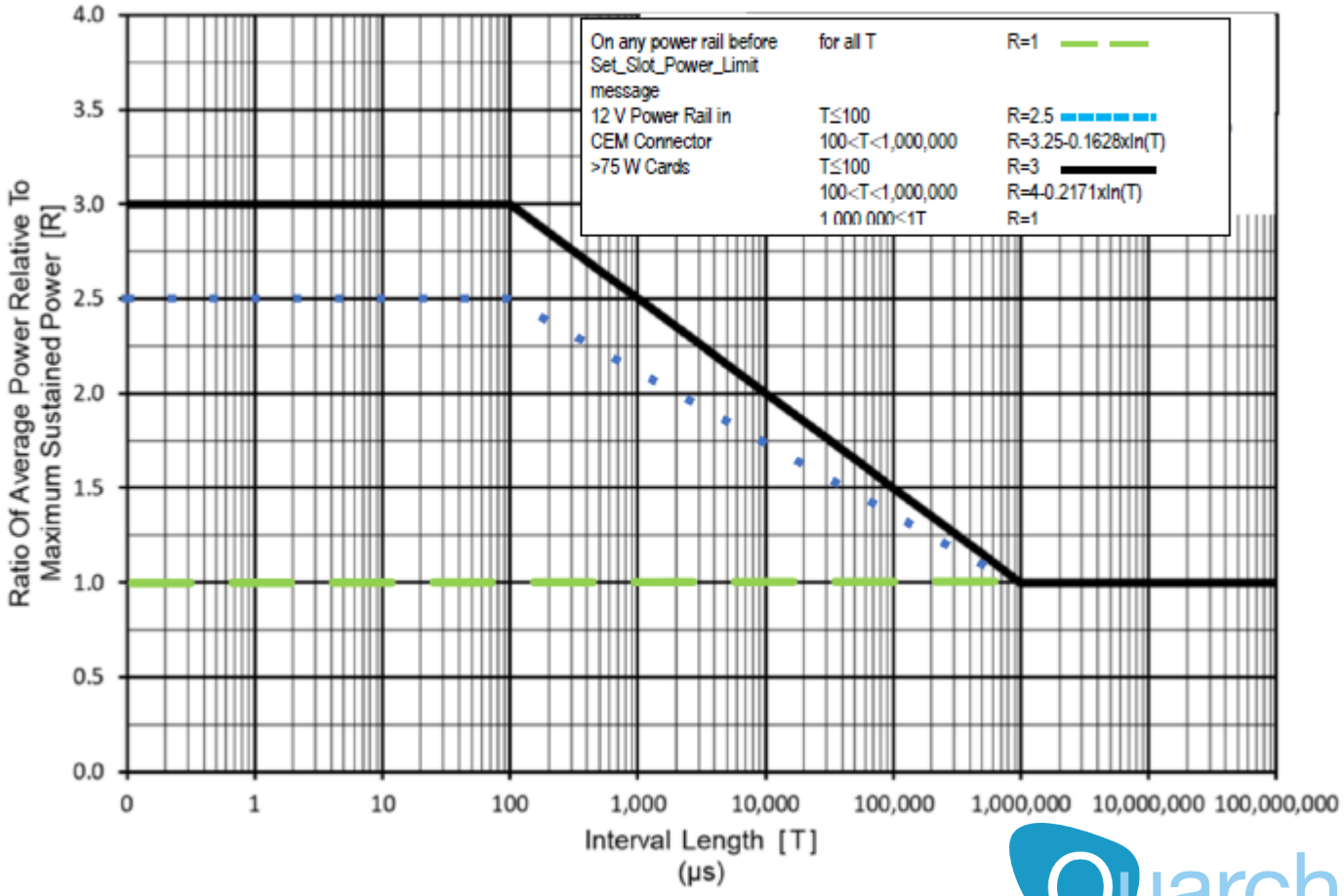
PCIe Add-in Cards that don't rely on aux power connector or use the newer 12V-2x6 or 48HPWR connectors.

This may be needed for high performance GPU, accelerator cards etc. where they need can experience spikes in power consumption during certain operations.

Only after Set_Slot_Power_Limit message received.

System must conform to power supply rail requirements.

Average Power Calculation Interval Length in microseconds (T)	Ratio of Average Power In Interval T Divided by Maximum Sustained Power (R) max		
	On Any Power Rail Before Set_Slot_Power_Limit Message	+12V Power Rail in CEM Connector for all Add-In Cards	Total Card Power for > 75 W Cards
≤100	1	2.5	3
>100 and <1,000,000		3.25 – 0.1628 x ln(T)	4 – 0.2171 x ln(T)
≥1,000,000	1		



How to Test Power Excursions

Must be able to look at all power rails on the add-in card. This includes Power, Current and Voltage for each rail.

