

# The Consequences of Infinite Storage Bandwidth

Allen Samuels

Engineering Fellow, Systems and Software Solutions

April 21, 2016

**SanDisk**<sup>®</sup>

# Disclaimer

During the presentation today, we may make forward-looking statements.

Any statement that refers to expectations, projections, or other characterizations of future events or circumstances is a forward-looking statement, including those relating to industry predictions and trends, future products and their projected availability, and evolution of product capacities. Actual results may differ materially from those expressed in these forward-looking statements due to a number of risks and uncertainties, including among others: industry predictions may not occur as expected, products may not become available as expected, and products may not evolve as expected; and the factors detailed under the caption “Risk Factors” and elsewhere in the documents we file from time to time with the SEC, including, but not limited to, our annual report on Form 10-K for the year ended January 3, 2016. This presentation contains information from third parties, which reflect their projections as of the date of issuance. We undertake no obligation to update these forward-looking statements, which speak only as of the date hereof or the date of issuance by a third party.

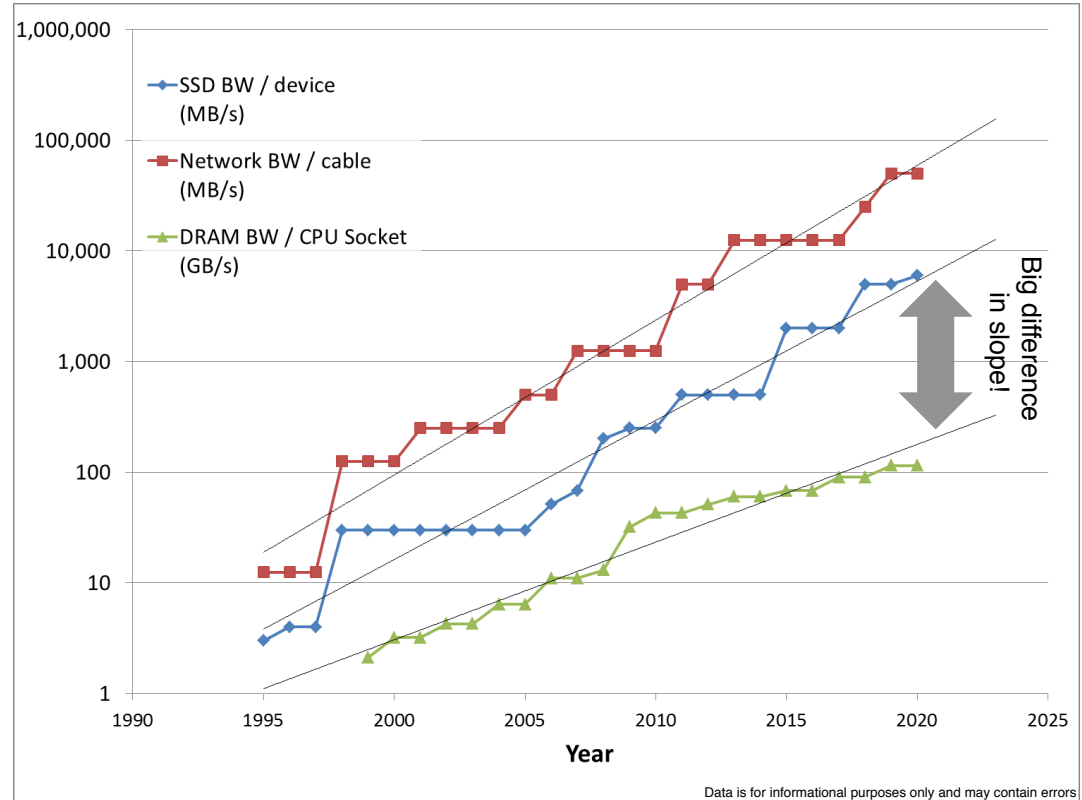
**What do I Mean By Infinite Bandwidth ?**

