Optimizing NAND Flash Performance

Ryan Fisher (rfisher@micron.com)
Sr. Applications Engineer
Micron Technology, Inc.

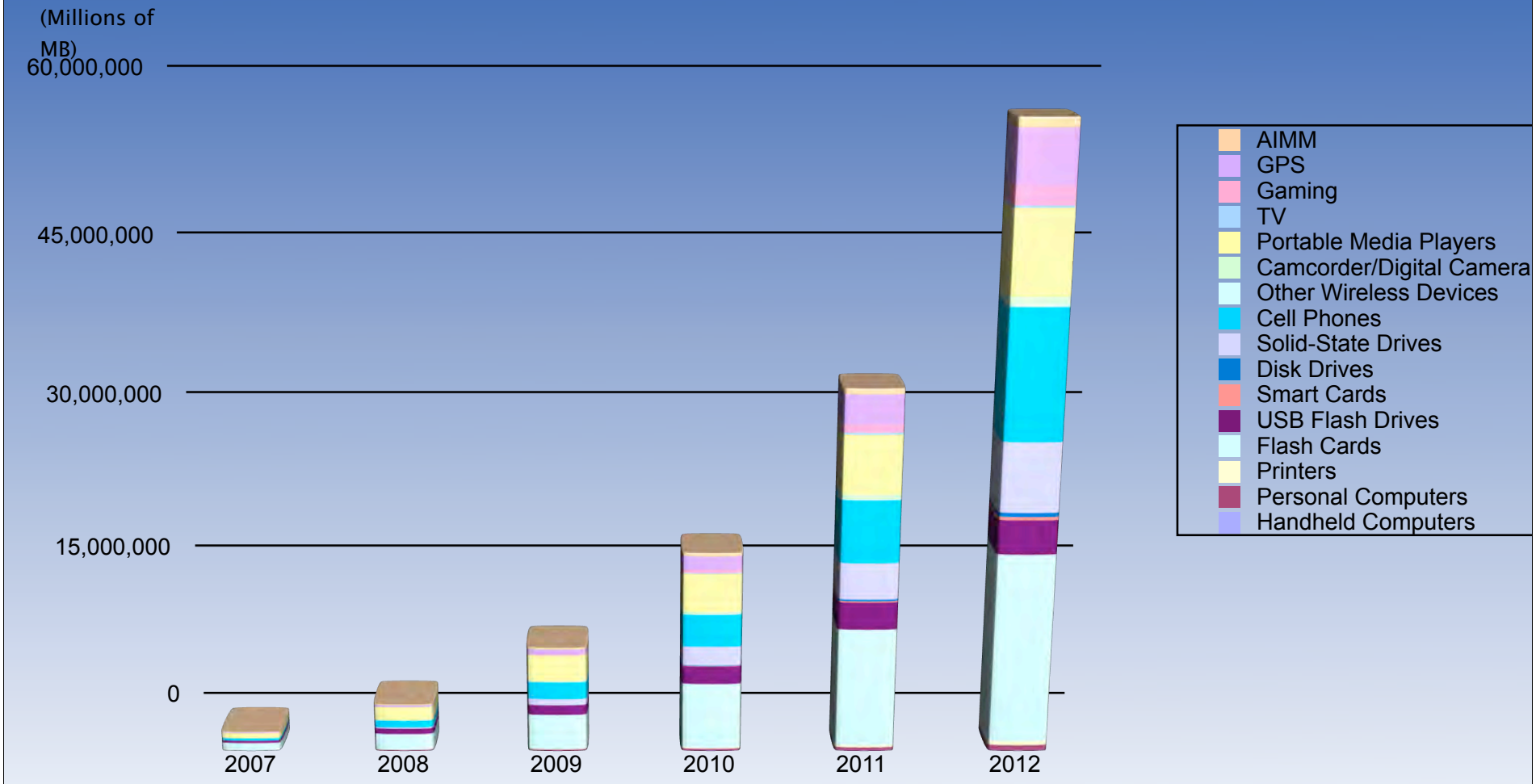
Santa Clara, CA USA
August 2008

NAND Usage by Application

NAND Flash is finding its way into increasingly diverse end systems and applications—all with different requirements when it comes to NAND Flash

- Total density
  - Single package vs. multiple packages
  - Package type (TSOP/BGA/LGA)
- Configuration of NAND within a system
  - Single-channel vs. multi-channel
- Performance requirements (system-level PROGRAM/READ)

NAND Usage by Application



Source: Gartner 2Q08

Micron defines AIMM as Automotive, Industrial, Medical, Military

NAND Array Performance

NAND array performance times (^{t}PROG, ^{t}R, ^{t}BERS) have a major influence on NAND throughput performance

- They're outside the control of the host system
- They vary for SLC and MLC NAND devices
  - Even varies from SLC to SLC device and MLC to MLC device

To proactively improve overall NAND throughput, the focus needs to be on specific characteristics of the NAND device and a system-level configuration of the NAND device that systems/applications can take advantage of

Improving NAND Performance

Various methods can be used to improve NAND Flash performance

- Optimize NAND AC timings
  - Using minimum timings available for a given NAND device
- Optimize NAND bus activity
  - Avoid long idle times on the NAND bus during operations
- Cache Operations
  - Read and program cache operations support data pipelining by reading or writing data while NAND array access operations occur in the background
- Multi-Plane Commands
  - Available on NAND devices with multi-plane architecture
  - Support operations (PROGRAM/READ/ERASE) to each plane simultaneously, increasing data throughput
- Interleave Operations
  - Available on NAND devices with more than one NAND die and/or more than one CE#
  - Support ongoing, simultaneous operations on more than one NAND die
- Multi-Channel Operations
  - Available when using multiple independent command and/or I/O channels to communicate with more than one NAND device in parallel operations
  - Available with some configurations of packaged NAND devices (BGA/LGA)

