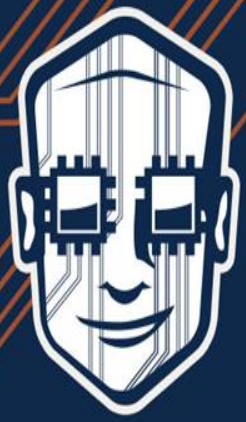
DesignCon® 2015



CRITICAL MEMORY PERFORMANCE METRICS FOR DDR4 SYSTEMS

Barbara Aichinger
Vice President New Business Development
FuturePlus Systems Corporation

UBM Tech

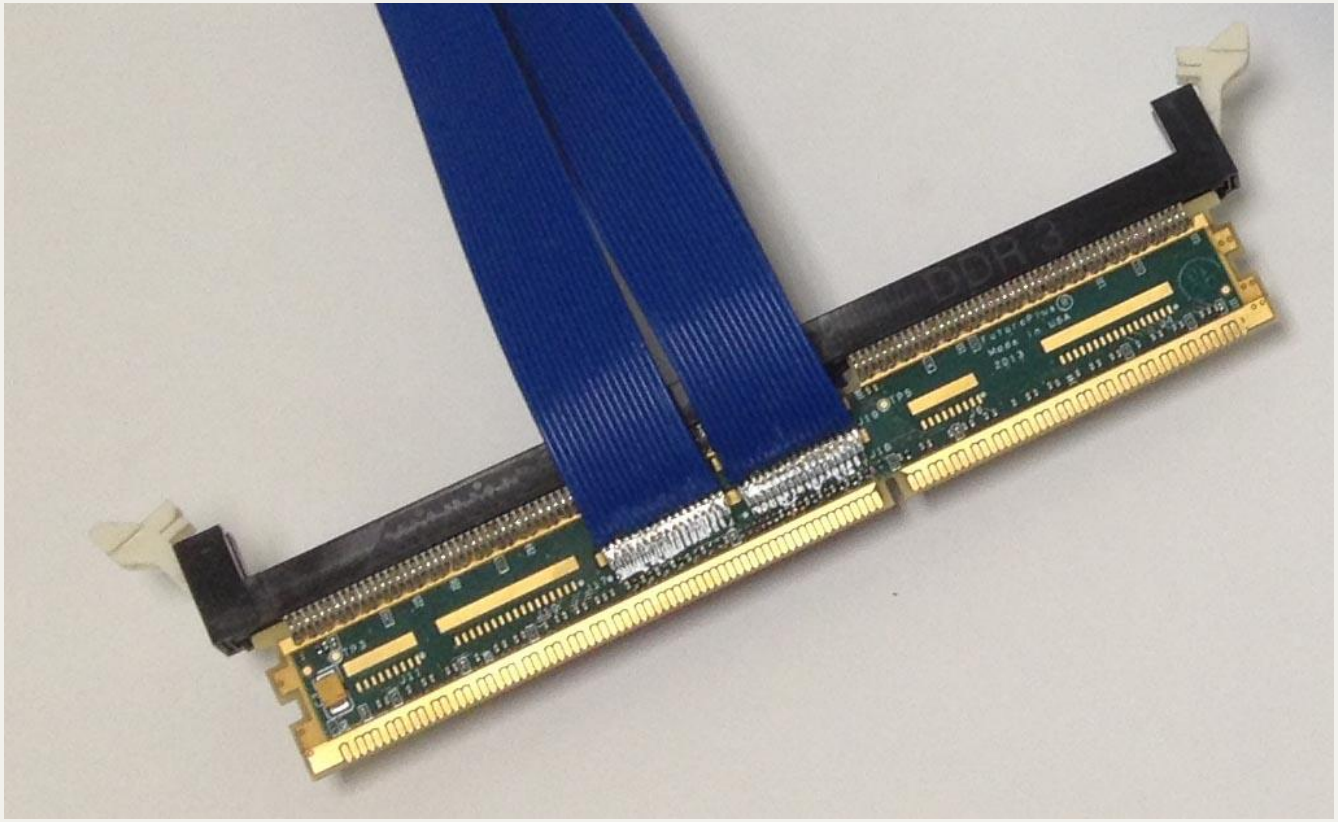
Outline

- How can we monitor the very fast low voltage DDR4 memory without effecting the system?
- Traditional Performance Metrics
  - Bandwidth, Latency and Power Management
- New Performance Metrics
  - Bus Mode Analysis
  - Page Hit Analysis
  - Multiple Open Banks Analysis
  - Bank Utilization Analysis
  - Row Hammer (excessive Activate) Detection
- Measured DDR4 Performance in a real target

How to Monitor the DDR4 Memory

- Use a slot interposer to 'listen' to the traffic between the DIMM and the Memory Controller
  - A small amount of current is 'tapped' off the bus
  - Only the Address, Command and Control bus needs to be monitored
  - Same probing method can be used for SODIMM and 'memory down' using a BGA adapter

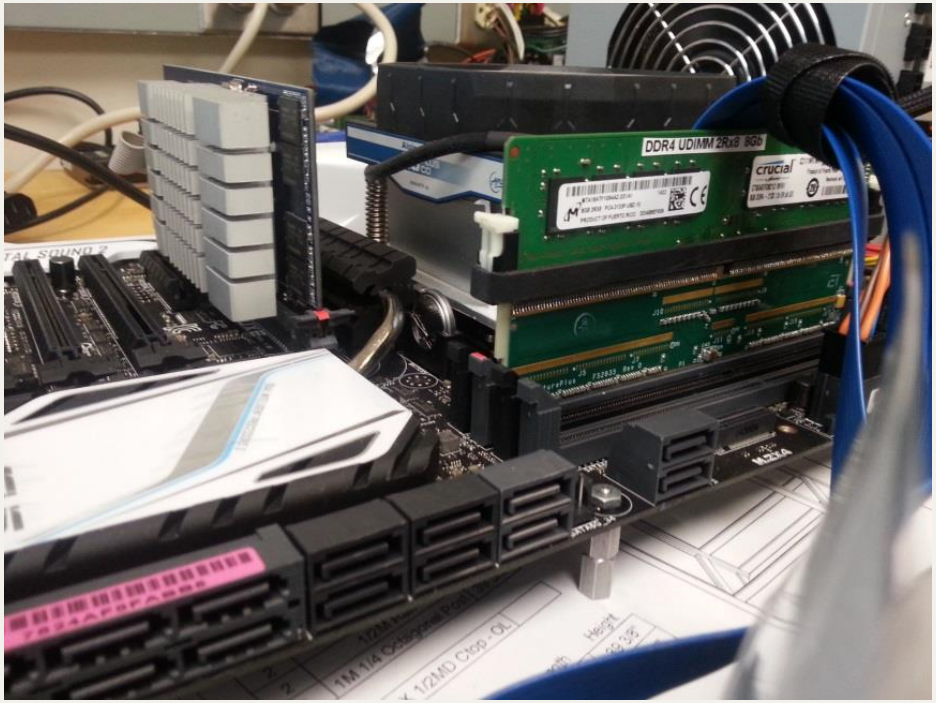
DDR4 DIMM Interposer

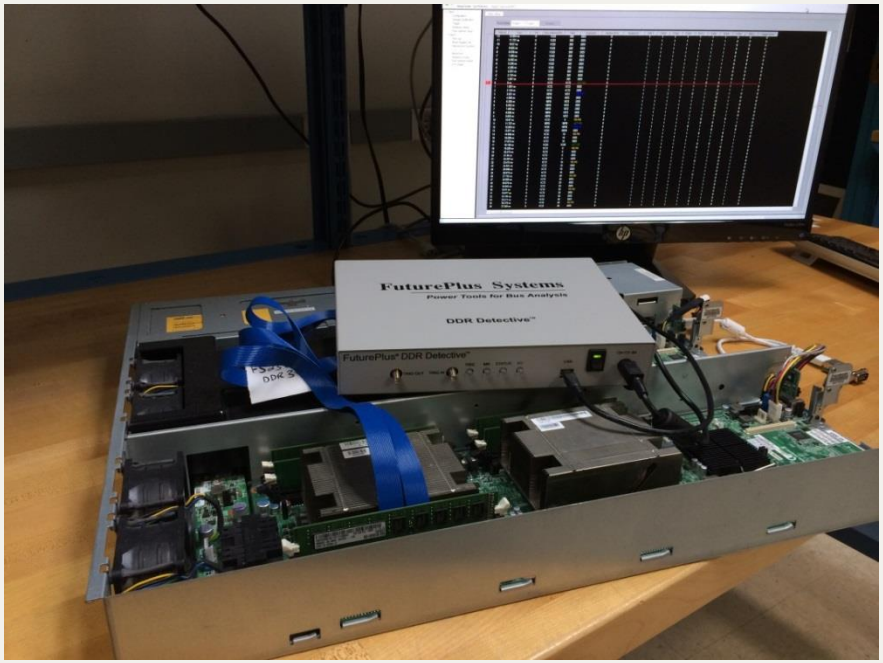


Monitoring the DDR4 Memory

- Highly attenuated signals are then amplified and the bus is ‘replicated’ inside the equipment

The system boots and runs never knowing the equipment is present





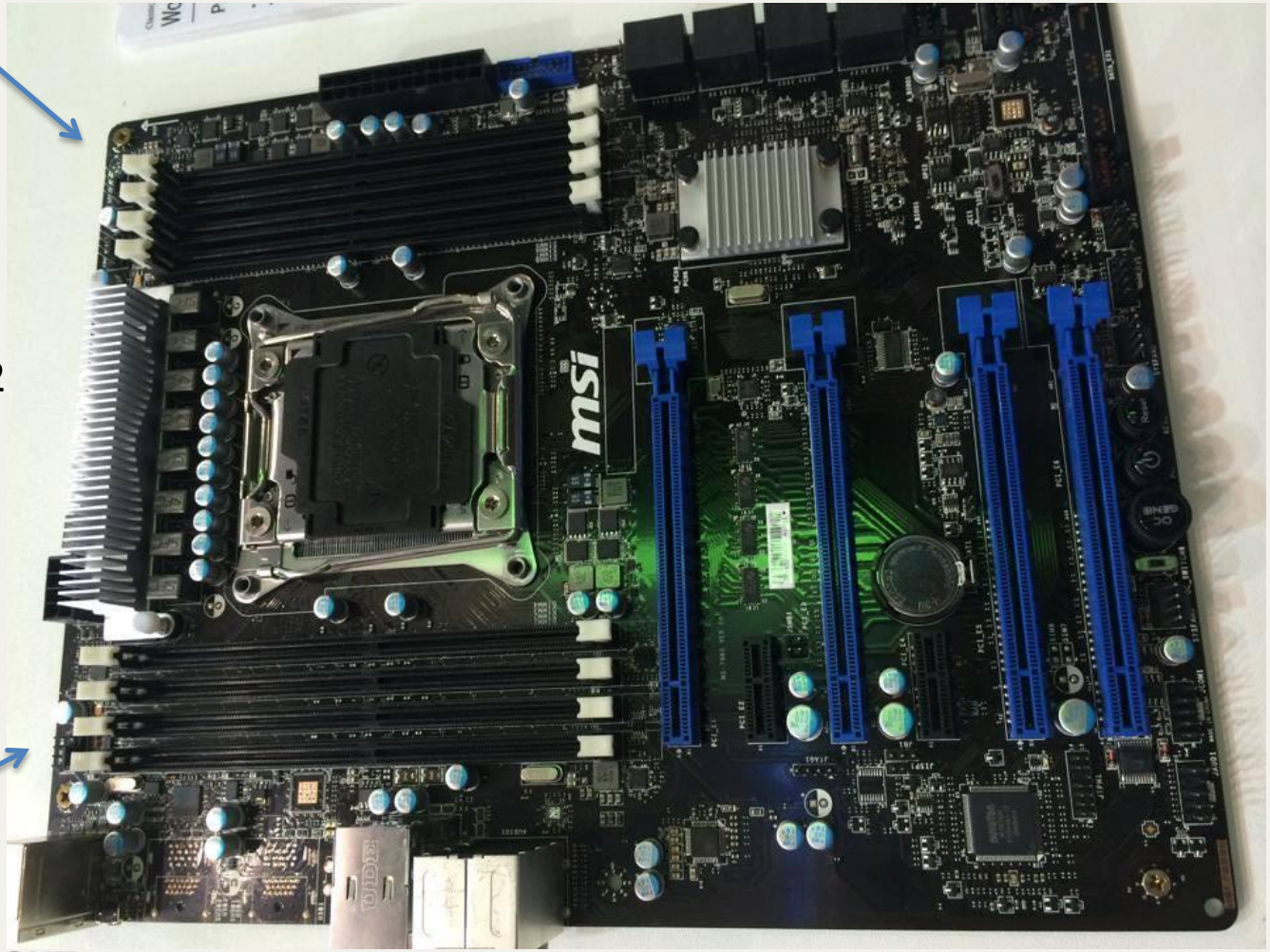
Performance Metrics

- Billions of clock cycles every second
- Old methods of measurement captures small amount of time in large trace buffers
- New method uses sophisticated event counters to:
  - Never miss a clock cycle
  - Count Performance Metrics for long periods of time REAL TIME

