Customer Report

Results of DDR Detective investigation on an early OCP system

Results from a Data Center visit in May of 2013 describe several JEDEC violations found on an early OCP Server. Customer was interested in deploying OCP based servers and wanted more information on the robustness of the design/manufacturing and configuration.

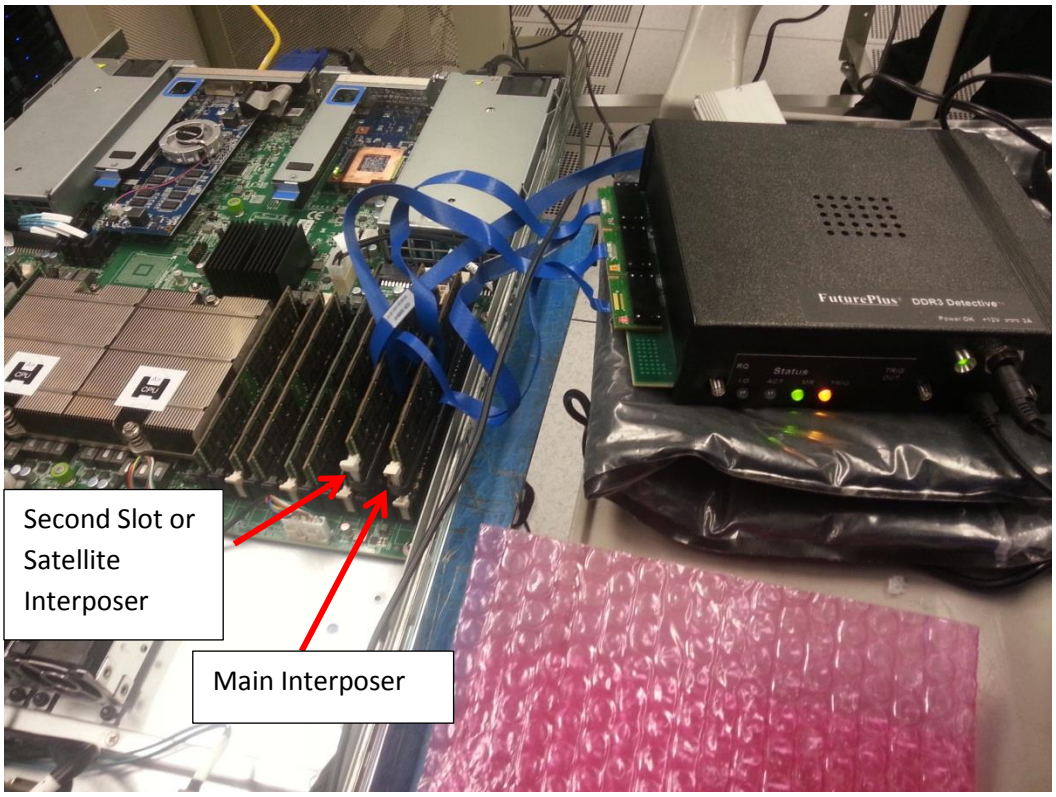
6/1/2013

June 4 2013

Hello [redacted]

Here are the results of my visit down to your [redacted] lab on May 28th 2013. Even though some of the violations were repeats from the ones I found at the OCP workshop I will repeat them here for completeness. As you recall we first ran the DDR3 Detective in the Penguin provided system. It was a AMD Roadrunner with Samsung 1600 16GB registered DIMMS. The motherboard was marked Quanta and the rework was marked A. We placed the DDR3 Detective DIMM interposer in the last slot and we ran the Google Stressapp. We did not see very high data bus utilization and it was below 20% for the time we were monitoring the system. Below you will see the violations we picked up and a trace showing each one. *NOTE: if this system has some kind of custom setup that is non standard to the JEDEC spec we will detect false positives. The tool is very programmable and the user can enter whatever custom setting the manufacturer wants the tool to test against.*

Here is a picture of the tool in the system. We first installed just one interposer in the last slot and then we installed a second interposer in an adjacent slot on the same channel so that we could see if there were any DIMM to DIMM violations. For a robust analysis we would advise repeating these tests for all slots in the system on each channel.



DDR3 Detective installed in the Penguin AMD RR system

Below is the general setup of the DDR3 Detective and the JEDEC parameters that we used. These are from the JEDEC spec and are based on the speed of the bus (we detected 800Mhz clock which is DDR3

1600), the size of the DIMMs and the Speed Bin and latency parameters. The latency parameters are picked up by monitoring the MRS transactions at boot time. The Speed Bin parameters are derived from those numbers and the specification. More details on the speed bin parameters can be seen on page 173 of the DDR3 JEDEC spec. I have checked off the violations we found.