AC Timing Optimization

Optimizing NAND AC timings (specifically ^{t}WC/^{t}RC) can have a dramatic impact on performance

- It can be a less costly/complex method of improving NAND Flash performance in non-optimized systems (i.e., updating a count/timing register in a host controller)
- It can have a significant impact on total time to perform NAND array (PROGRAM/READ) operations
  - This becomes more important as the size of the NAND page increases (2KB/4KB/8KB, etc.)
  - This is important for READ operations because time for data output from the NAND is greater than ^{t}R
  - This issue is addressed by the ONFI 2.0 synchronous interface; attend Session T2A on Weds at 2:40 p.m., presented by Michael Abraham, for more information on faster I/O performance with ONFI 2.0
- It sets the stage for optimal use of interleaved operations

AC Timing Optimization

Example effects of optimized vs. non-optimized AC timings

Five mass storage class devices with different NAND controllers were analyzed to determine what ^{t}WC timings were being utilized given a NAND data sheet specification of 20ns ^{t}WC MIN:

Based on NAND Data sheet ^{t}PROG for a single page program

Time % increase for page program		^{t}WC	Time to input 4KB NAND page data	Controllers analyzed that used timing
SLC	MLC			
35%	10%	45ns	~185µs	1
14%	4%	30ns	~123µs	3
7%	2%	25ns	~103µs	1

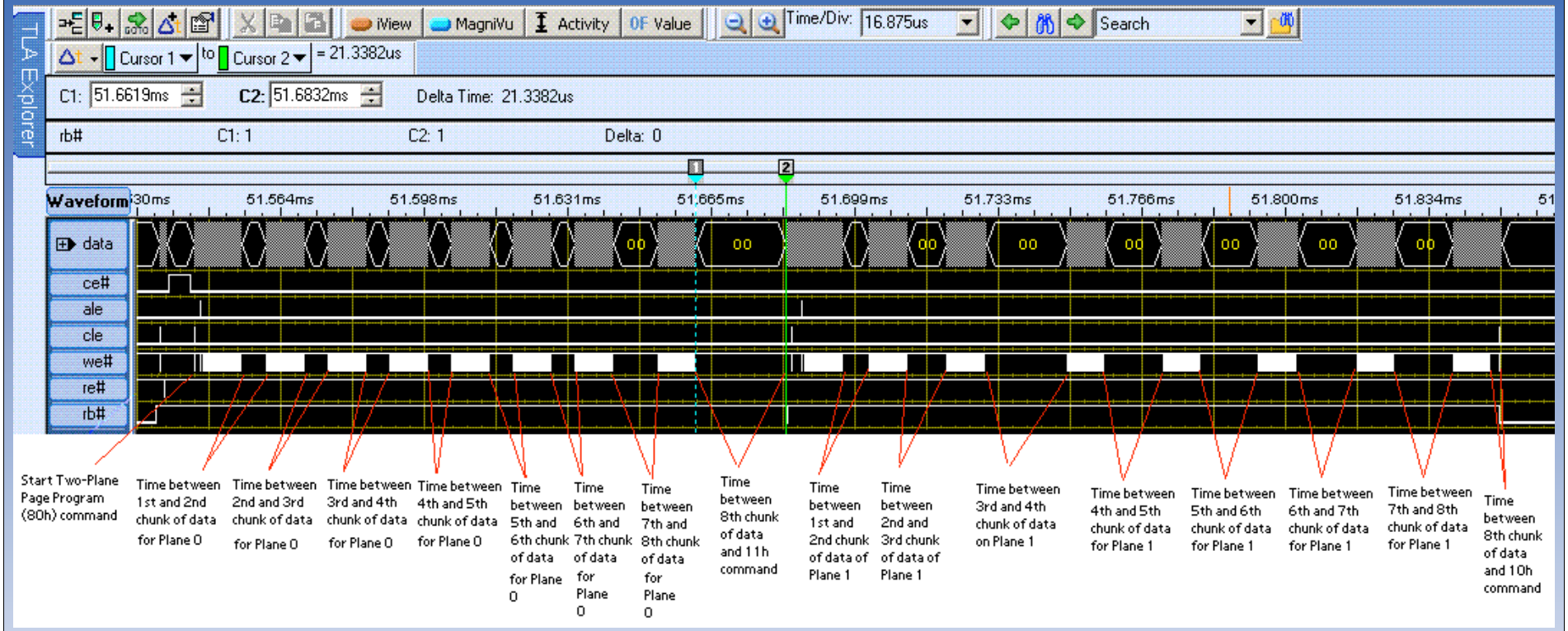
NAND Bus Activity Optimization

Optimizing NAND bus activity to avoid excessive idle time represents another system-level interaction with the NAND device that is sometimes overlooked when trying to improve overall NAND Flash performance

Common reasons for NAND bus idle time

1. System limitations dealing with the amount of buffer space available to communicate with the NAND cache register
2. System-level multi-tasking operations that limit the amount of time dedicated to communicating with the NAND device at a given time
3. Higher-level system interactions (OS level) with the NAND controller that do not communicate in the most efficient way