Work Smarter not Harder

- For Performance metrics the DDR Detective® uses counters instead the traditional trace memory
  - To capture a second of DDR4 traffic would take 4.5Gbytes of logic analyzer/protocol analyzer trace depth $$$$!
    - 1 hour = 270 Gbytes of trace depth and then time to sift through it and post process!
  - By using large counters and counting events and the time between events we can achieve hours and days worth of metrics with no trace buffer memory and with no time consuming post processing

4 Memory Channels each channel is 2 slots



ASUS X99 DDR4 Motherboard

Traditional Measurements

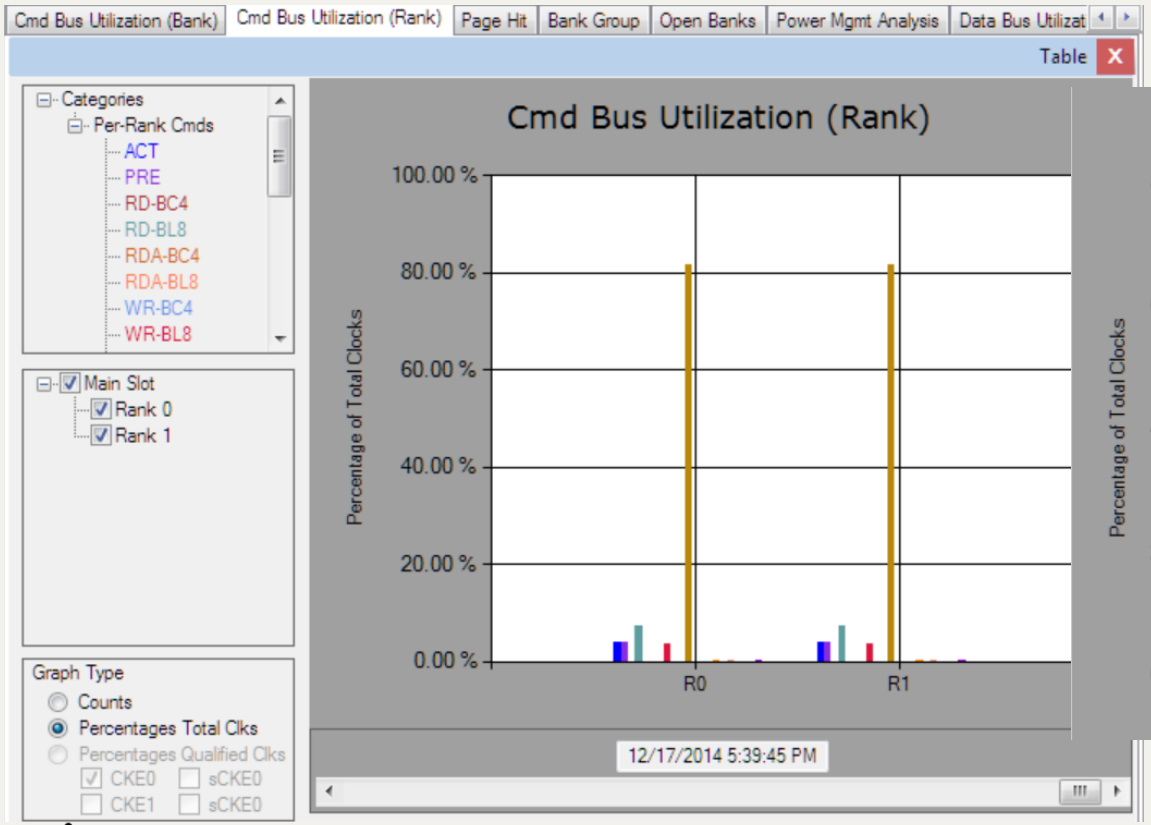
- Bandwidth
  - Command Bus Utilization
  - Data Bus Utilization
- Power Management
- Latency

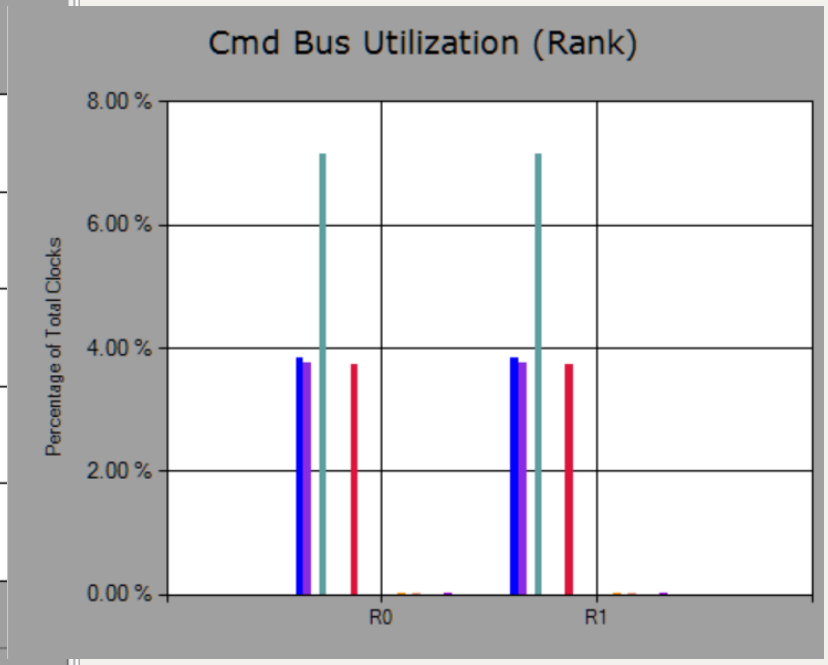
Bandwidth

- Overhead
  - Any use of the bus other than a Read or a Write
  - Command Bus Utilization
- Data Bus
  - Utilization: the % of the time that Read or Write Data is being transferred
  - Bandwidth: the amount of data transferred per second