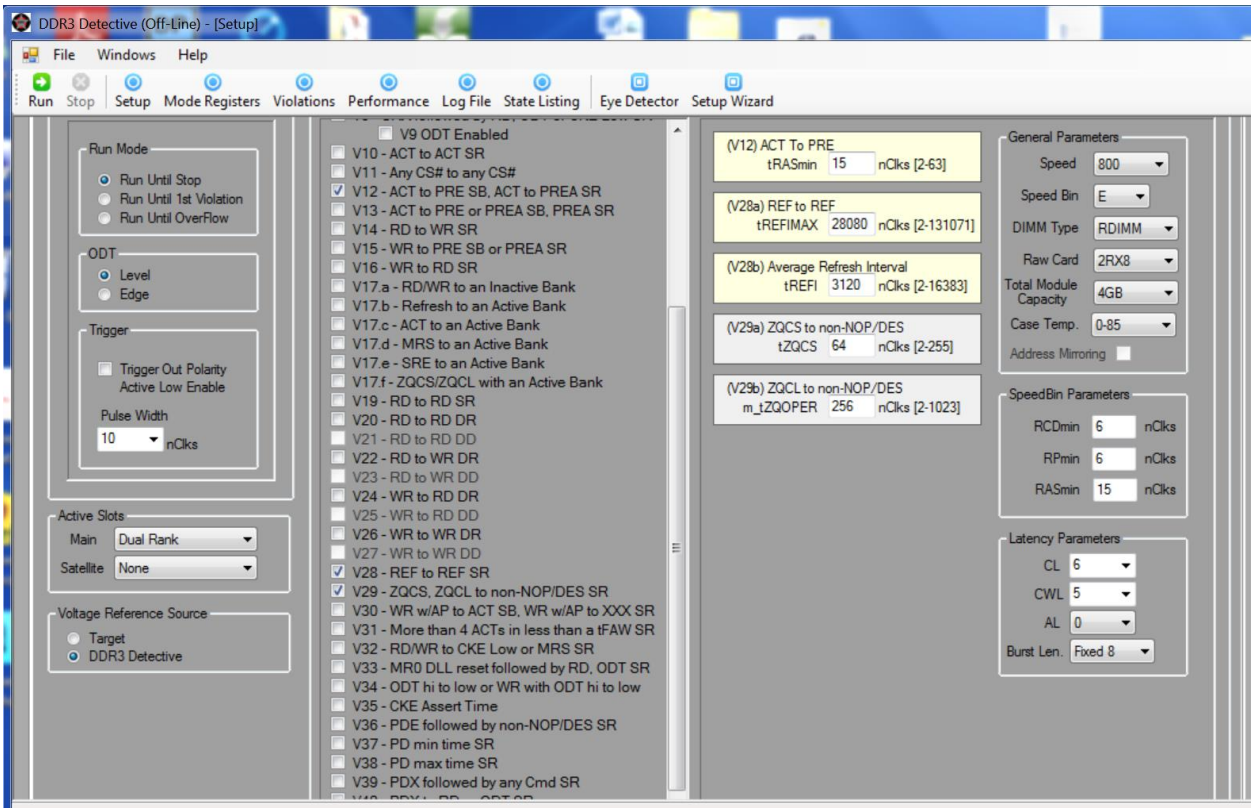
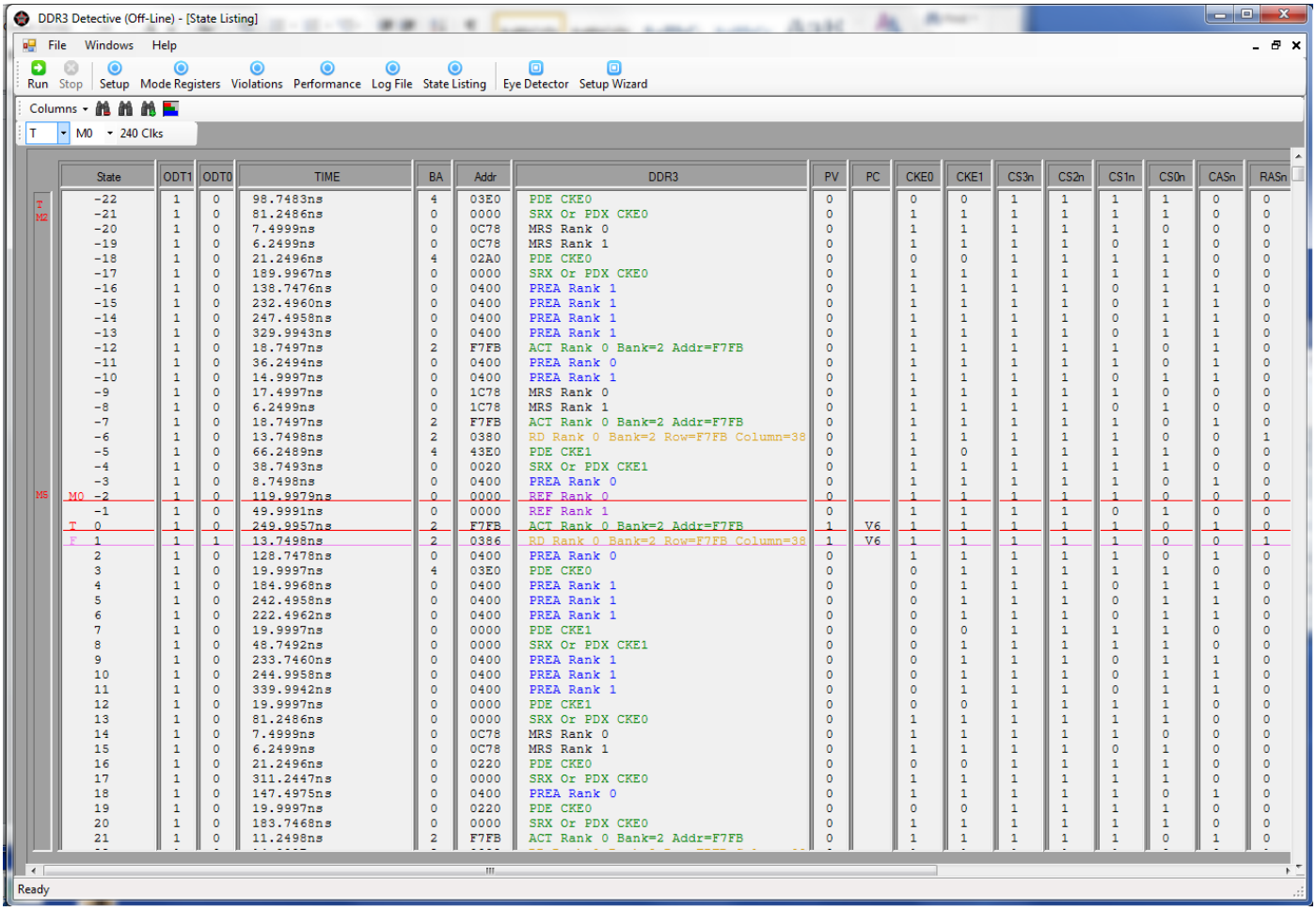