Be able to observe power usage over periods long enough to capture the full length of excursion/ catch the excursion happening.

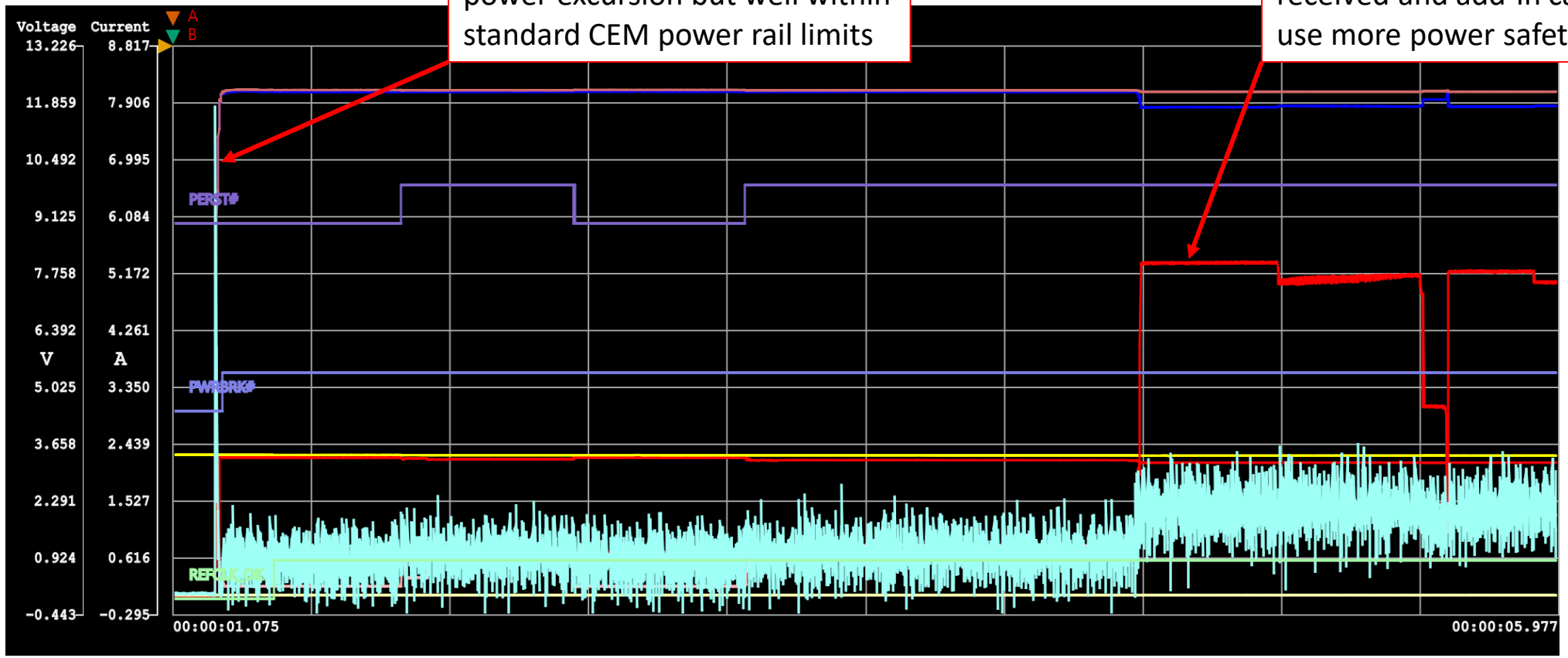
Use a software tool that can run workloads that will cause power excursions.

Be able to look at power up of the device under test to ensure it remains within specification.

Start-up Sequencing and Power Draw

In-rush current not covered by power excursion but well within standard CEM power rail limits