Command Bus Utilization

2400MT/s

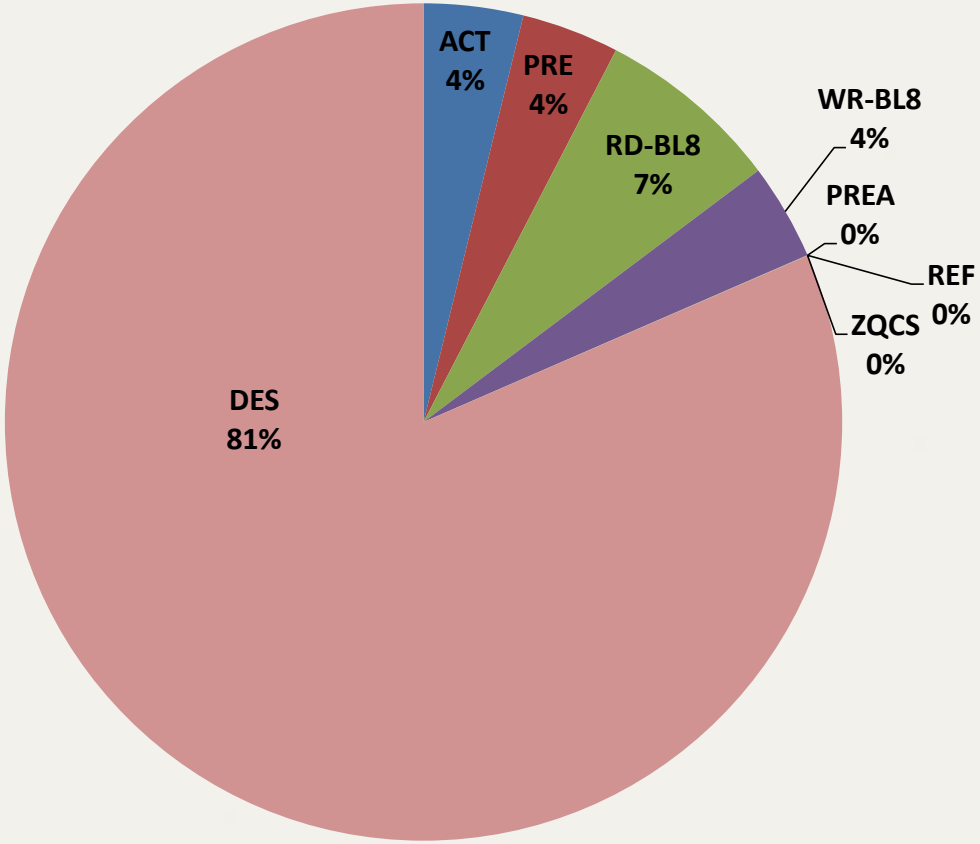




KEYSIGHT TECHNOLOGIES
Strategic Solutions Partner

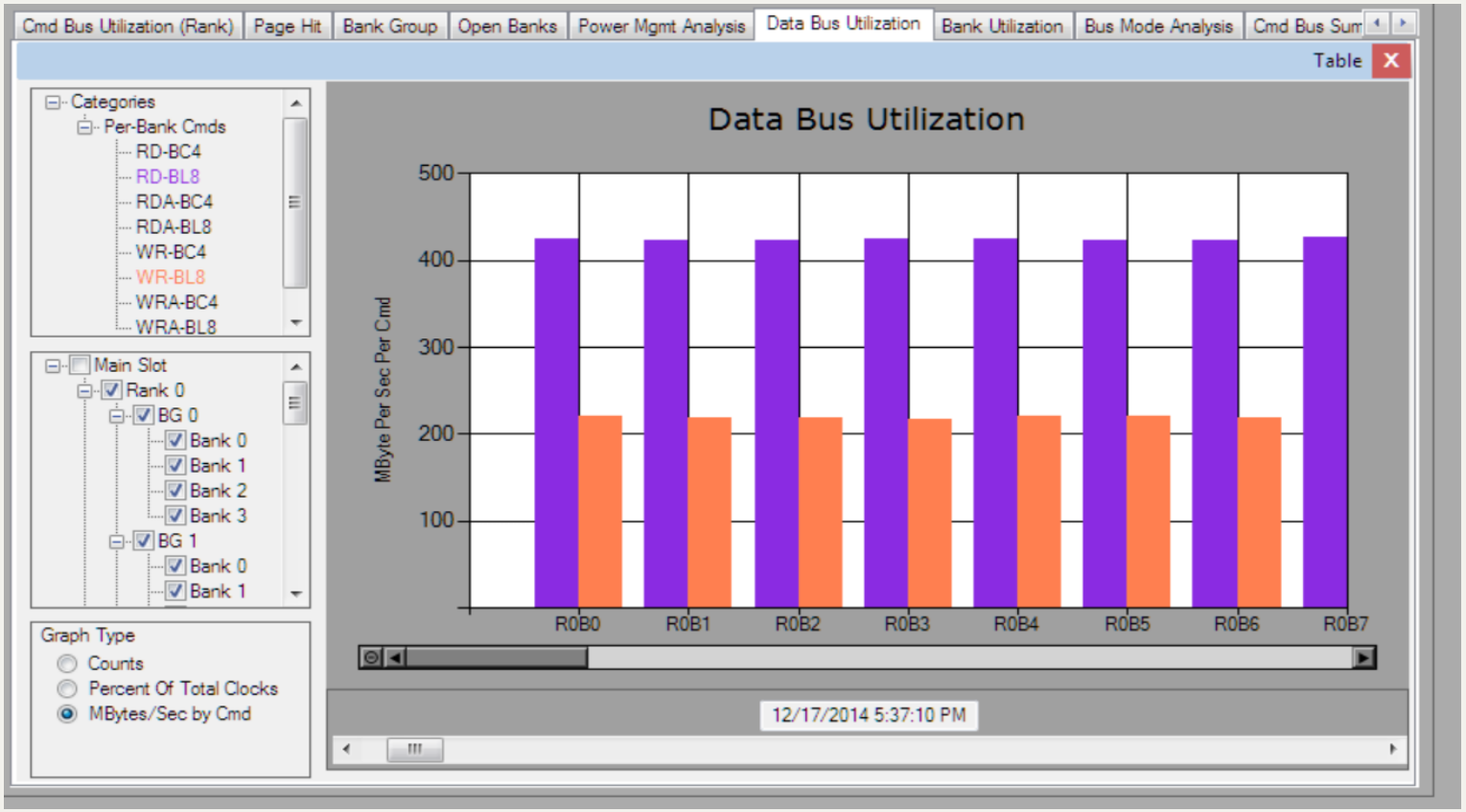
FuturePlus Systems
Power Tools for Bus Analysis