Figure 1: The DDR3 Detective Setup for the AMD OCP RoadRunner System

V6 - Refresh to non-NOP/Deselect including CKE low is less than tRFC

After a Refresh command to a particular Rank the DRAM needs time to recover so the memory controller may not target commands at it for a certain amount of time after a Refresh Command. We saw the system violate this on several occasions with different commands. In the trace we saved we saw it violated on an REF (meaning REF to REF),ACT, RD and MRS. We saw this violation on both Rank 0 and Rank1.

Below is a screen shot of one of these violations.

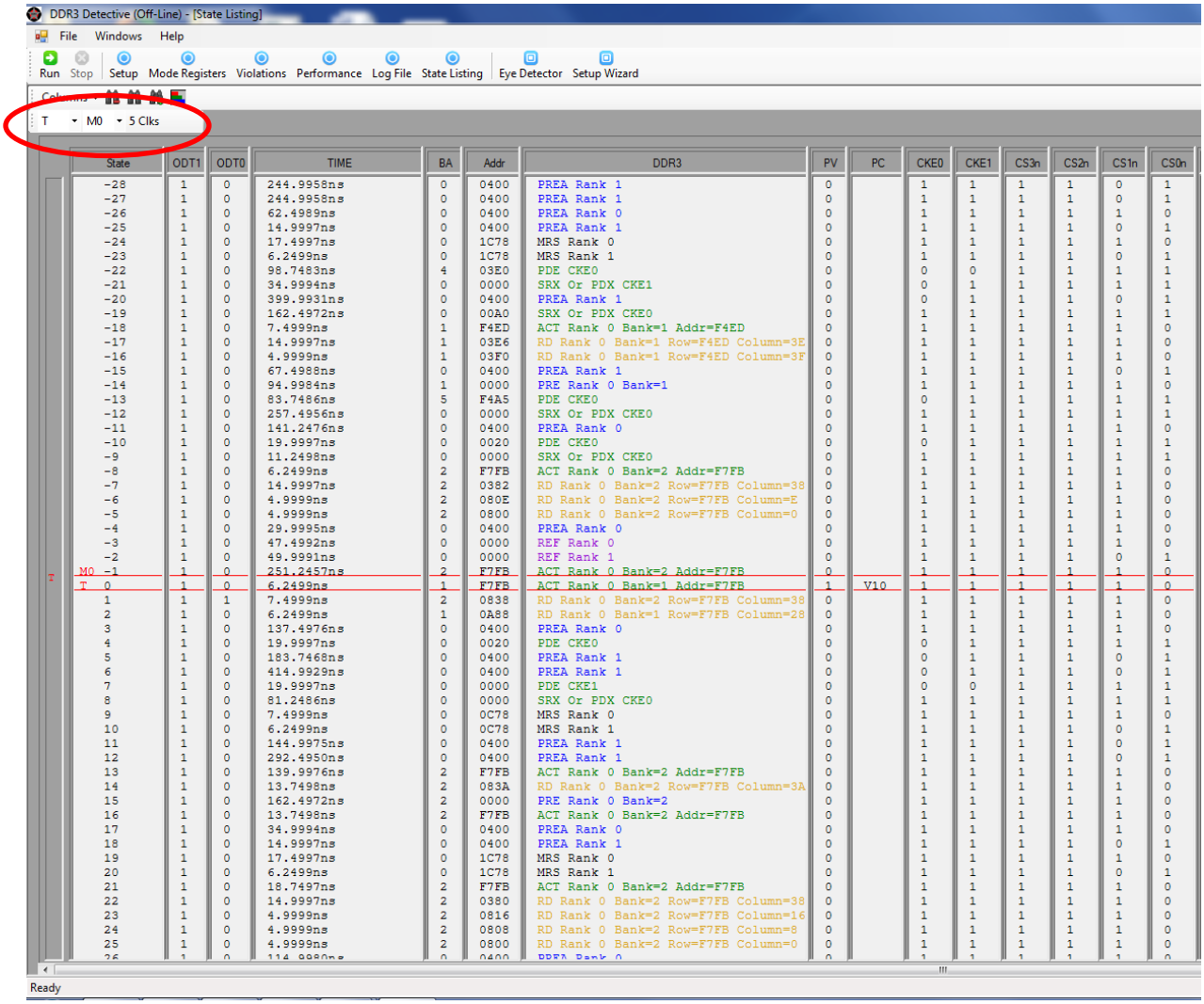


	State	ODT1	ODT0	TIME	BA	Addr	DDR3	PV	PC	CKE0	CKE1	CS3n	CS2n	CS1n	CS0n	CASn	RASn
T	-22	1	0	98.7483ns	4	03E0	PDE CKE0	0		0	0	1	1	1	1	0	0
M2	-21	1	0	81.2486ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1	0	0
	-20	1	0	7.4999ns	0	0C78	MRS Rank 0	0		1	1	1	1	1	0	0	0
	-19	1	0	6.2499ns	0	0C78	MRS Rank 1	0		1	1	1	1	0	1	0	0
	-18	1	0	21.2496ns	4	02A0	PDE CKE0	0		0	0	1	1	1	1	0	0
	-17	1	0	189.9967ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1	0	0
	-16	1	0	138.7476ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1	1	0
	-15	1	0	232.4960ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1	1	0
	-14	1	0	247.4958ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1	1	0
	-13	1	0	329.9943ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1	1	0
	-12	1	0	18.7497ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0	1	0
	-11	1	0	36.2494ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0	1	0
	-10	1	0	14.9997ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1	1	0
	-9	1	0	17.4997ns	0	1C78	MRS Rank 0	0		1	1	1	1	1	0	0	0
	-8	1	0	6.2499ns	0	1C78	MRS Rank 1	0		1	1	1	1	0	1	0	0
	-7	1	0	18.7497ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0	1	0
	-6	1	0	13.7498ns	2	0380	RD Rank 0 Bank=2 Row=F7FB Column=38	0		1	1	1	1	1	0	0	1
	-5	1	0	66.2489ns	4	43E0	PDE CKE1	0		1	0	1	1	1	1	0	0
	-4	1	0	38.7493ns	0	0020	SRX Or PDX CKE1	0		1	1	1	1	1	1	0	0
	-3	1	0	8.7498ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0	1	0
M5	M0 -2	1	0	119.9979ns	0	0000	REF Rank 0	0		1	1	1	1	1	0	0	0
	-1	1	0	49.9991ns	0	0000	REF Rank 1	0		1	1	1	1	0	1	0	0