NAND Bus Activity Optimization



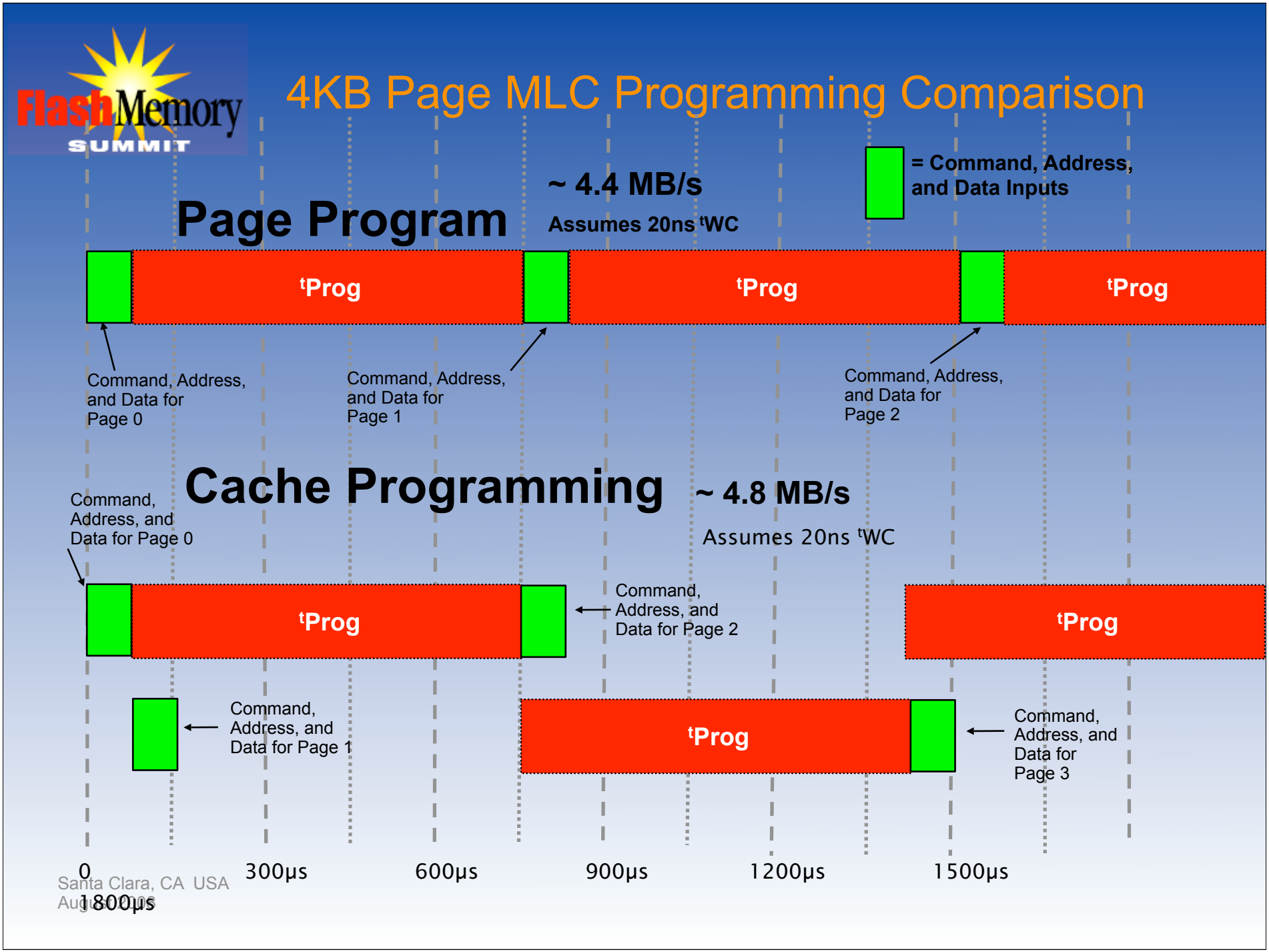
Example of two-plane page program operation issued to a NAND Flash device in customer application

- Total extra time added for inputting of data due to “idle” times shown above is ~166µs
- This equates to a 95% time increase compared to the time required for a optimized two-plane page program operation data input sequence