Command Bus Utilization

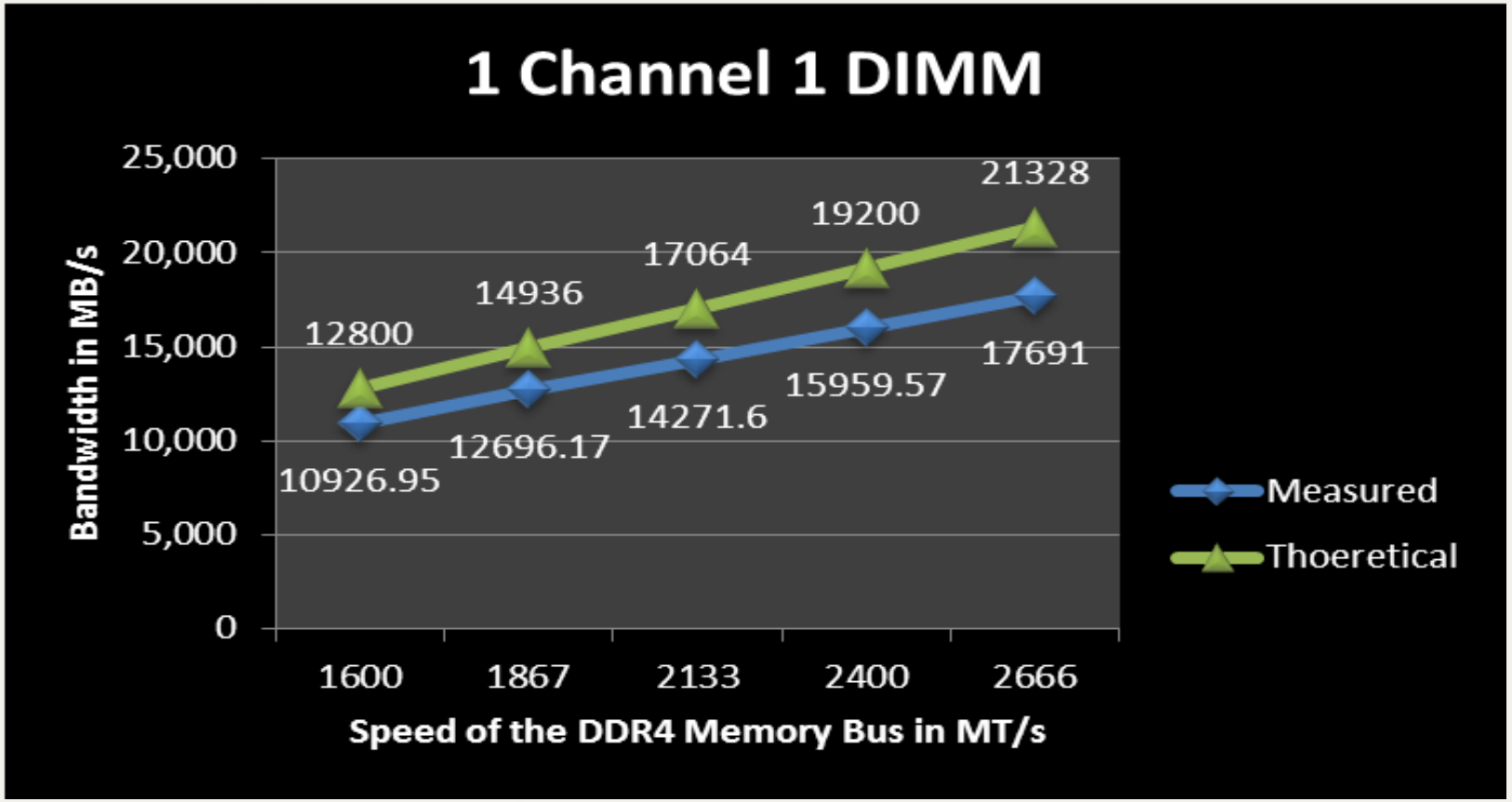


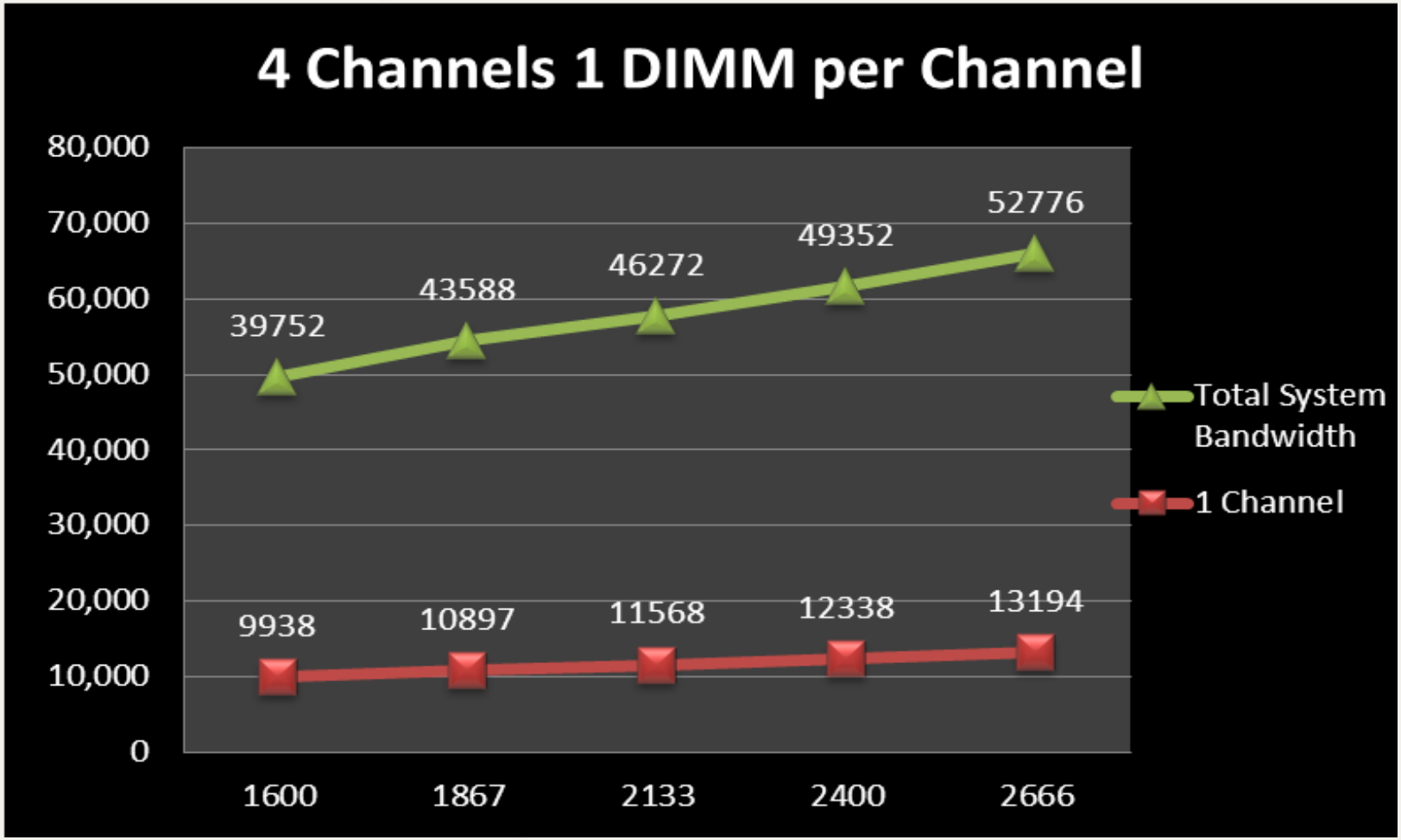
Data Bus Utilization

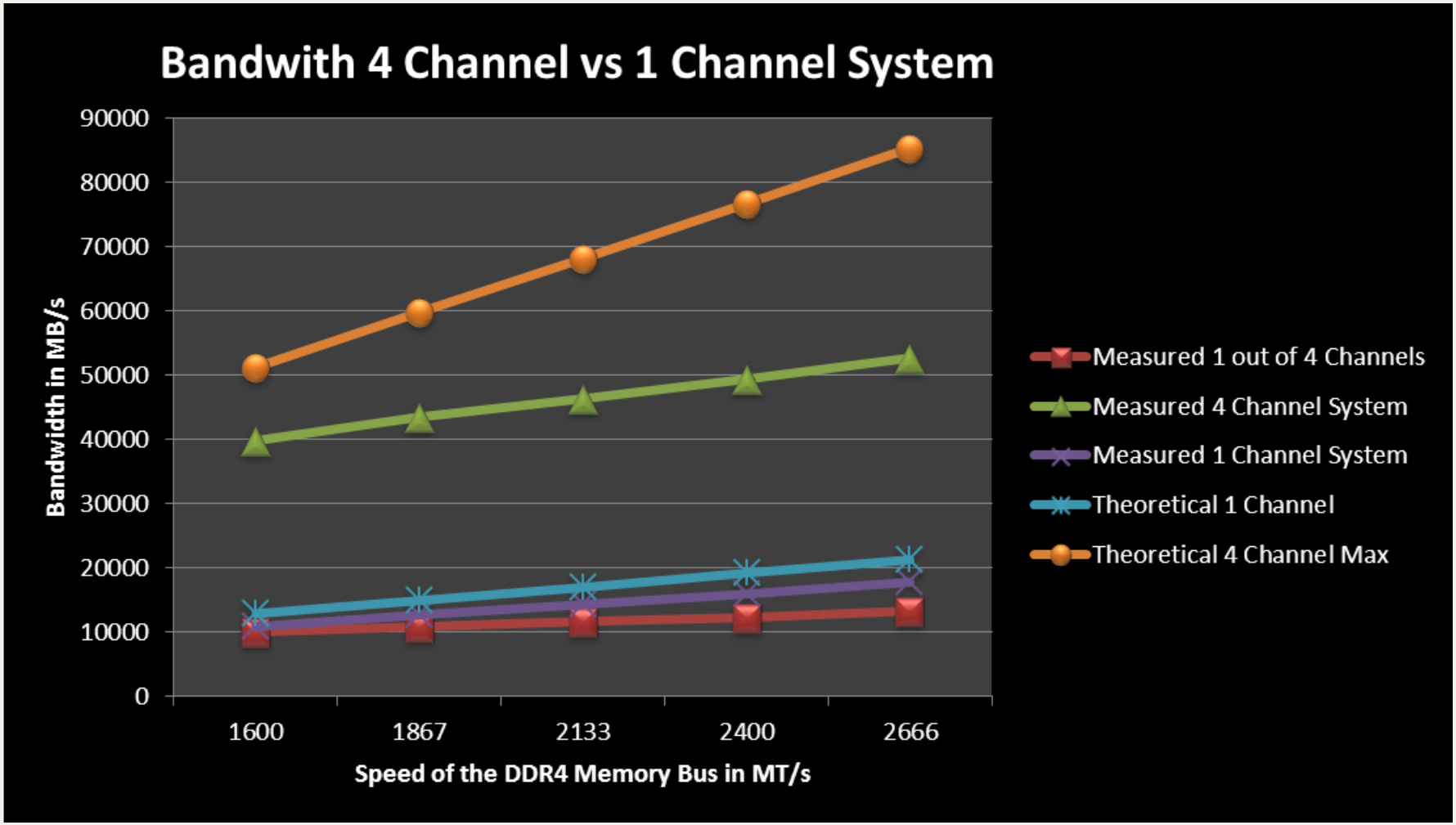
2400MT/s



DDR4 Bandwidth

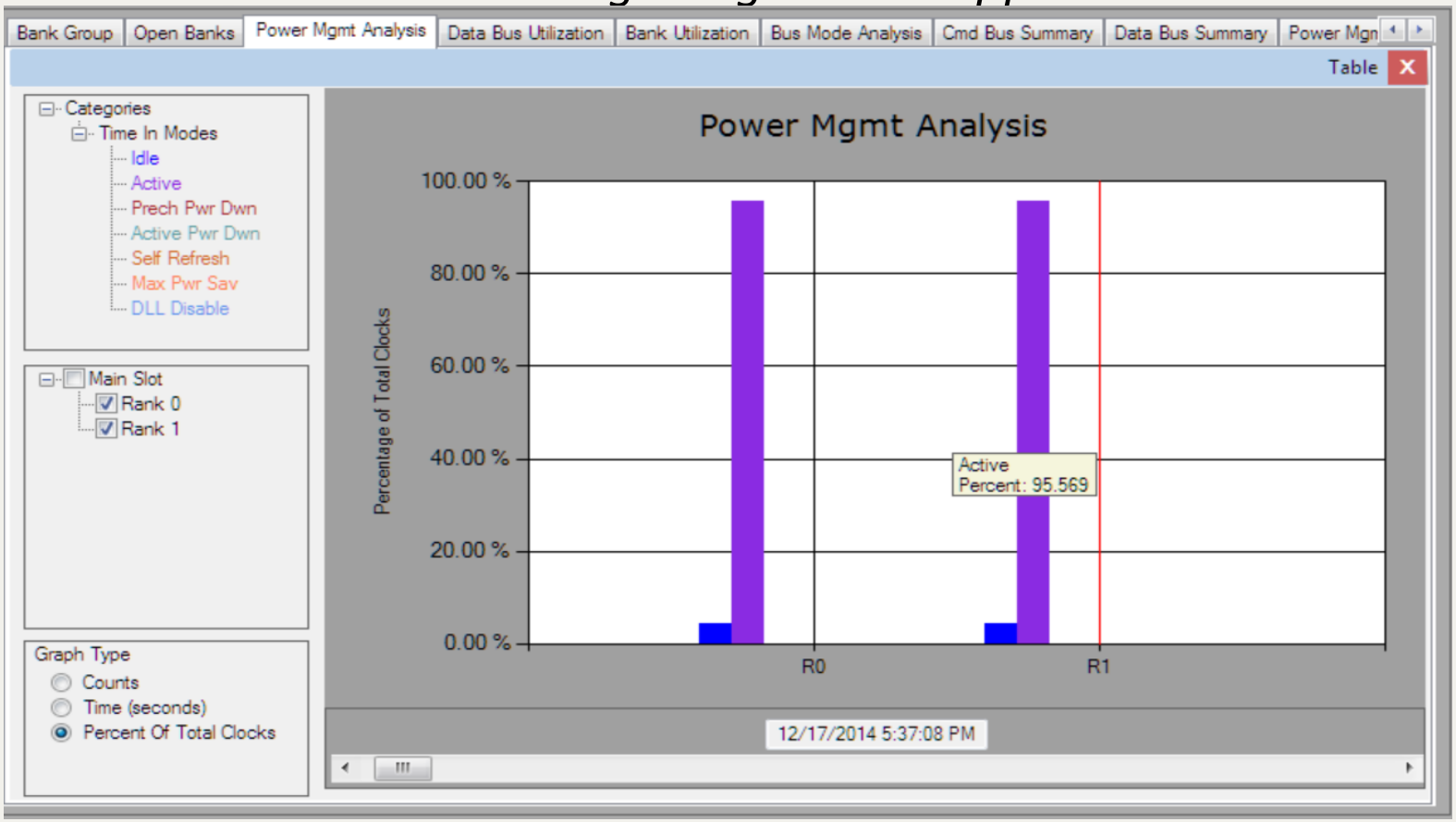



4 Channels 1 DIMM per Channel
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
39752
43588
46272
49352
52776
9938
10897
11568
12338
13194
1600
1867
2133
2400
2666
Total System Bandwidth
1 Channel


Bandwith 4 Channel vs 1 Channel System
Bandwidth in MB/s
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
1600
1867
2133
2400
2666
Speed of the DDR4 Memory Bus in MT/s
Measured 1 out of 4 Channels
Measured 4 Channel System
Measured 1 Channel System
Theoretical 1 Channel
Theoretical 4 Channel Max