	T 0	1	0	249.9957ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	1	V6	1	1	1	1	1	0	1	0
	F 1	1	1	13.7498ns	2	0386	RD Rank 0 Bank=2 Row=F7FB Column=38	1	V6	1	1	1	1	1	0	0	1
	2	1	0	128.7478ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0	1	0
	3	1	0	19.9997ns	4	03E0	PDE CKE0	0		0	1	1	1	1	1	0	0
	4	1	0	184.9968ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1	1	0
	5	1	0	242.4958ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1	1	0
	6	1	0	222.4962ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1	1	0
	7	1	0	19.9997ns	0	0000	PDE CKE1	0		0	0	1	1	1	1	0	0
	8	1	0	48.7492ns	0	0000	SRX Or PDX CKE1	0		0	1	1	1	1	1	0	0
	9	1	0	233.7460ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1	1	0
	10	1	0	244.9958ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1	1	0
	11	1	0	339.9942ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1	1	0
	12	1	0	19.9997ns	0	0000	PDE CKE1	0		0	0	1	1	1	1	0	0
	13	1	0	81.2486ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1	0	0
	14	1	0	7.4999ns	0	0C78	MRS Rank 0	0		1	1	1	1	1	0	0	0
	15	1	0	6.2499ns	0	0C78	MRS Rank 1	0		1	1	1	1	0	1	0	0
	16	1	0	21.2496ns	0	0220	PDE CKE0	0		0	0	1	1	1	1	0	0
	17	1	0	311.2447ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1	0	0
	18	1	0	147.4975ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0	1	0
	19	1	0	19.9997ns	0	0220	PDE CKE0	0		0	0	1	1	1	1	0	0
	20	1	0	183.7468ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1	0	0
	21	1	0	11.2498ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0	1	0

Figure x: V6 Refresh too close to an ACT (240 clocks vs 280) and Read (251 vs 280) command Rank 0

V10-ACT to ACT same Rank is less than tRRD

The Activate command opens the bank within a Rank. The JEDEC specification requires that within the same Rank ACTIVATE commands be spaced out property. This system violated it by 1 clock cycle. The specification says 6 and the DDR3 Detective recorded 5. Relatively speaking this was not a frequent violation but occurred on both Ranks of the slot we were monitoring. What are the possible consequences? If the DRAM is not expecting the second ACT to occur so close to the first one it may miss the second ACT and not open the bank. Subsequent Reads and Writes to this bank would then not occur as expected.