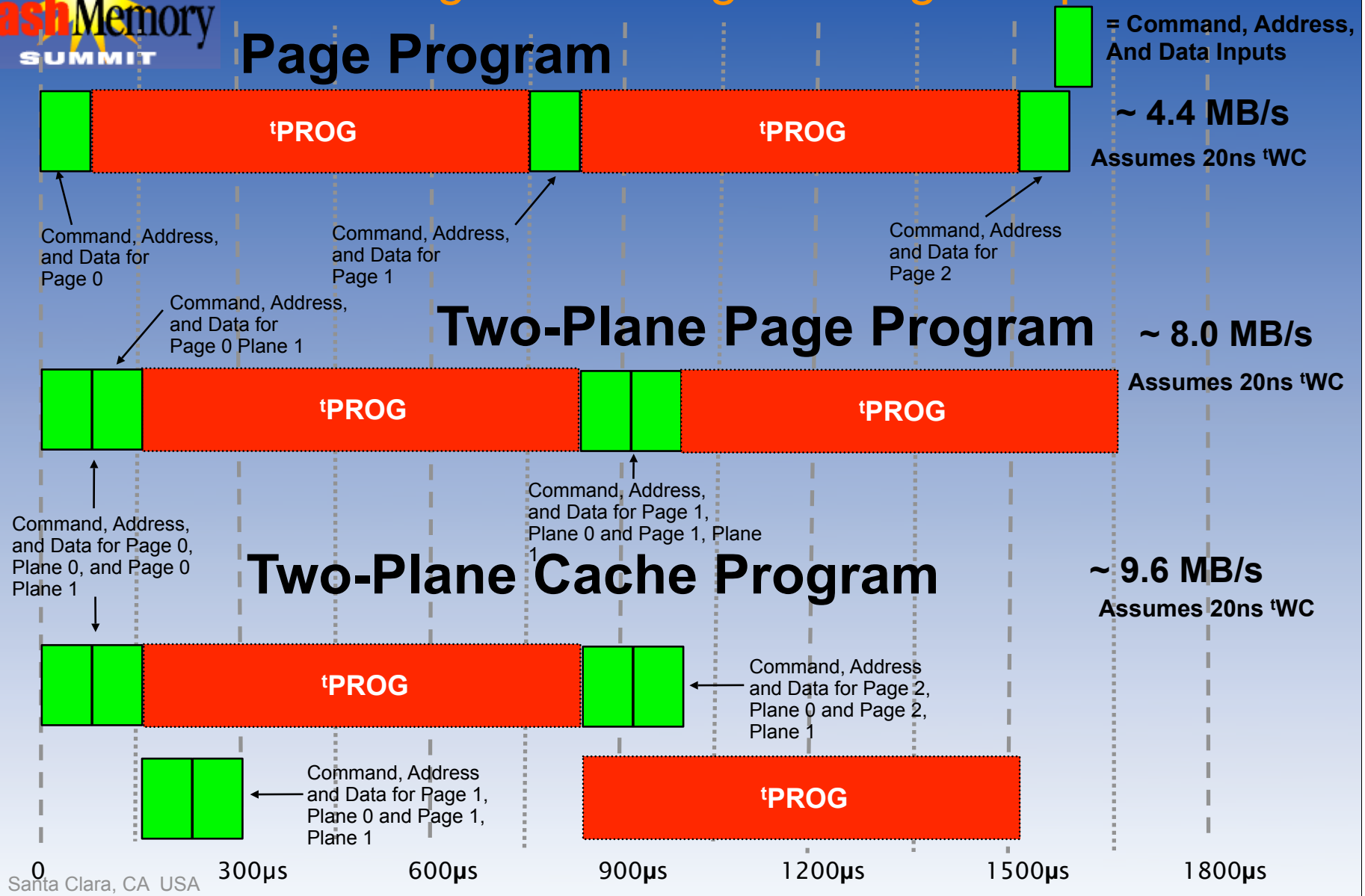
Cache and Multi-Plane Operations

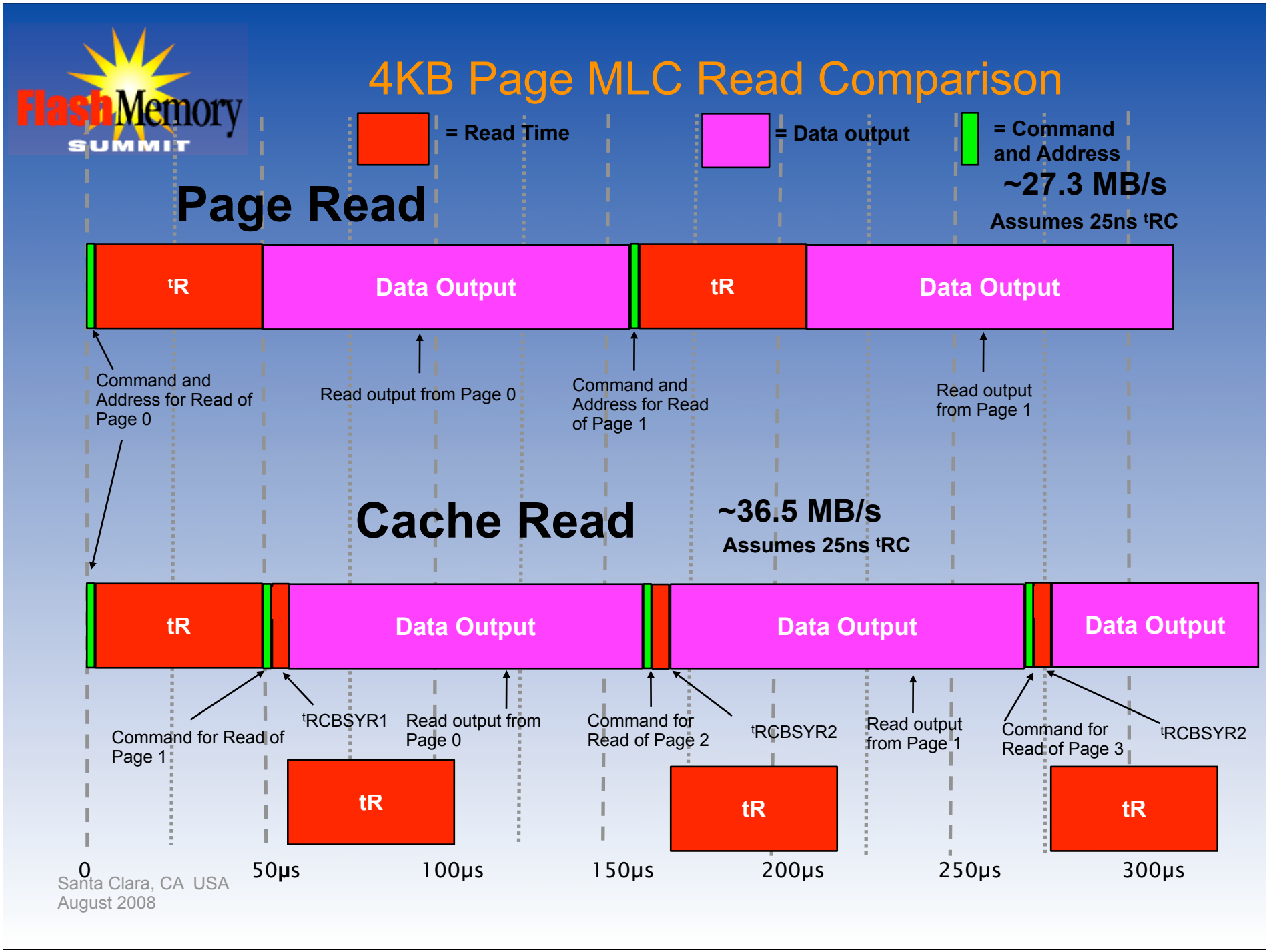
Cache and multi-plane operations represent another means of improving NAND Flash device performance