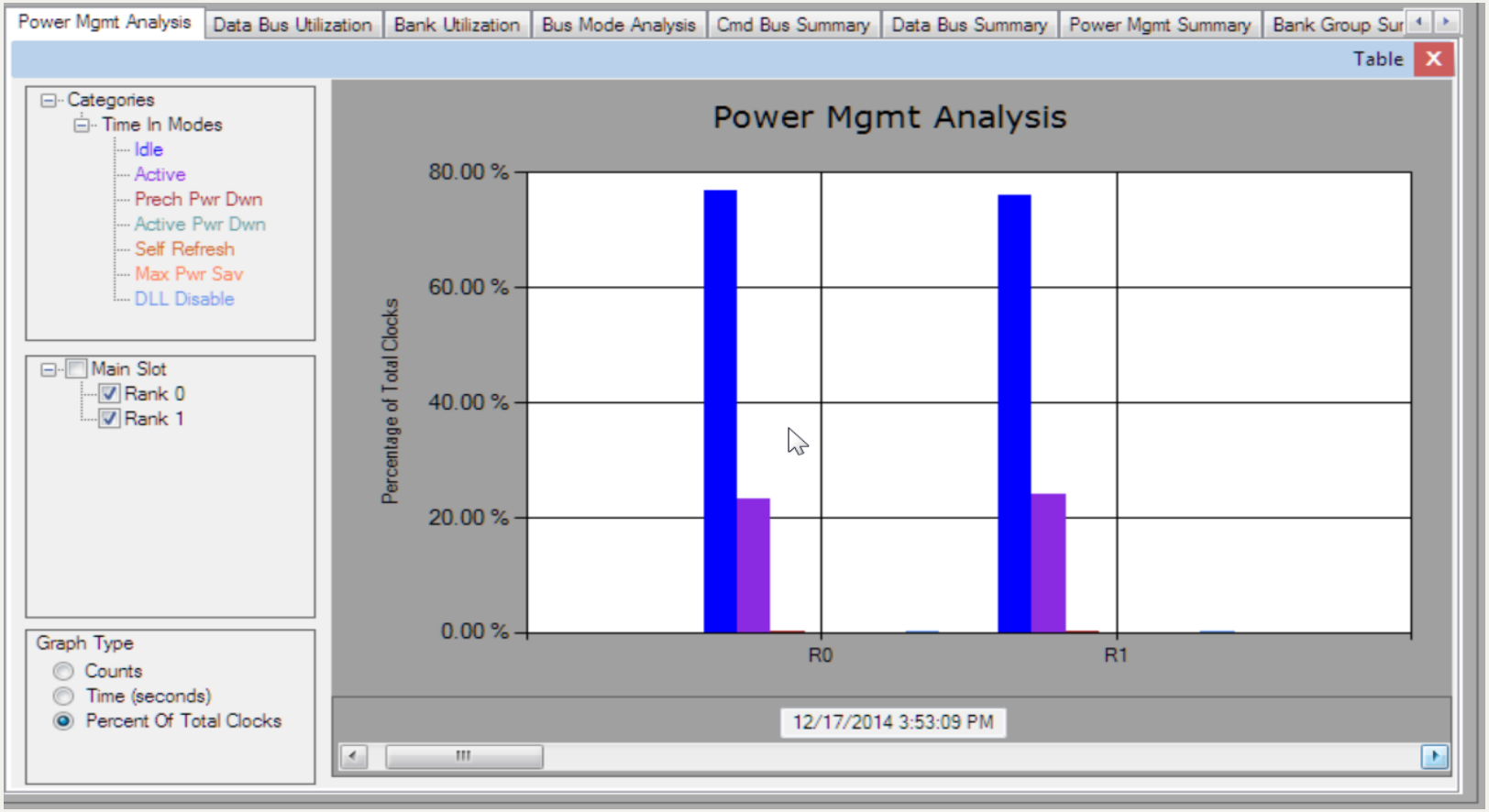
Power Management

while running Google StressApp



Power Management Analysis

Ubuntu Boot



Power Management

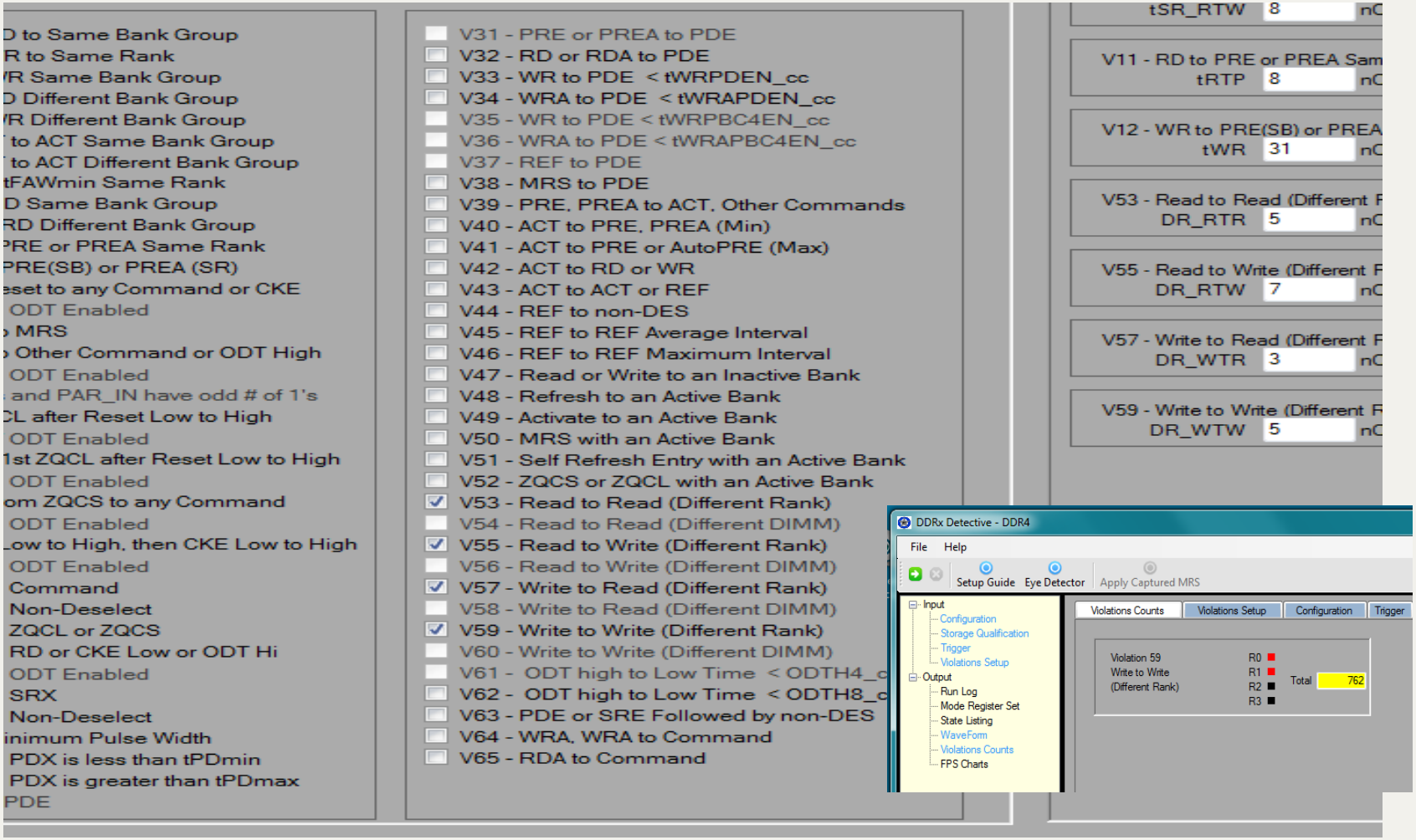
- ~50M servers Servers World Wide
- Each Server averages 16-24 DIMMs
  - 800M to 1.2B DIMMs
- Even small power savings per DIMM can add up

*Every time **Facebook**'s data center engineers figure out a way to **reduce** server consumption **by a single watt**, the improvement, at Facebook's scale, has the potential to add **millions of dollars** to the company's bottom line.*

Yevgeny Sverdlik
Editor in Chief
Data Center Knowledge

Latency

- Several Jedec Paramters apply:
  - RD to WR same rank tSR_RTW
  - RD to PRE/PREA same Rank tRTP
  - WR to PRE(SB) or PREA (SR) tWR
  - Read to Read different Rank tDR_RTR
  - Read to Write different Rank DR_RTW
  - Write to Read different Rank tDR_WTR
  - Write to Write different Rank tDR_WTW



KEYSIGHT TECHNOLOGIES
Strategic Solutions Partner

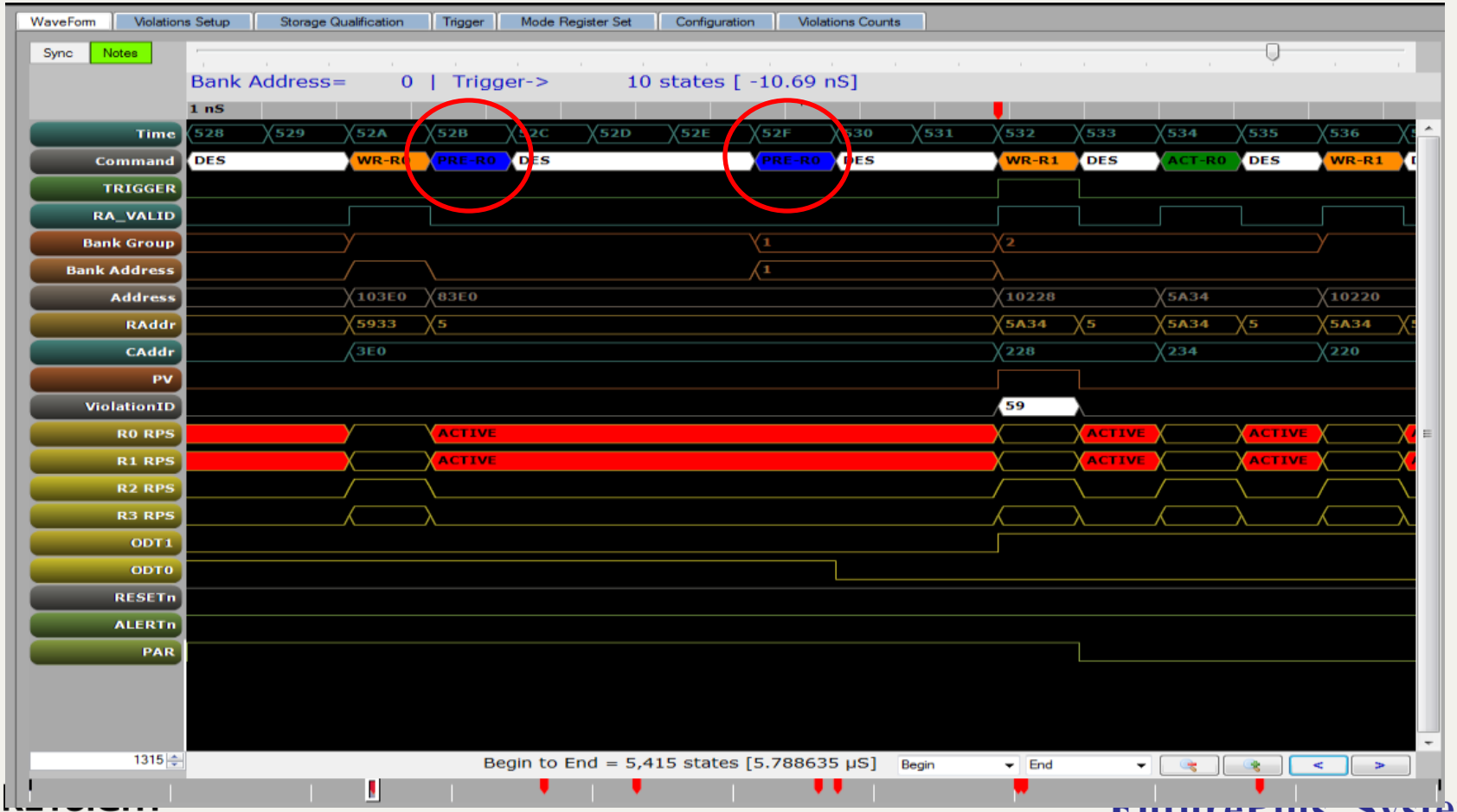
FuturePlus Systems
Power Tools for Bus Analysis

Latency Measurements

measurement made at 1867

V#	Parameter	Description	Spec	Measured
V2	tSR_RTW	RD to WR same Rank	8	10
V11	tRTP	RD to PRE same Rank	8	8
V12	tWR	WR to PRE SB or PREA SR	31	31
V53	tDR_RTR	RD to RD diff Rank	5	6
V57	tDR_WTR	WR to RD diff Rank	3	6
V59	tDR_WTW	WR to WR diff Rank	5	8

Intervening Commands



Latency Measurements

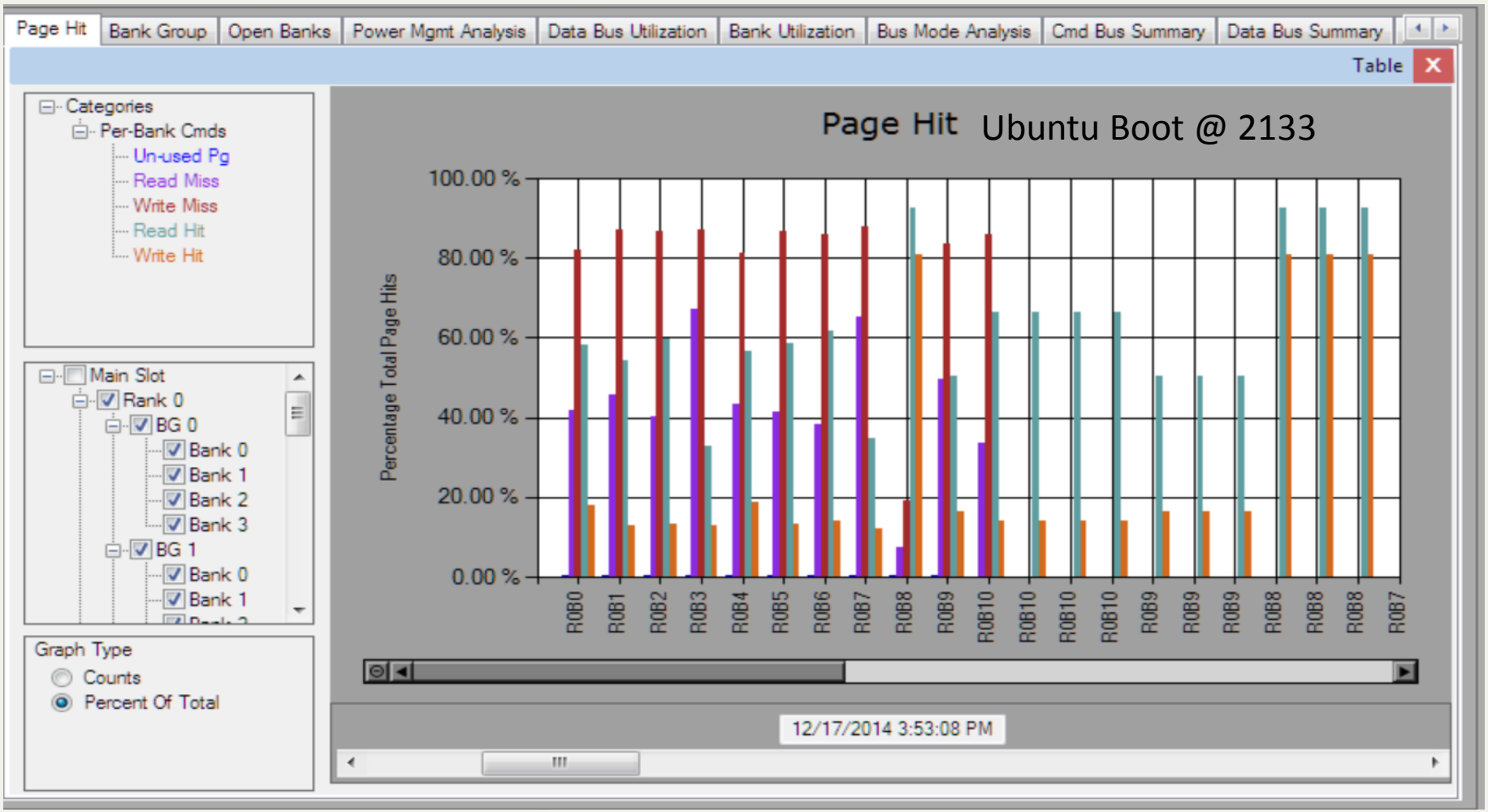
V#	Parameter	Description	Spec	Measured
V1	tCCD_L	RD to RD Same Bank Group	5	6
V3	tCCD_L	WR to WR Same Bank Group	5	6
V4	tCCD_S	RD to RD diff Bank Group	4	4
V5	tCCD_S	WR to WR diff Bank Group	4	4
V6	tRRD_L	ACT to ACT Same Bank Group	5	5
V7	tWTR_L	ACT to ACT diff Bank Group	4	4
V9	tWTR_L	WR to RD Same Bank Group	22	23
V10	tWTR_S	WR to RD Diff Bank Group	17	19