	State	ODT1	ODT0	TIME	BA	Addr	DDR3	PV	PC	CKE0	CKE1	CS3n	CS2n	CS1n	CS0n
	-28	1	0	244.9958ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
	-27	1	0	244.9958ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
	-26	1	0	62.4989ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0
	-25	1	0	14.9997ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
	-24	1	0	17.4997ns	0	1C78	MRS Rank 0	0		1	1	1	1	1	0
	-23	1	0	6.2499ns	0	1C78	MRS Rank 1	0		1	1	1	1	0	1
	-22	1	0	98.7483ns	4	03E0	PDE CKE0	0		0	0	1	1	1	1
	-21	1	0	34.9994ns	0	0000	SRX Or PDX CKE1	0		0	1	1	1	1	1
	-20	1	0	399.9931ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1
	-19	1	0	162.4972ns	0	00A0	SRX Or PDX CKE0	0		1	1	1	1	1	1
	-18	1	0	7.4999ns	1	F4ED	ACT Rank 0 Bank=1 Addr=F4ED	0		1	1	1	1	1	0
	-17	1	0	14.9997ns	1	03E6	RD Rank 0 Bank=1 Row=F4ED Column=3E	0		1	1	1	1	1	0
	-16	1	0	4.9999ns	1	03F0	RD Rank 0 Bank=1 Row=F4ED Column=3F	0		1	1	1	1	1	0
	-15	1	0	67.4988ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
	-14	1	0	94.9984ns	1	0000	PRE Rank 0 Bank=1	0		1	1	1	1	1	0
	-13	1	0	83.7486ns	5	F4A5	PDE CKE0	0		0	1	1	1	1	1
	-12	1	0	257.4956ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
	-11	1	0	141.2476ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0
	-10	1	0	19.9997ns	0	0020	PDE CKE0	0		0	1	1	1	1	1
	-9	1	0	11.2498ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
	-8	1	0	6.2499ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0
	-7	1	0	14.9997ns	2	0382	RD Rank 0 Bank=2 Row=F7FB Column=38	0		1	1	1	1	1	0
	-6	1	0	4.9999ns	2	080E	RD Rank 0 Bank=2 Row=F7FB Column=E	0		1	1	1	1	1	0
	-5	1	0	4.9999ns	2	0800	RD Rank 0 Bank=2 Row=F7FB Column=0	0		1	1	1	1	1	0
	-4	1	0	29.9995ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0
	-3	1	0	47.4992ns	0	0000	REF Rank 0	0		1	1	1	1	1	0
	-2	1	0	49.9991ns	0	0000	REF Rank 1	0		1	1	1	1	0	1
	M0 -1	1	0	251.2457ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0
T	T 0	1	0	6.2499ns	1	F7FB	ACT Rank 0 Bank=1 Addr=F7FB	1	V10	1	1	1	1	1	0
	1	1	1	7.4999ns	2	0838	RD Rank 0 Bank=2 Row=F7FB Column=38	0		1	1	1	1	1	0
	2	1	0	6.2499ns	1	0A88	RD Rank 0 Bank=1 Row=F7FB Column=28	0		1	1	1	1	1	0
	3	1	0	137.4976ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0
	4	1	0	19.9997ns	0	0020	PDE CKE0	0		0	1	1	1	1	1
	5	1	0	183.7468ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1
	6	1	0	414.9929ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1
	7	1	0	19.9997ns	0	0000	PDE CKE1	0		0	0	1	1	1	1
	8	1	0	81.2486ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
	9	1	0	7.4999ns	0	0C78	MRS Rank 0	0		1	1	1	1	1	0
	10	1	0	6.2499ns	0	0C78	MRS Rank 1	0		1	1	1	1	0	1
	11	1	0	144.9975ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
	12	1	0	292.4950ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
	13	1	0	139.9976ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0
	14	1	0	13.7498ns	2	083A	RD Rank 0 Bank=2 Row=F7FB Column=3A	0		1	1	1	1	1	0
	15	1	0	162.4972ns	2	0000	PRE Rank 0 Bank=2	0		1	1	1	1	1	0
	16	1	0	13.7498ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0
	17	1	0	34.9994ns	0	0400	PREA Rank 0	0		1	1	1	1	1	0
	18	1	0	14.9997ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
	19	1	0	17.4997ns	0	1C78	MRS Rank 0	0		1	1	1	1	1	0
	20	1	0	6.2499ns	0	1C78	MRS Rank 1	0		1	1	1	1	0	1
	21	1	0	18.7497ns	2	F7FB	ACT Rank 0 Bank=2 Addr=F7FB	0		1	1	1	1	1	0
	22	1	0	14.9997ns	2	0380	RD Rank 0 Bank=2 Row=F7FB Column=38	0		1	1	1	1	1	0
	23	1	0	4.9999ns	2	0816	RD Rank 0 Bank=2 Row=F7FB Column=16	0		1	1	1	1	1	0
	24	1	0	4.9999ns	2	0808	RD Rank 0 Bank=2 Row=F7FB Column=8	0		1	1	1	1	1	0
	25	1	0	4.9999ns	2	0800	RD Rank 0 Bank=2 Row=F7FB Column=0	0		1	1	1	1	1	0

Figure y: ACTIVATE to ACTIVATE too close together on the same Rank

V12 - ACTIVATE to a PRECHARGE is less than tRASmin

This is a violation that we saw at the OCP workshop and AMD said was a BIOS setting issue. Here it is again on the system we tested at your data center. This time the numbers are a bit different because this system was running at 1600 and the system at the OCP workshop was running at 800. However the violation is the same, 1 clock cycle.

Based the settings 29 clocks is the closest these two commands can ever appear on the bus. We saw this violated thousands of times. Here is the trace that shows one such violation. Note M2 to T measures 28 clocks and the violation is 29. So these two commands are one clock period too close together.