Set Power Limit Message received and add-in card can use more power safety



Analyzing Power Data

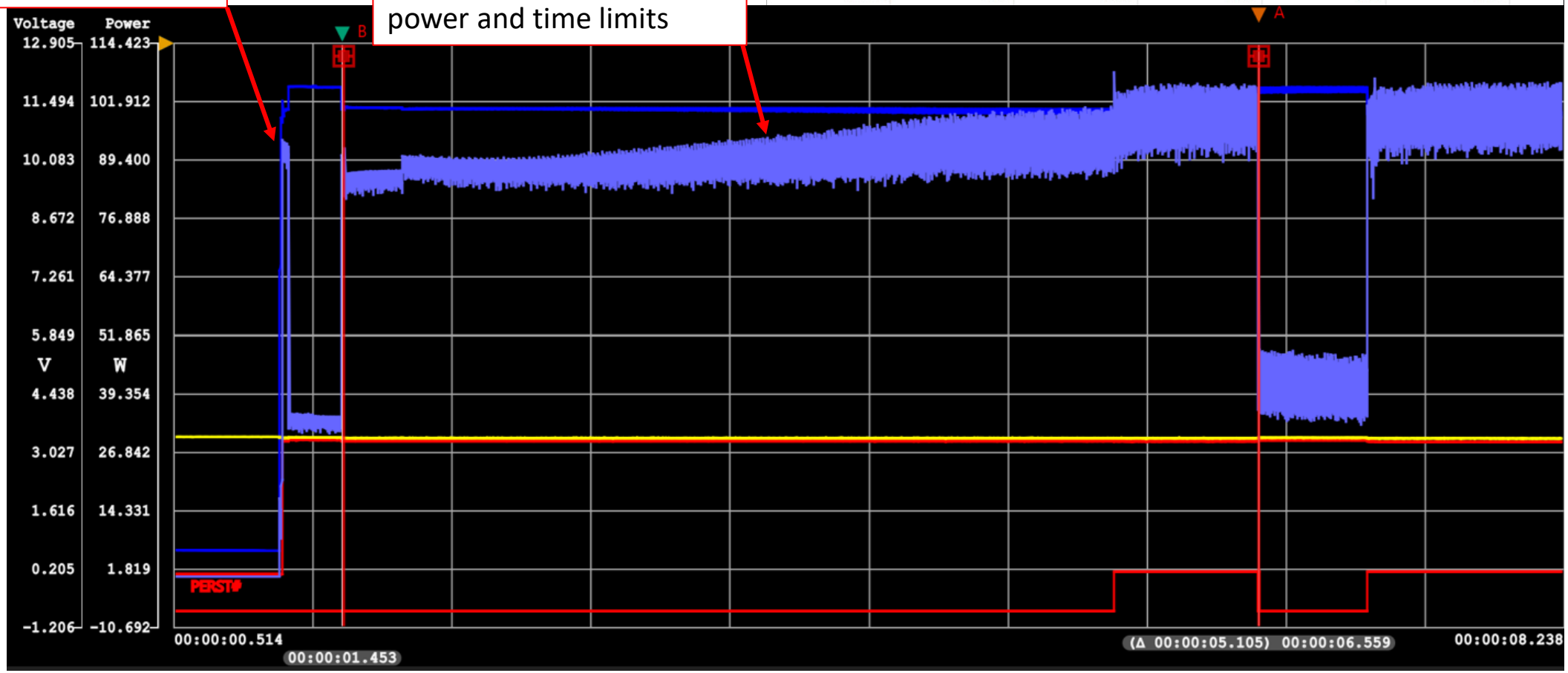
Too large an inrush before set power limit received.

Large constant power consumption violating power and time limits

Statistics

File Edit

Region	Voltage +12V	Voltage +12V	Current +12V	Current +12V	Power +12V	Pow +12
Duration S	Min V	Max V	Max A	RMS A	RMS W	Min
1.458	0.63	11.90	8.487	2.00	23.00	
5.088	11.14	11.86	9.733	8.10	91.59	35,7
45.681	2.30	11.88	9.617	8.67	97.43	



L0p Introduction

Low Power substate to reduce power consumption without interrupting data flow.