Latency

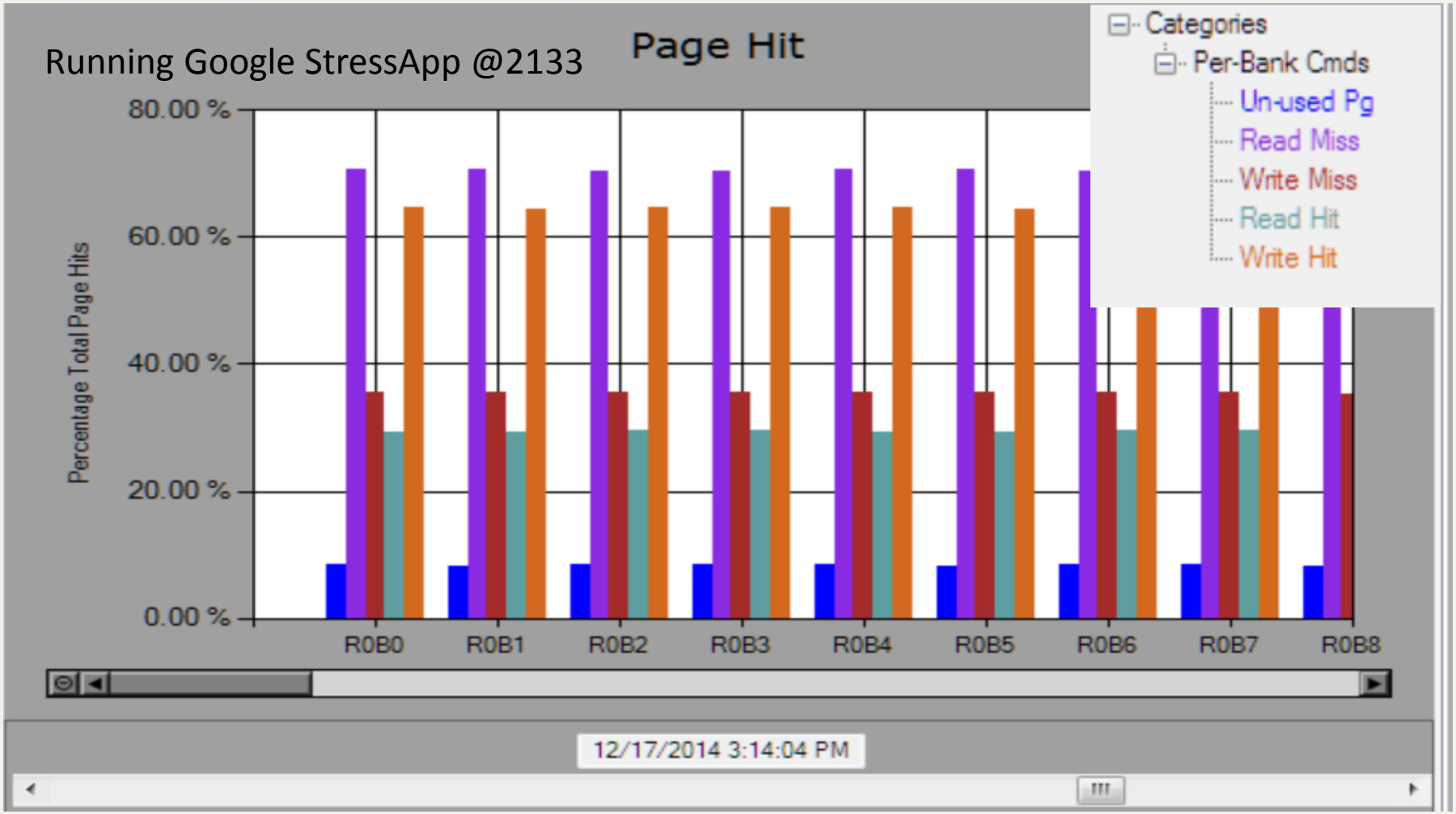
- Good designs operate on the edge of the spec
- Architectural tradeoffs will occur
- Do I need margin?
  - Design for the worst case and buy quality parts

New Performance Metrics

- Page Hit Analysis
  - Read Hit: Page was Open
  - Read Miss : Page was not Open, Transaction was preceded by an ACT
  - Write Hit: Page was Open
  - Write Miss: Page was not Open, Transaction was preceded by an ACT
  - Unused: Page was opened and closed and never accessed
- Multiple Open Banks
  - Open Banks make for faster access IF your going to that bank on the next access…performance hit if your not
  - Power hit when banks are open
- Bank Group Analysis
  - New for DDR4: Back to back access to same bank is a performance hit
  - Faster to have back to back accesses to different bank groups

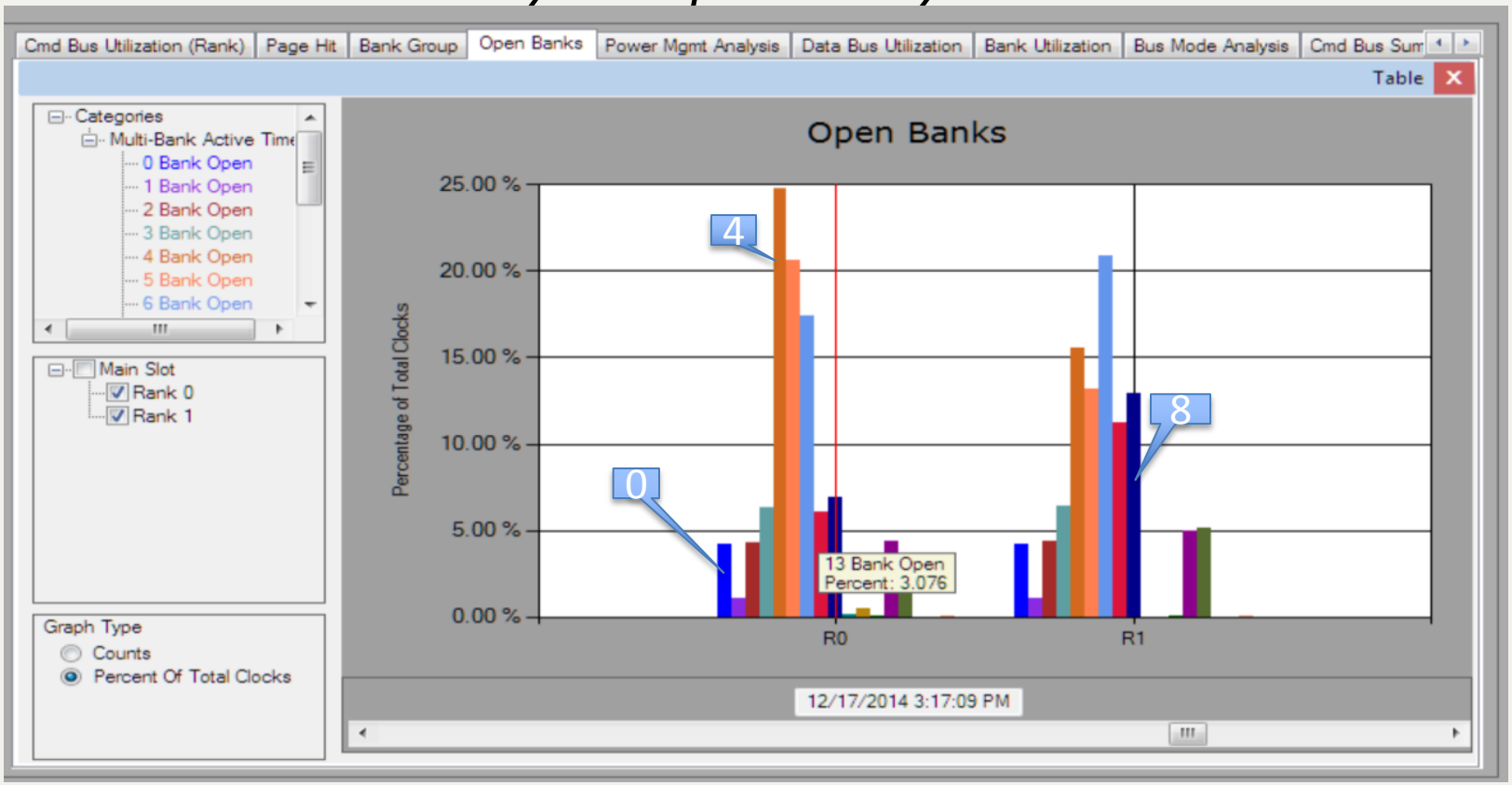

Page Hit
Bank Group
Open Banks
Power Mgmt Analysis
Data Bus Utilization
Bank Utilization
Bus Mode Analysis
Cmd Bus Summary
Data Bus Summary
Table
Categories
Per-Bank Cmds
Un-used Pg
Read Miss
Write Miss
Read Hit
Write Hit
Main Slot
Rank 0
BG 0
Bank 0
Bank 1
Bank 2
Bank 3
BG 1
Bank 0
Bank 1
Graph Type
Counts
Percent Of Total
Page Hit Ubuntu Boot @ 2133
Percentage Total Page Hits
100.00 %
80.00 %
60.00 %
40.00 %
20.00 %
0.00 %
R0B0
R0B1
R0B2
R0B3
R0B4
R0B5
R0B6
R0B7
R0B8
R0B9
R0B10
R0B10
R0B10
R0B10
R0B9
R0B9
R0B9
R0B8
R0B8
R0B8
R0B7
12/17/2014 3:53:08 PM