**SanDisk®**

# Network, Storage and DRAM Trends

Log scale

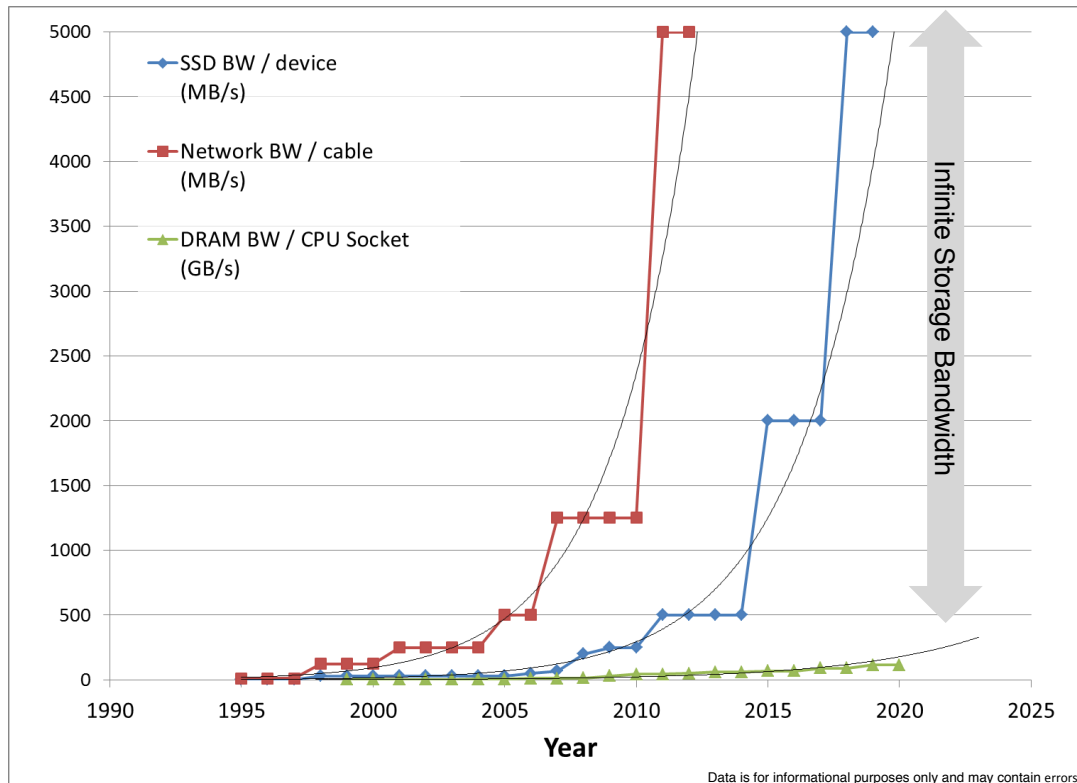
- Use DRAM Bandwidth as a proxy for CPU throughput
- Reasonable approximation for DMA heavy, and/or poor cache hit performance workloads (e.g. Storage)