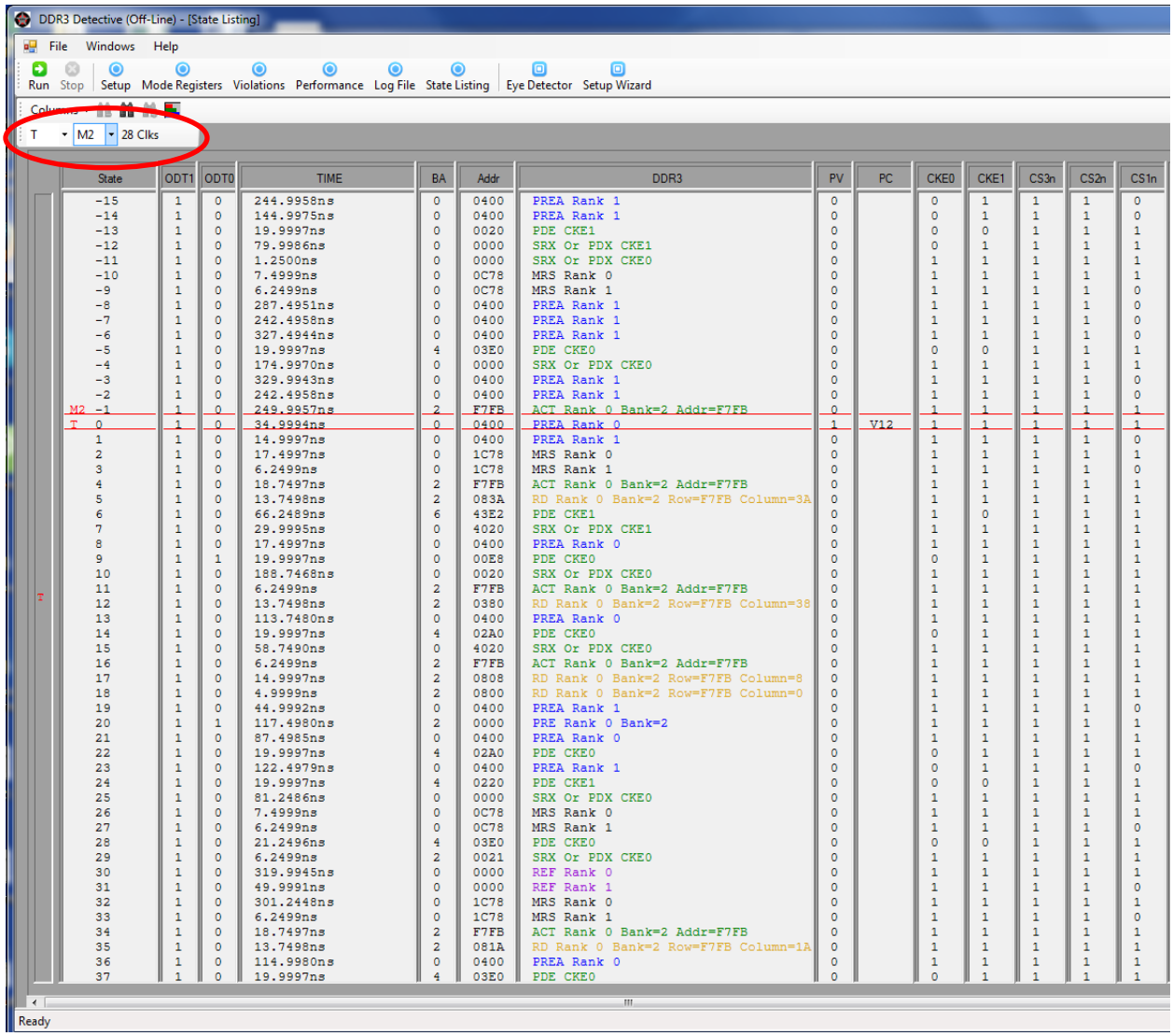


Figure 2: ACT command too close to a PRECHARGE command same bank

The JEDEC specification is a DRAM spec. That means it describes the behavior that the DRAM expects. An ACT command opens a bank and a PRECHARGE command closes it. The DRAM does not expect these

transactions to be 28 clock periods apart. At a bare minimum it expects them to be 29 clocks apart at these speeds. The effect this has on the DRAM is unknown. One hypothesis is that it may not properly close the bank. If that is the case data corruption could occur.

V28b – Average Refresh Interval tREFI

Next up is a Refresh command violation V28b. We have seen this one in the past and I know AMD had some concern about the 128 interval that we chose to do the average over. AMD is correct that 128 as the number of Refresh intervals that we do the average over is not in the specification. In fact no guidance is given in the spec for what number of Refresh intervals to perform the average over. We chose 128 after consulting with a few vendors. It would be great if we had an industry wide standards body that would define this but we don't. We would be happy to take AMD's input on what this number should be. In any event Refresh is important. DRAM stands for *Dynamic* Random Access Memory and as such they need to be Refreshed. If the DRAM is not refreshed properly it will lose state and data corruption will occur. This is tricky to measure as the spec allows for some leeway. What we do is average the number of clocks of the refresh interval over 128 rolling refresh intervals. To be specific we count the clocks between refreshes then over 128 intervals we average then. Every refresh interval we average the last 128 intervals. If the average is greater than the spec we flag a violation. We saw this violation on each of the two ranks of the slot we were in. I captured the failure for Rank 0 and it is shown below. Note the tool has sophisticated hardware filtering so I am able to accurately capture only the refreshes with proper time tags.