Running Google StressApp @2133



Multiple Open Banks

How many are open at any one time



KEYSIGHT TECHNOLOGIES
Strategic Solutions Partner

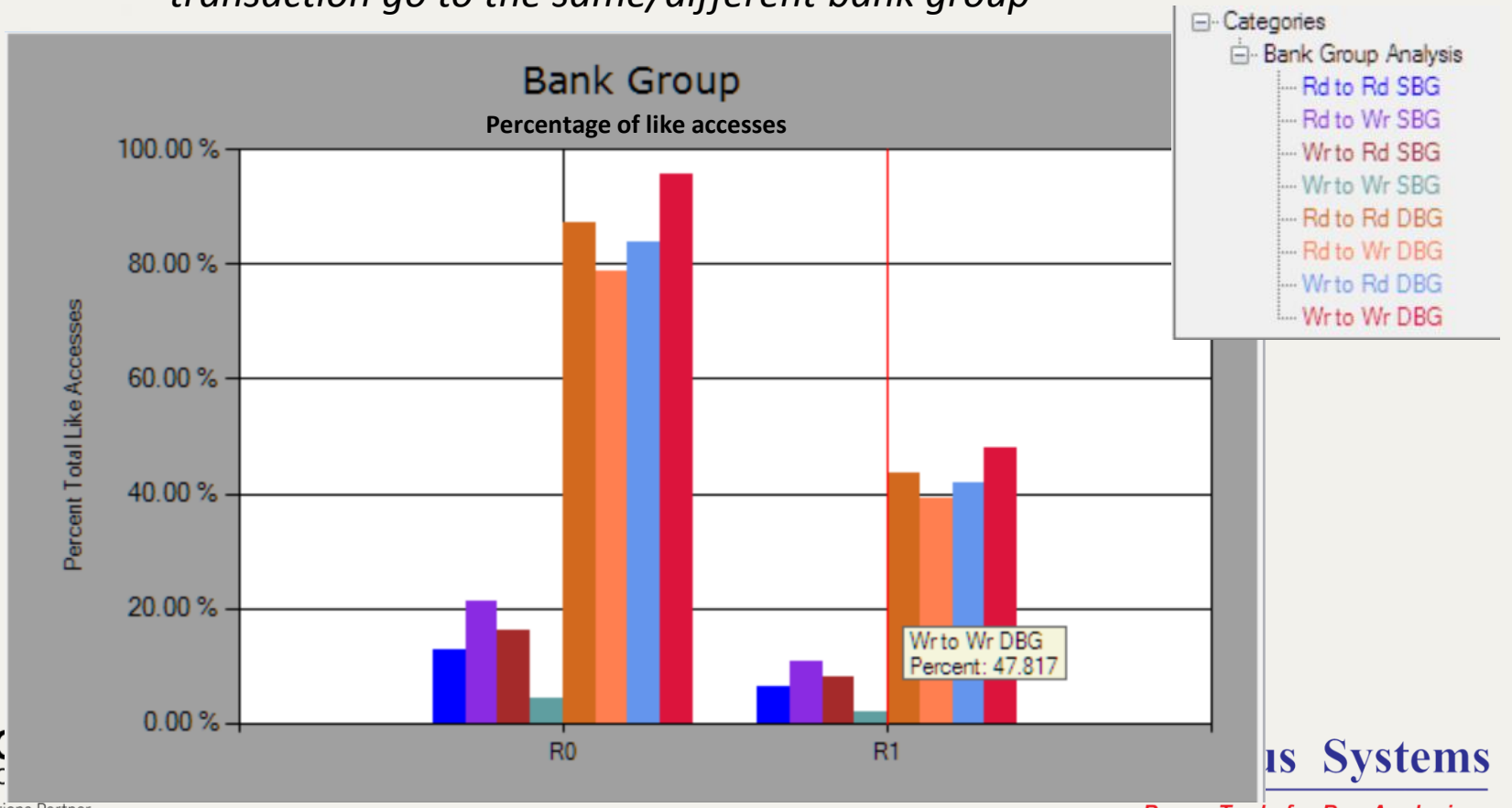
FuturePlus Systems
Power Tools for Bus Analysis

Bank Group Access Analysis

- tCCD_L
  - Takes longer for back to back RD/WR accesses to the same bank group
- tCCD_S
  - Can reduce latency by going to different bank groups

Bank Group Access Analysis

Relative to the previous transaction how many times did the following transaction go to the same/different bank group

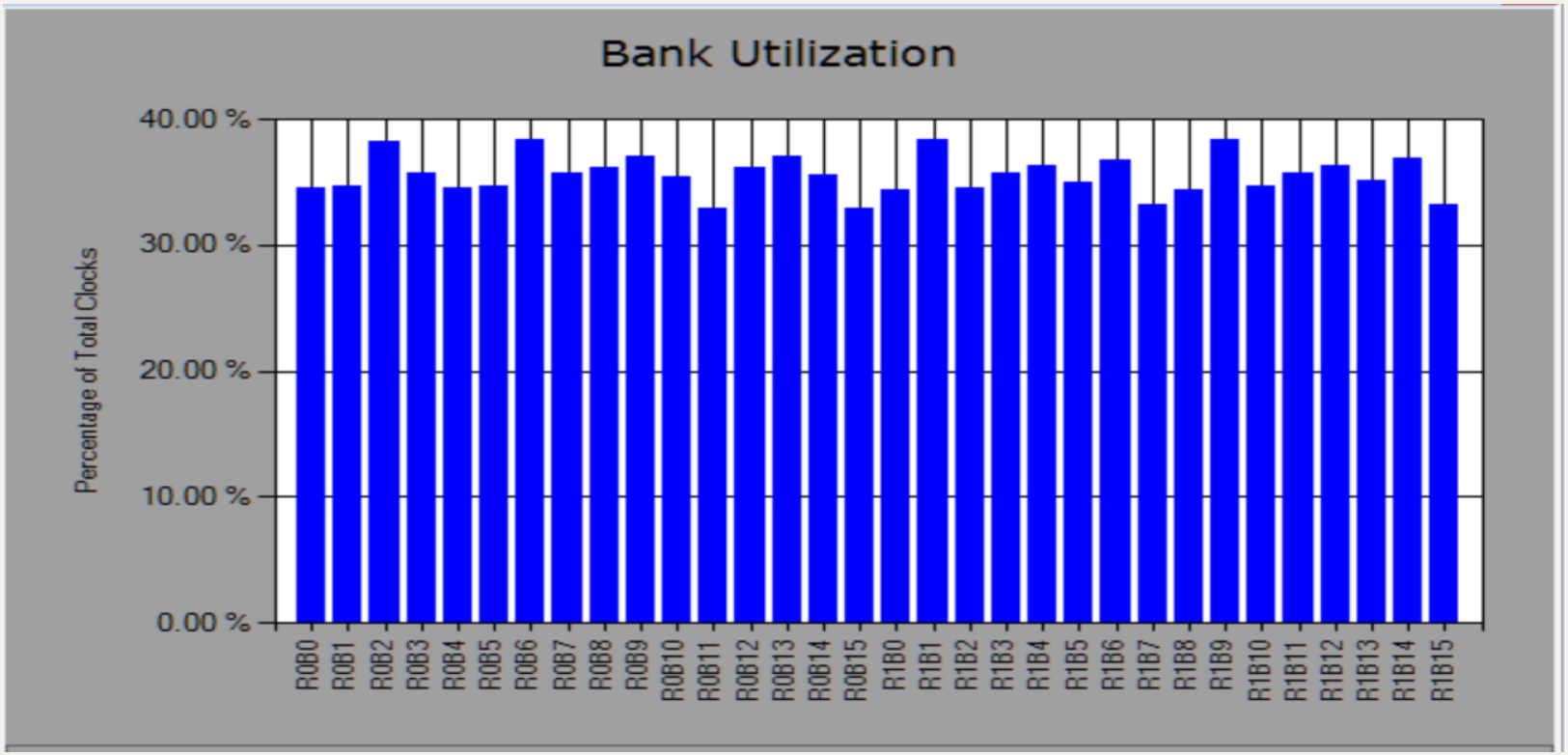


More Performance Metrics!

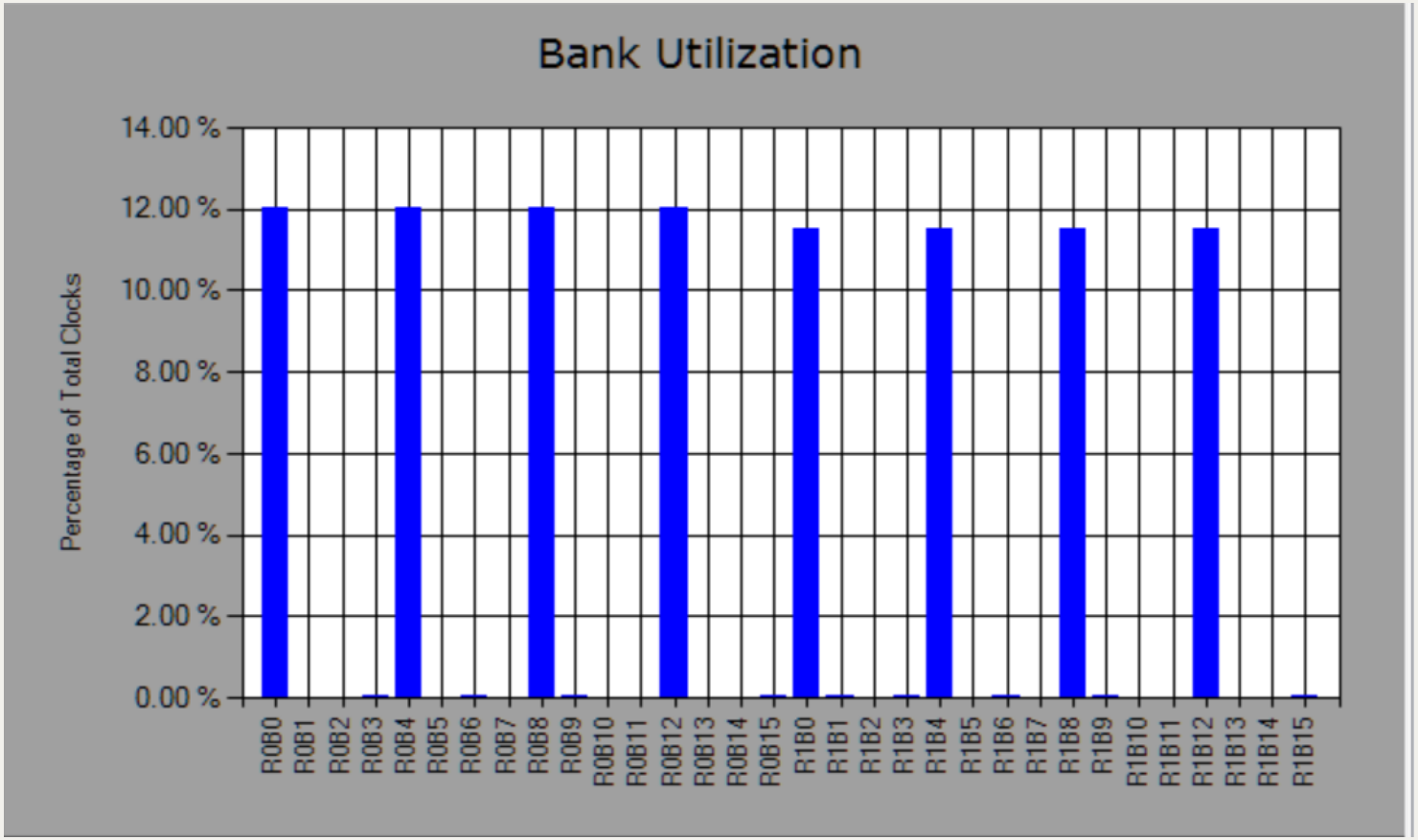
- Bank Utilization
  - What happens during a chip kill or page retirement scenario?
  - How does the traffic reallocate?
  - What are the performance implications?
- Do I have system hot spots?
  - Row Hammer (excessive Activates)
- Fast Boot
  - Why does the system take so long to boot?