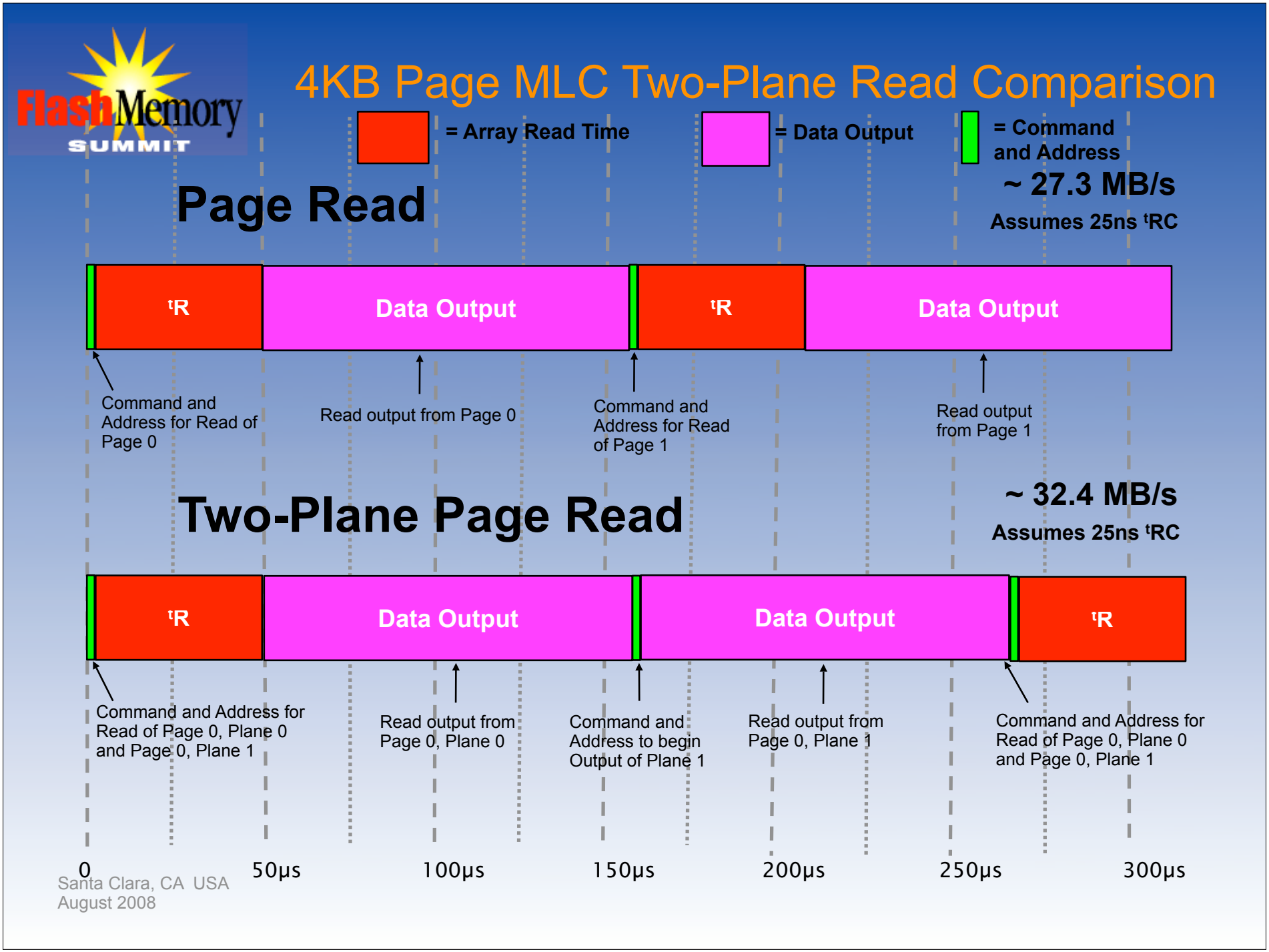
- Cache operations
  - Require slightly different NAND Flash commands than normal PROGRAM/READ operations
- Multi-plane operations
  - Require different NAND command codes than normal PROGRAM/READ/ERASE operations
  - Have address restrictions that must be followed
  - A host buffer (used to communicate with the NAND device) is recommended, but not required, to be at least 2X the size of the NAND page


Flash Memory SUMMIT
4KB Page MLC Programming Comparison
= Command, Address, and Data Inputs
Page Program
~ 4.4 MB/s
Assumes 20ns tWC
tProg
tProg
tProg
Command, Address, and Data for Page 0
Command, Address, and Data for Page 1
Command, Address, and Data for Page 2
Cache Programming
~ 4.8 MB/s
Assumes 20ns tWC
Command, Address, and Data for Page 0
tProg
Command, Address, and Data for Page 2
tProg
Command, Address, and Data for Page 1
tProg
Command, Address, and Data for Page 3
0
300µs
600µs
900µs
1200µs
1500µs
1800µs