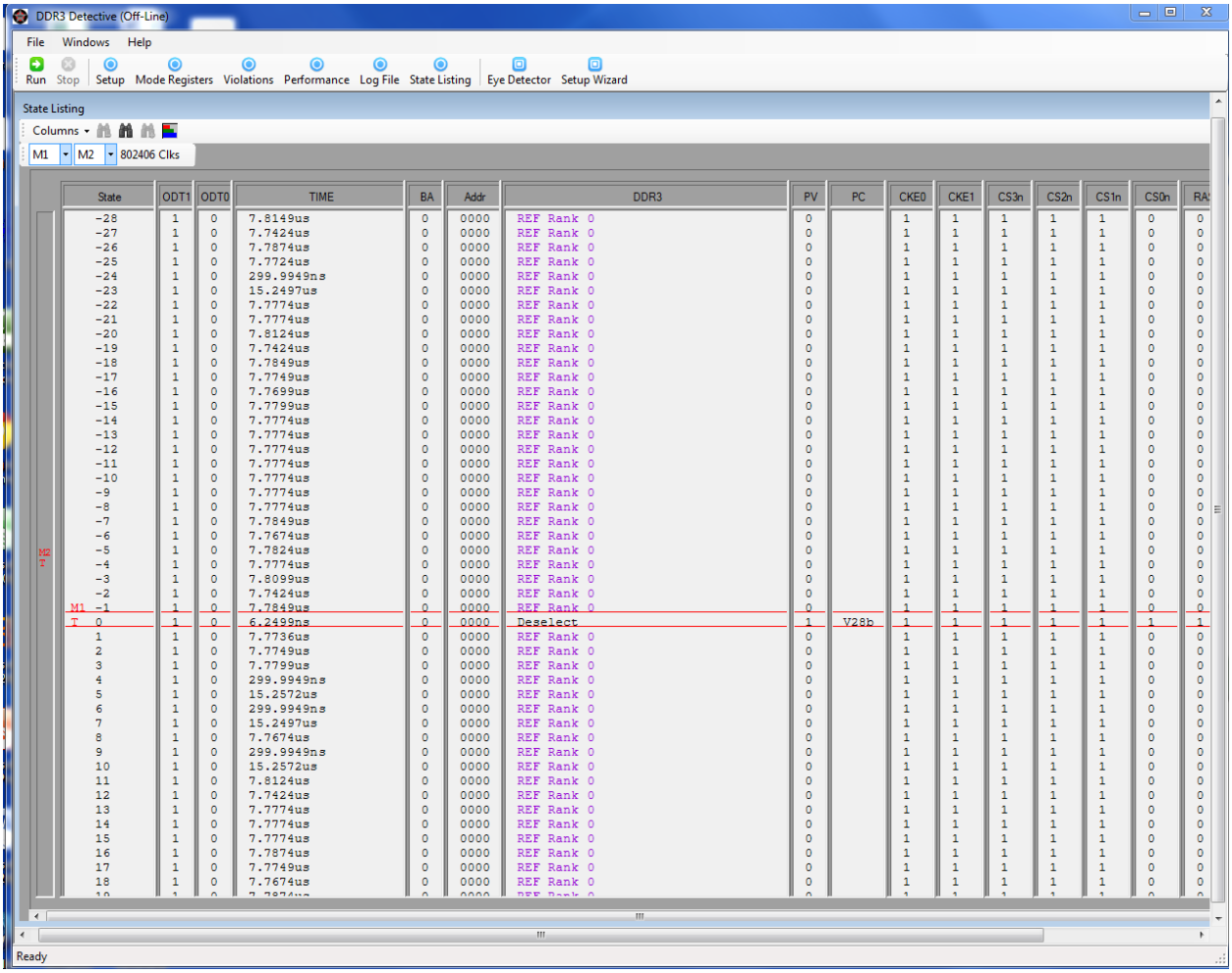
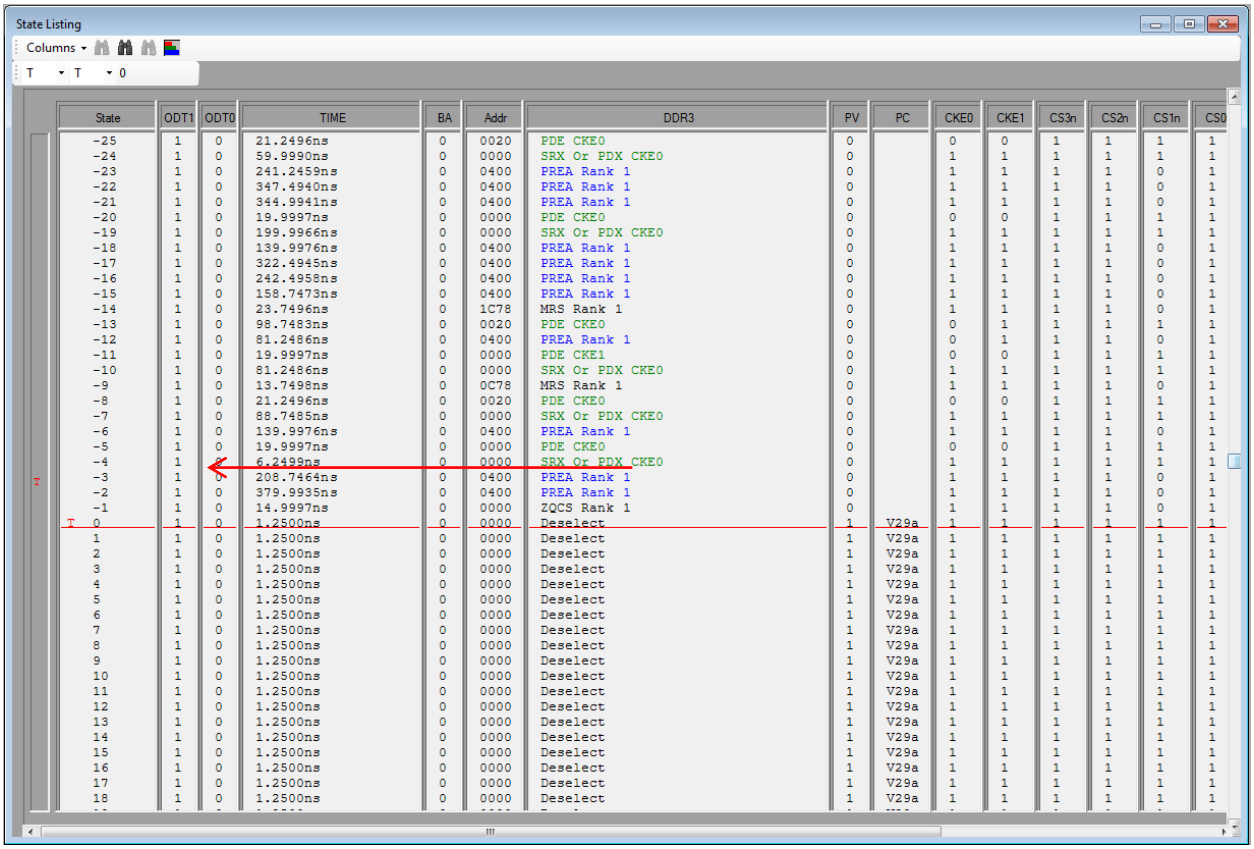