Bank Utilization

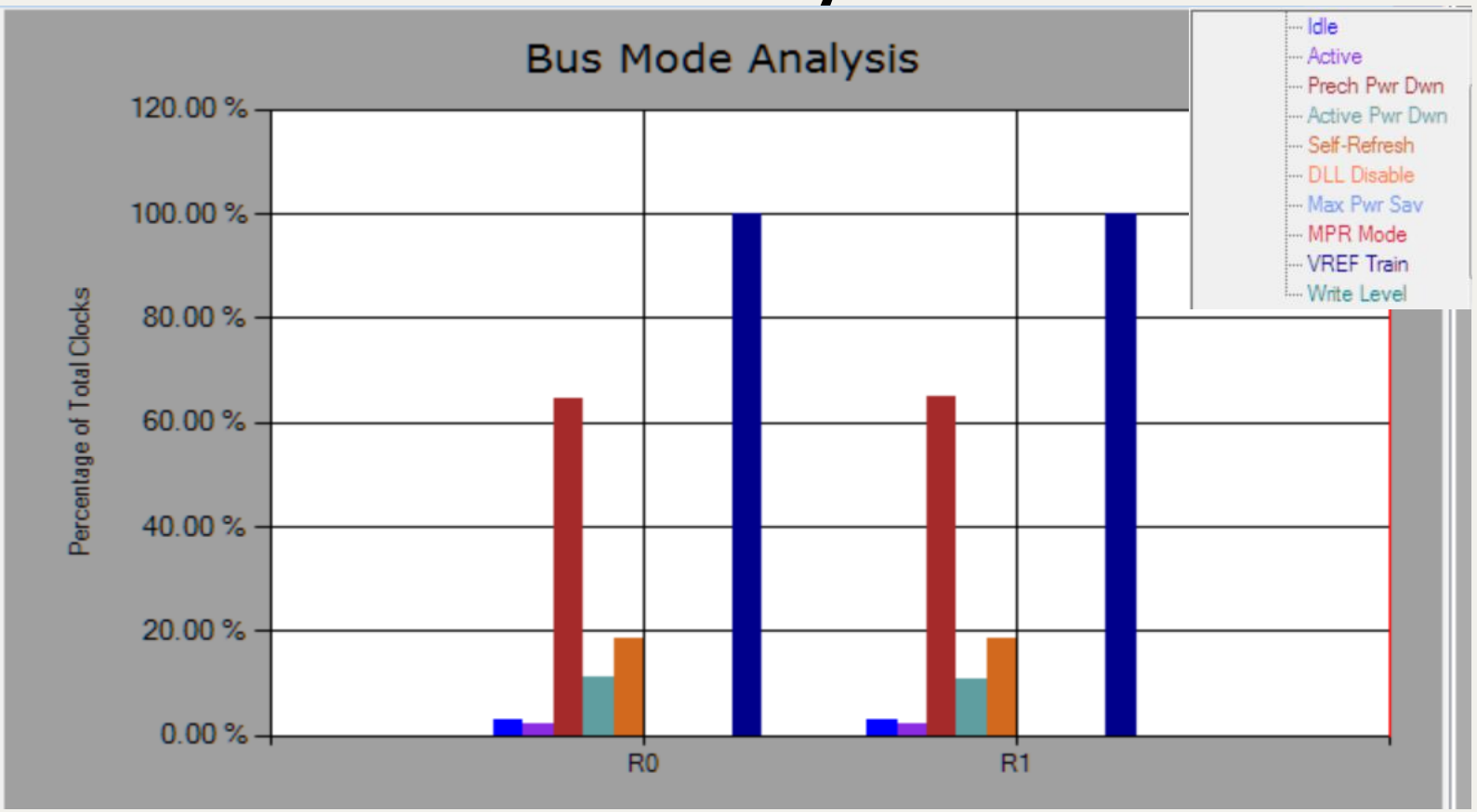
Percentage of total cycles the banks are open (running Google Stress App)



Bank Utilization



Boot Analysis



Row Hammer

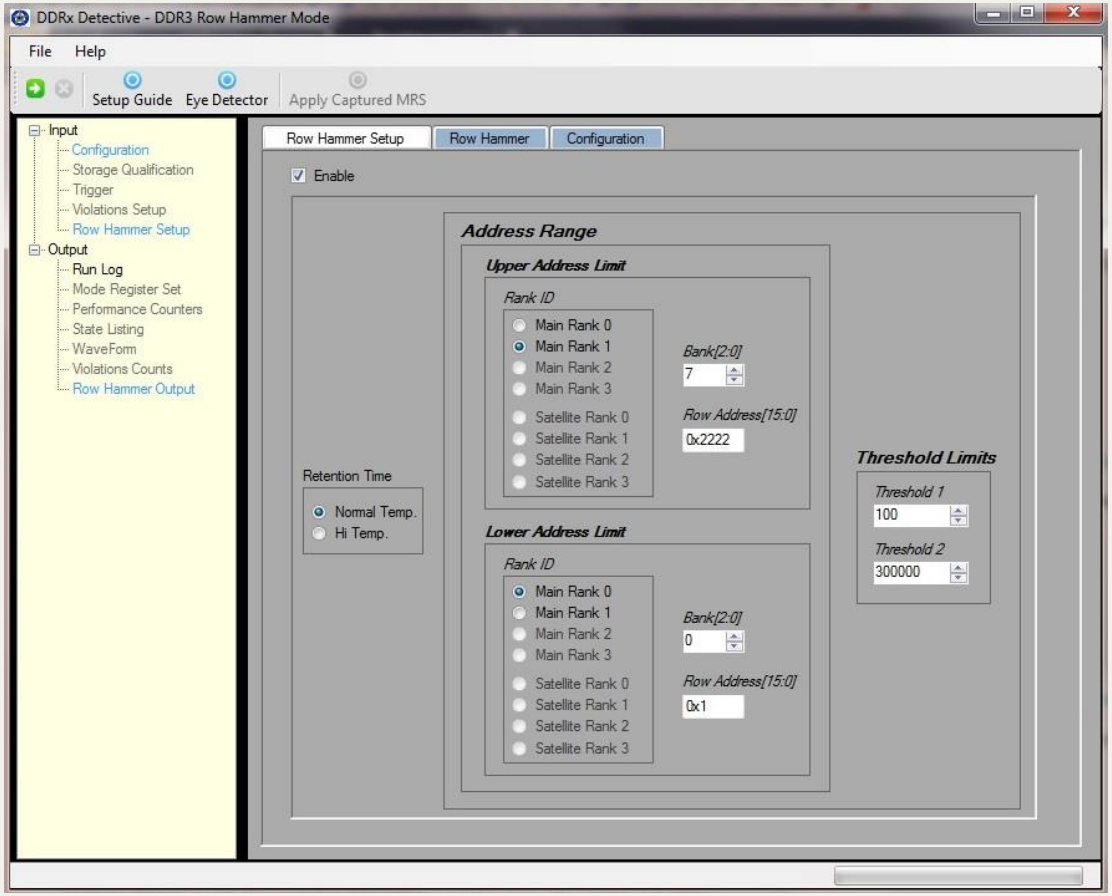
Excessive ACTIVATE commands to a single Row

- An identified failure mechanism in DDR3 memory
- The result of charge leakage
- Current work around is to increase the refresh rate
  - Burns power, lowers performance
  - Lowers the statistical probability of an error

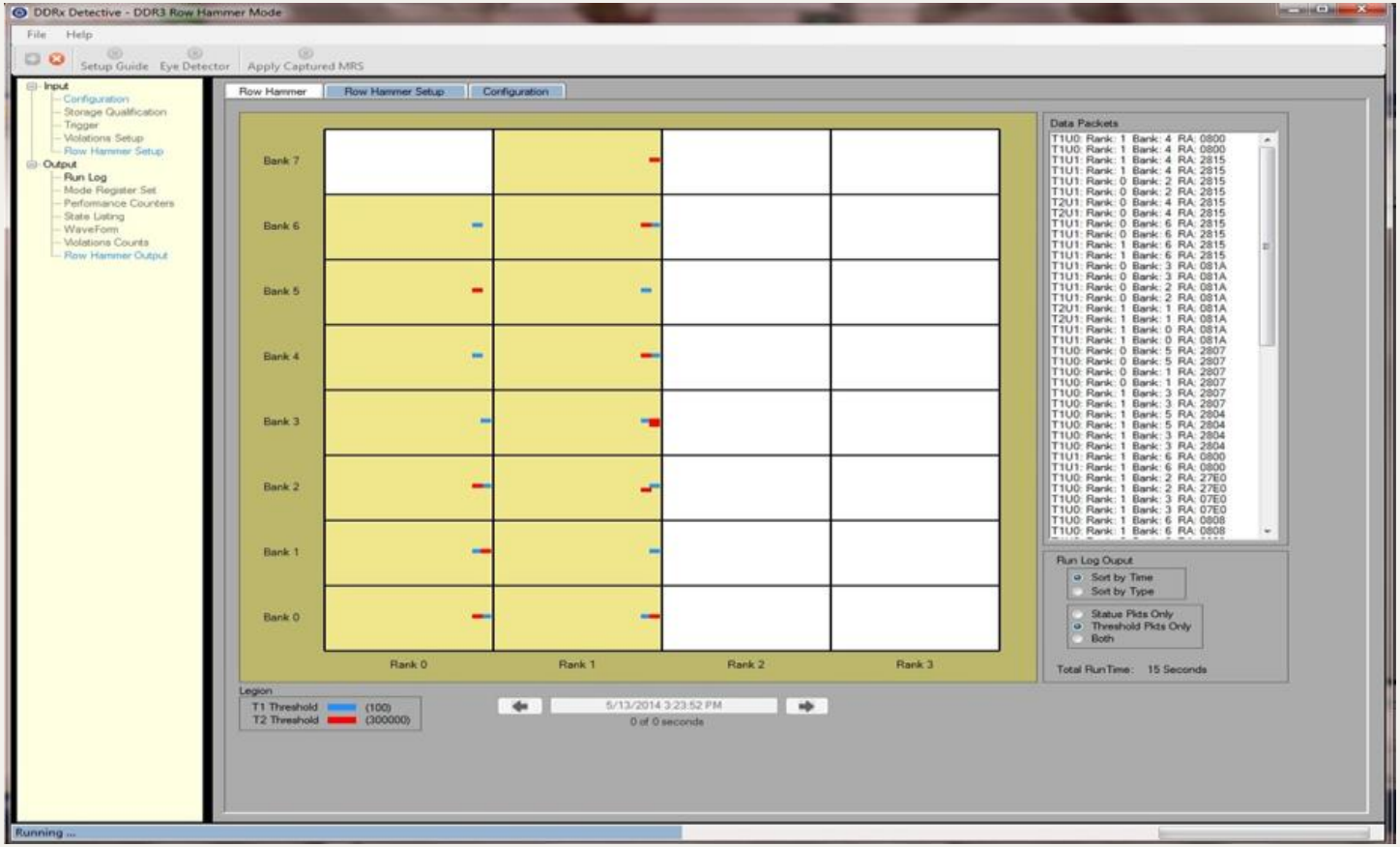
Critical Applications cannot tolerate Row Hammer failures

- Memory Vendors are specifying how many ACTIVATES per Retention cycle to a single row the memory can tolerate
- However there is no way the memory controller or the memory can count this
- Critical Software applications should be tested to ensure that the Row Hammer Threshold is not reached

One method to detect Row Hammer



Detecting Excessive ACTIVATE Commands



Critical DDR4 Performance Metrics

- Memory Controller/System Architecture
  - Can this insight lead to better designs?
  - Benchmark Servers Memory Performance
- Workload Analysis
  - Should the Memory Controller settings be based on criteria set by the workload?
  - Can compilers be made better?
  - Can critical applications be written to avoid Row Hammer?
- Do we all need a DDR5?
  - Work Smarter not Harder and understand what we have

Summary

- Power Management, Bandwidth, Latency
- NEW Metrics:
  - Page Hit Analysis
  - Multiple Open Banks
  - Bank Group Analysis
  - Bank Utilization
  - Boot Analysis
  - Row Hammer Detection
  - New Measurements give insight into new designs and better architectures

Contact Information

Barbara Aichinger

Vice President New Business Development

FuturePlus Systems

Barb.Aichinger@FuturePlus.com

www.FuturePlus.com

Check out our new website dedicated to DDR Memory! www.DDRDetective.com