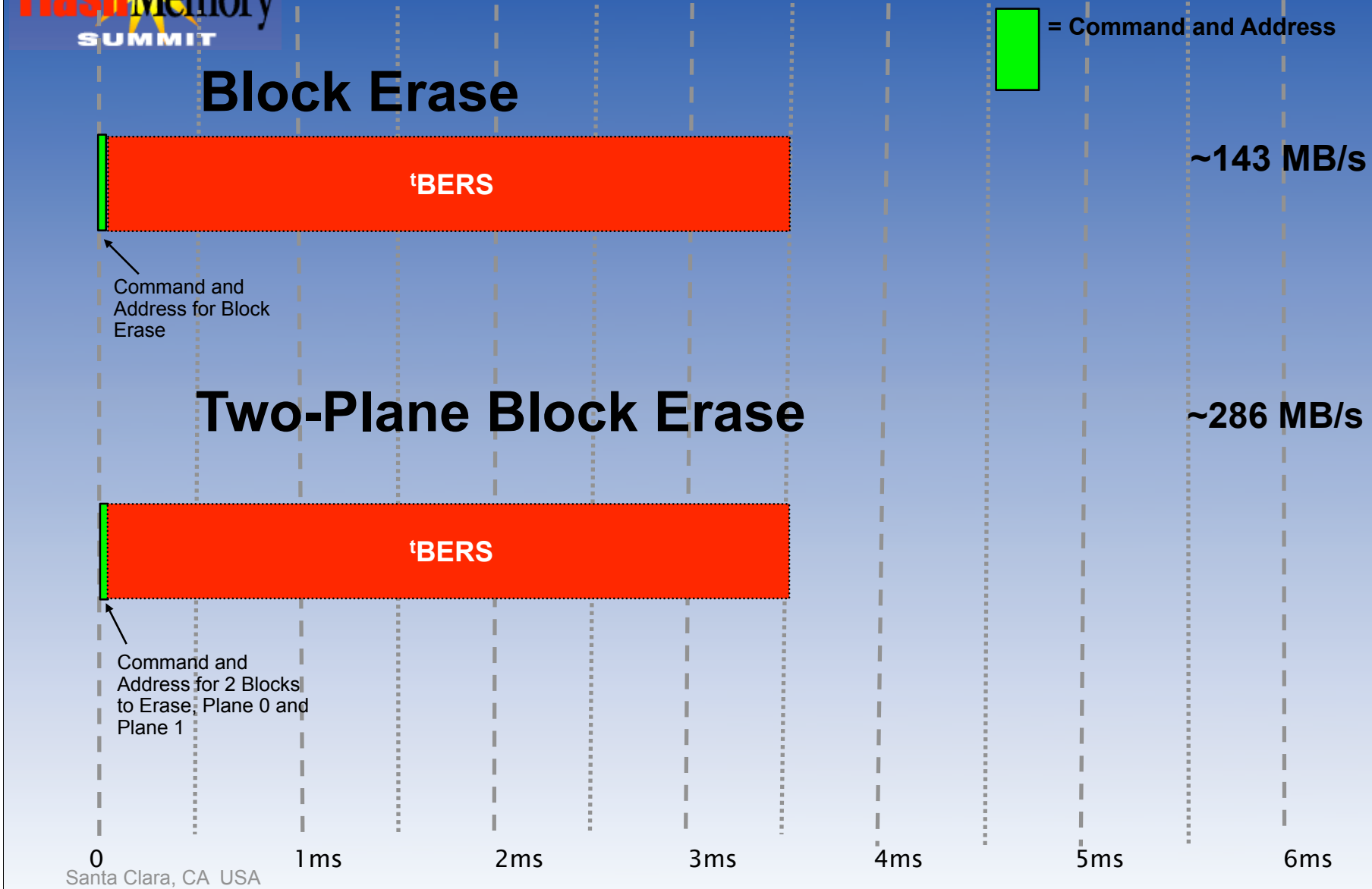
4KB Page MLC Programming Comparison




Flash Memory SUMMIT
4KB Page MLC Read Comparison
= Read Time
= Data output
= Command and Address
~27.3 MB/s
Assumes 25ns tRC
Page Read
tR
Data Output
tR
Data Output
Command and Address for Read of Page 0
Read output from Page 0
Command and Address for Read of Page 1
Read output from Page 1
Cache Read
~36.5 MB/s
Assumes 25ns tRC
tR
Data Output
Data Output
Data Output
Command for Read of Page 1
tRCBSYR1
Read output from Page 0
Command for Read of Page 2
tRCBSYR2
Read output from Page 1
Command for Read of Page 3
tRCBSYR2
tR
tR
tR
0
50µs
100µs
150µs
200µs
250µs
300µs
Santa Clara, CA USA
August 2008


Flash Memory SUMMIT
4KB Page MLC Two-Plane Read Comparison
= Array Read Time
= Data Output
= Command and Address
Page Read
~ 27.3 MB/s
Assumes 25ns tRC
tR
Data Output
tR
Data Output
Command and Address for Read of Page 0
Read output from Page 0
Command and Address for Read of Page 1
Read output from Page 1
Two-Plane Page Read
~ 32.4 MB/s
Assumes 25ns tRC
tR
Data Output
Data Output
tR
Command and Address for Read of Page 0, Plane 0 and Page 0, Plane 1
Read output from Page 0, Plane 0
Command and Address to begin Output of Plane 1
Read output from Page 0, Plane 1
Command and Address for Read of Page 0, Plane 0 and Page 0, Plane 1
0
50µs
100µs
150µs
200µs
250µs
300µs
Santa Clara, CA USA
August 2008