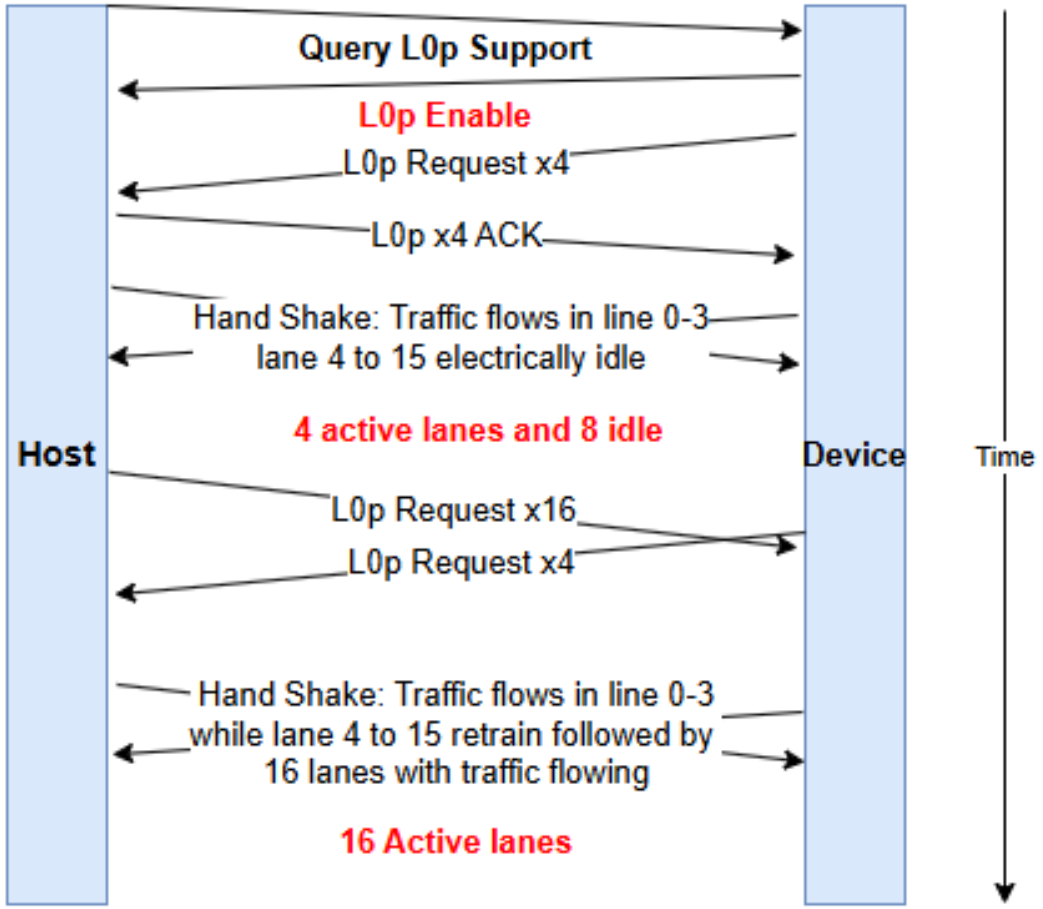
L0p allows for the link to dynamically adjust its bandwidth by reducing the number of active lanes.

Optional and exclusively available for Flit mode and L0s is not supported in Flit mode.

Both TX and RX sides of the link scale their lane usage together.

L0p can be initiated by either side with the other side ACK or NAK.

Does not require a full link retraining process



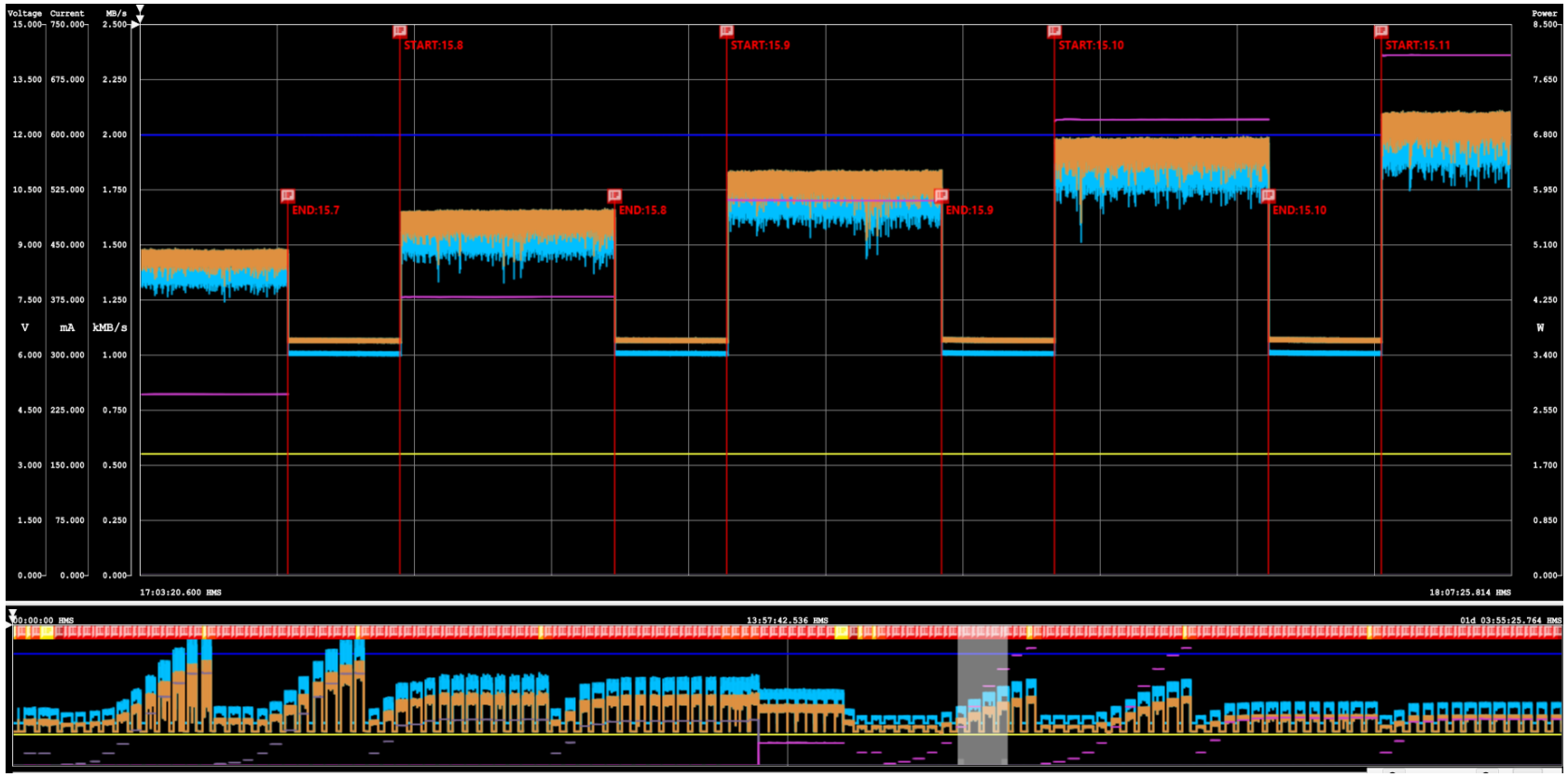
How to test L0p Performance and Power Savings