Figure x: Refresh interval should average 6240 clocks but averages 6269 clocks over the last 128 intervals.

I looked at all 128 intervals of this particular trace and I calculated that for this trigger the average worked out to be 6269. If you look back up at Figure 2 for the spec value you will see 6240 clocks. So on average this system was 29 clocks too slow per interval with its refreshes to Rank 0 of this DIMM. In the system we looked at during the OCP workshop it was 8 clocks on average too slow with its refreshes. Given the speed differences the raw time works out to be 36.25ns for this system and 20ns for the system running at 800MT/s. This is an average time per interval too slow on the Refreshes.

This next violation was the same one we saw at the OCP Workshop : Calibrate Command V29a. During operation the system issues ZQCS commands so that the DRAM can recalibrate itself to deal with drift associated with voltage and temperature. During this time period the bus must be quite and ODT must be disabled. What the tool detected was that indeed the bus is quite with only Deselect commands present but ODT for that rank was enabled. According to the spec JESD79-3E section 5.5.2 Note 2: "On-die termination must be disabled via the ODT signal or MRS during the calibration procedure". What is seen is that the ODT signal for rank 1 is enabled.



State	ODT1	ODT0	TIME	BA	Addr	DDR3	PV	PC	CKE0	CKE1	CS3n	CS2n	CS1n	CS0
-25	1	0	21.2496ns	0	0020	PDE CKE0	0		0	0	1	1	1	1
-24	1	0	59.9990ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
-23	1	0	241.2459ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-22	1	0	347.4940ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-21	1	0	344.9941ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-20	1	0	19.9997ns	0	0000	PDE CKE0	0		0	0	1	1	1	1
-19	1	0	199.9966ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
-18	1	0	139.9976ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-17	1	0	322.4945ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-16	1	0	242.4958ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-15	1	0	158.7473ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-14	1	0	23.7496ns	0	1C78	MRS Rank 1	0		1	1	1	1	0	1
-13	1	0	98.7483ns	0	0020	PDE CKE0	0		0	1	1	1	1	1
-12	1	0	81.2486ns	0	0400	PREA Rank 1	0		0	1	1	1	0	1
-11	1	0	19.9997ns	0	0000	PDE CKE1	0		0	0	1	1	1	1
-10	1	0	81.2486ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
-9	1	0	13.7498ns	0	0C78	MRS Rank 1	0		1	1	1	1	0	1
-8	1	0	21.2496ns	0	0020	PDE CKE0	0		0	0	1	1	1	1
-7	1	0	88.7485ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
-6	1	0	139.9976ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-5	1	0	19.9997ns	0	0000	PDE CKE0	0		0	0	1	1	1	1
-4	1	0	6.2499ns	0	0000	SRX Or PDX CKE0	0		1	1	1	1	1	1
-3	1	0	208.7464ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-2	1	0	379.9935ns	0	0400	PREA Rank 1	0		1	1	1	1	0	1
-1	1	0	14.9997ns	0	0000	ZQCS Rank 1	0		1	1	1	1	0	1
0	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
1	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
2	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
3	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
4	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
5	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
6	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
7	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
8	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
9	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
10	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
11	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
12	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
13	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
14	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
15	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
16	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
17	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1
18	1	0	1.2500ns	0	0000	Deselect	1	V29a	1	1	1	1	1	1

Figure 3: ODT1 is enabled during a ZQCS command to Rank 1

AMD's response was This is due to our 3 DIMM per channel configuration. Due to the number of ODT pins, we tie ODT1 on the DIMM high. This will show as spec violation on the analyzer, but is OK because we send MR command to disable RTT_Nom. Indeed the spec would allow for ODT to be ignored if RTT Nom was disabled. However we check for that. We monitor the Mode Register Settings and if RTT Nom was disabled we would have not flagged this violation. However I did not save the MR settings as we acquired them so I have no proof that the MR was not set. In order to further investigate this one I would have to repeat the test and this time save the Mode Register information. I would make it as inconclusive for now.

Below is a picture of the tool, the FS2400 DDR3 Detective®

Figure 4: The DDR3 Detective® with DIMM interposer from FuturePlus Systems