4KB Page MLC Erase Comparison



Interleaved Operations

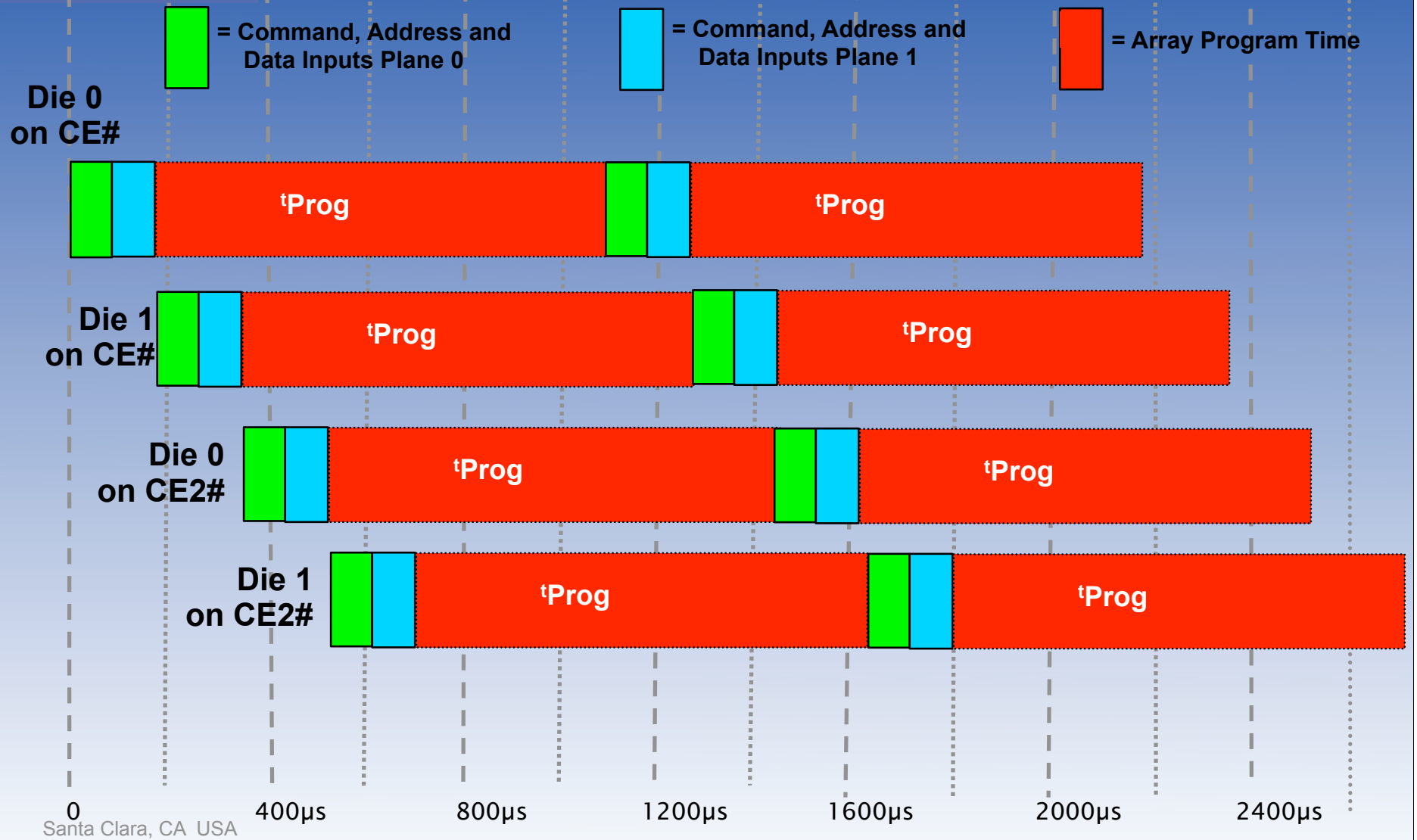
Interleaved operations can be used to improve NAND Flash array performance

- Depending on the system configuration, interleaved operations may be easier to implement that multi-plane operations
- Interleaved operations have the benefit of requiring:
  - Little extra support from the host system
  - No address restrictions within the additional NAND Flash die being used

Specifically, there are two types of interleaved operations

- Interleaved operations between multiple NAND die on the same CE#
- Interleaved operations between multiple NAND die on different CE#s within the same packaged device or multiple packaged devices

Interleaved Operation Two-Plane Page Program

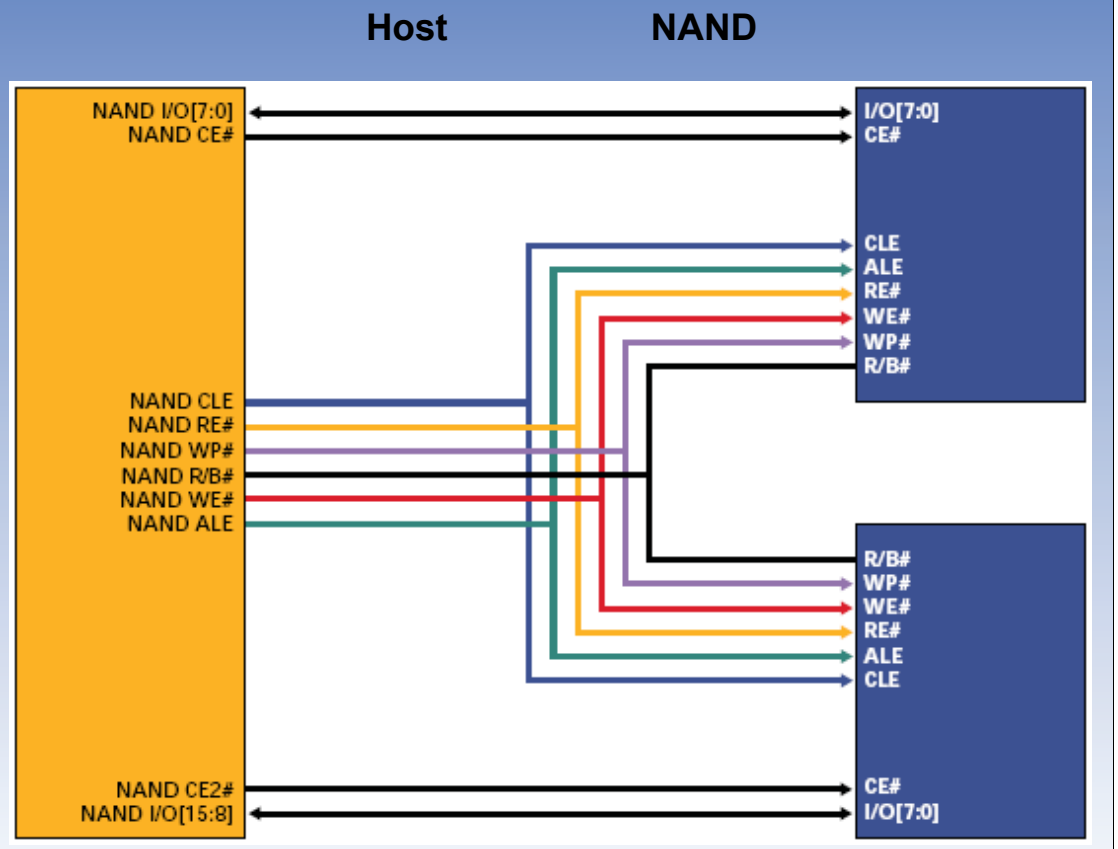


Multi-Channel Operations

Multi-Channel Operations

- Offer true operation parallelism for improving NAND Flash performance in a system
- Require increased number of signals from the host system to support this configuration