Must be able to look at all power rails on the device. This includes Power, Current and Voltage for each rail.

Be able to observe power usage over periods long enough to capture the full length of the workload being carried out.

Use a software tool that can run different workloads and able to know what lane width L0p is in.

Run similar work loads without L0p enabled to be able to gauge how much power savings it achieve while not losing on performance.



Potential Power Saving

Work loads ran on Gen5 with drive force into x1, x2 and x4 due to limited Gen6 capable systems.

Idle less relevant for L0p as L0p is aim at power saving while traffic is flowing.

Shows some possible power savings.

Gen 5 U.2 through Gen5 switch on Gen 3 host																
Random	Average IOPS			Average MB/s			Average Power (W)			Average (MB/s)/Watt			Efficency findings from		Idle power (W)		
Reads	x1	x2	x4	x1	x2	x4	x1	x2	x4	x1	x2	x4	Average (MB/s)/Watt		x1	x2	x4
4k 128m	12200	12211	12166	47.66	47.7	47.52	5.87	5.93	6.03	8.12	8.03	7.87	x1 3.2% more efficent than x4		5.8	5.86	5.96
16k 128m	10866	11229	11137	169.79	175.46	174.03	6.13	6.21	6.3	27.68	28.26	27.6	x2 2.4% more efficent than x4				
16k 256m	10881	11231	11219	170.02	175.49	175.3	6.13	6.21	6.33	27.7	28.26	27.68	x2 2.1% more efficent than x4				
32k 256m	9648	10019	10243	301.5	313.12	320.1	6.4	6.49	6.62	47.09	48.24	48.29					

Gen 5 E3 on Gen5 Host																	
Random	Average IOPS			Average MB/s			Average Power (W)			Average (MB/s)/Watt			Efficency findings from		Idle power (W)		
Reads	x1	x2	x4	x1	x2	x4	x1	x2	x4	x1	x2	x4	Average (MB/s)/Watt		x1	x2	x4
4k 128m	118352	128301	128466	462.32	501.18	501.82	4.26	4.99	5.11	108.46	100.4	98.12	x1 10.5% more efficent than x4		3.78	3.88	4
16k 128m	89968	112669	112761	1405.75	1760.46	1761.9	4.7	4.8	5.09	298.65	367.07	346.29	x2 6% more efficent than x4				
32k 128m	83607	90020	94065	326.59	351.64	367.44	4.65	4.62	5.08	70.25	76.05	72.31	x2 5.2% more efficent than x4				
64k 128m	66825	69201	71375	1044.14	1081.27	1115.24	4.94	4.96	5.08	211.36	217.89	219.53	x4 most efficent				
16k 256m	83607	92346	94065	321.12	431.67	433.85	4.38	4.81	5.05	73.35	89.65	85.94	x2 4.3% more efficent than x4				
32k 256m	66825	69201	71375	1182.6	1437.91	1454.71	4.31	4.62	4.9	274.48	311.2	296.67	x2 4.9% more efficent than x4				

Thank You