Two-Channel Configuration Example

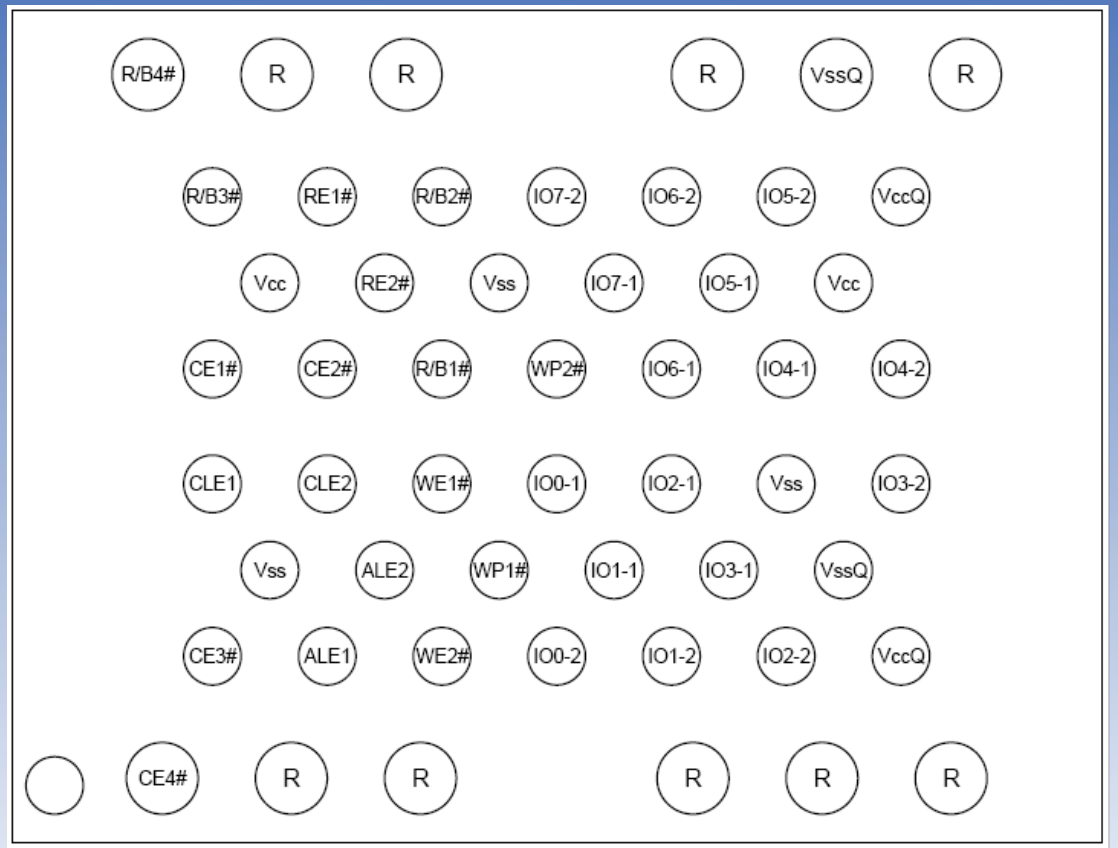


Multi-Channel Operations

Some packages with multiple NAND die offer multi-channel commands and I/O setups

- Two separate sets of command signals
  - CLE1, ALE1, WE1#, RE1#, WP1#, CE1#
  - CLE2, ALE2, WE2#, RE2#, WP2#, CE2#
- Two separate sets of I/O signals
  - I/O[7:0]-1
  - I/O[7:0]-2

Signal Assignments for ONFI-Defined x8 LGA



Source: ONFI 2.0 Specification (www.onfi.org)

Improving NAND Performance

Any of the techniques discussed will improve NAND Flash performance, and combining two or more of these techniques will provide further performance enhancements

Some of the more common combinations of these techniques are:

- Multi-channel configuration with multi-plane operations
- Multi-channel configuration with interleaved operations
- Multi-plane operations with interleaved operations
- Multi-plane operations with optimized NAND AC timings

Resources for Improving NAND Flash Performance

Micron offers additional information for improving NAND Flash performance, available on the NAND Flash Technical Notes web page at: www.micron.com/product/nand/technotes

- TN-29-01: NAND Flash Performance Increase – Using the Micron PAGE READ CACHE MODE Command
- TN-29-14: NAND Flash Performance Increase with PROGRAM PAGE CACHE MODE Command
- TN-29-25: Improving NAND Flash Performance Using Two-Plane Command Enabled Micron Device
- TN-29-28: Memory Management in NAND Flash Arrays
  - Recommended for anyone considering using multi-plane/interleaved/multi-channel configurations

Thank You

Santa Clara, CA USA
August 2008

Biography

- Ryan Fisher has a BSEE from the University of Florida and has worked in the semiconductor industry for 7 years
- Positions held:
  - Motorola SPS (Austin, TX): Applications Engineer for Wireless Baseband Processors
  - Micron Technology (Boise, ID): Sr. Applications Engineer for NAND Flash memory