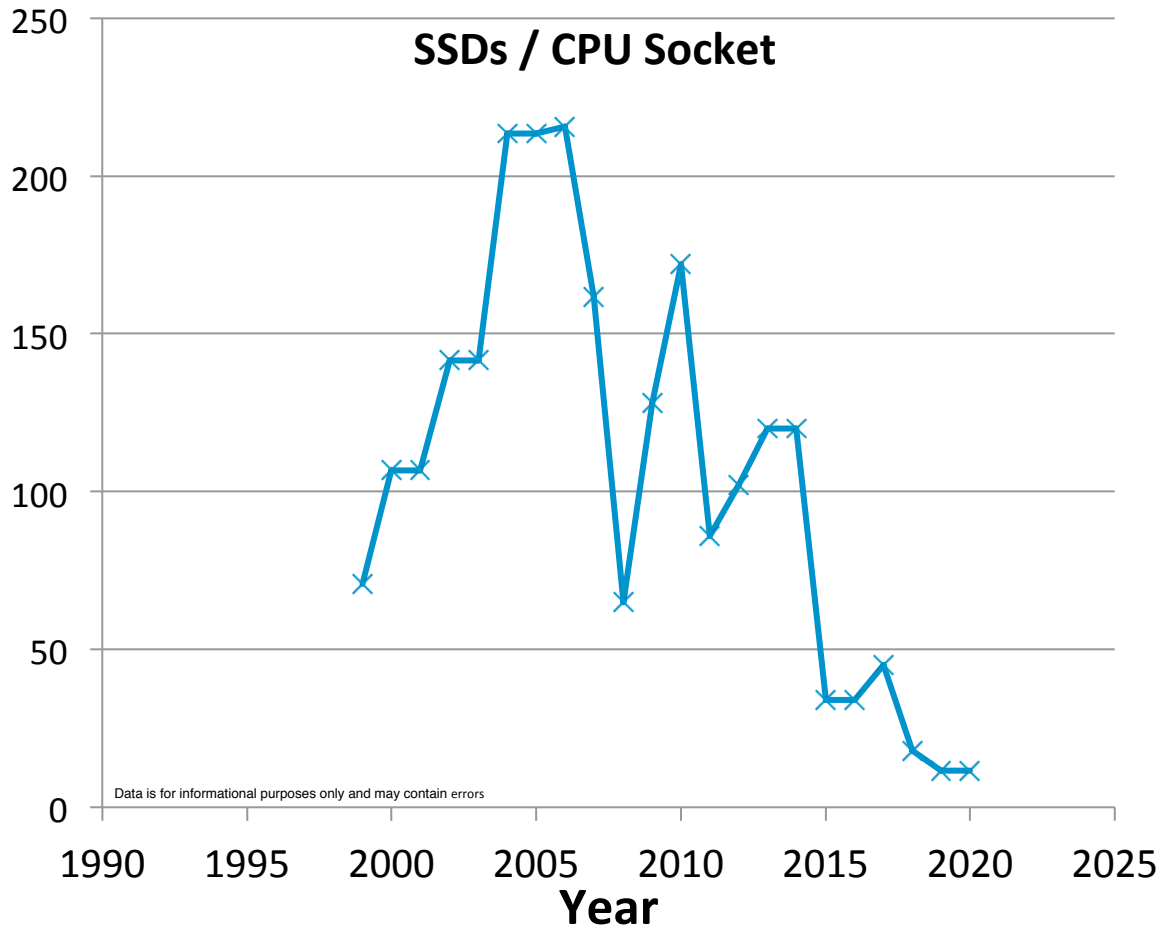


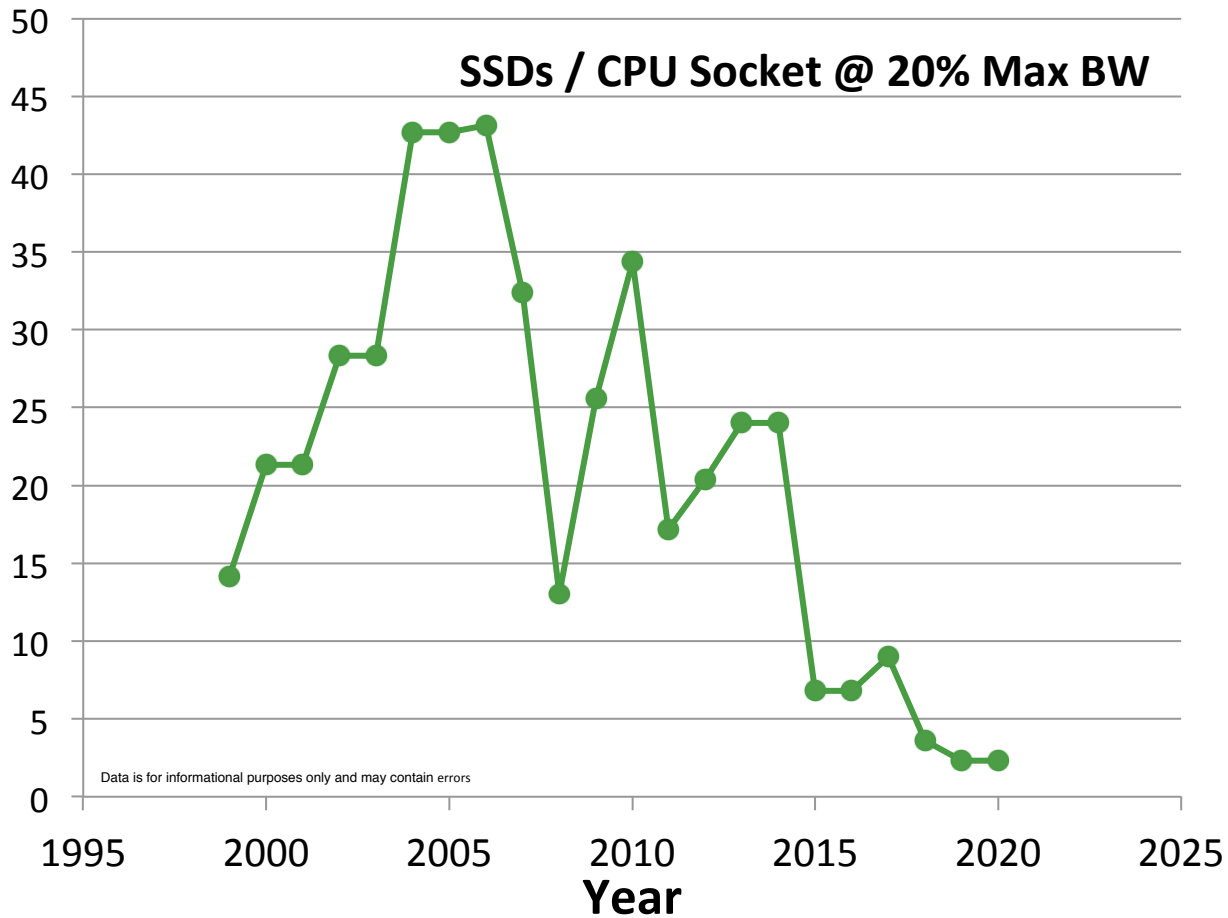
# Network, Storage and DRAM Trends

## Linear scale

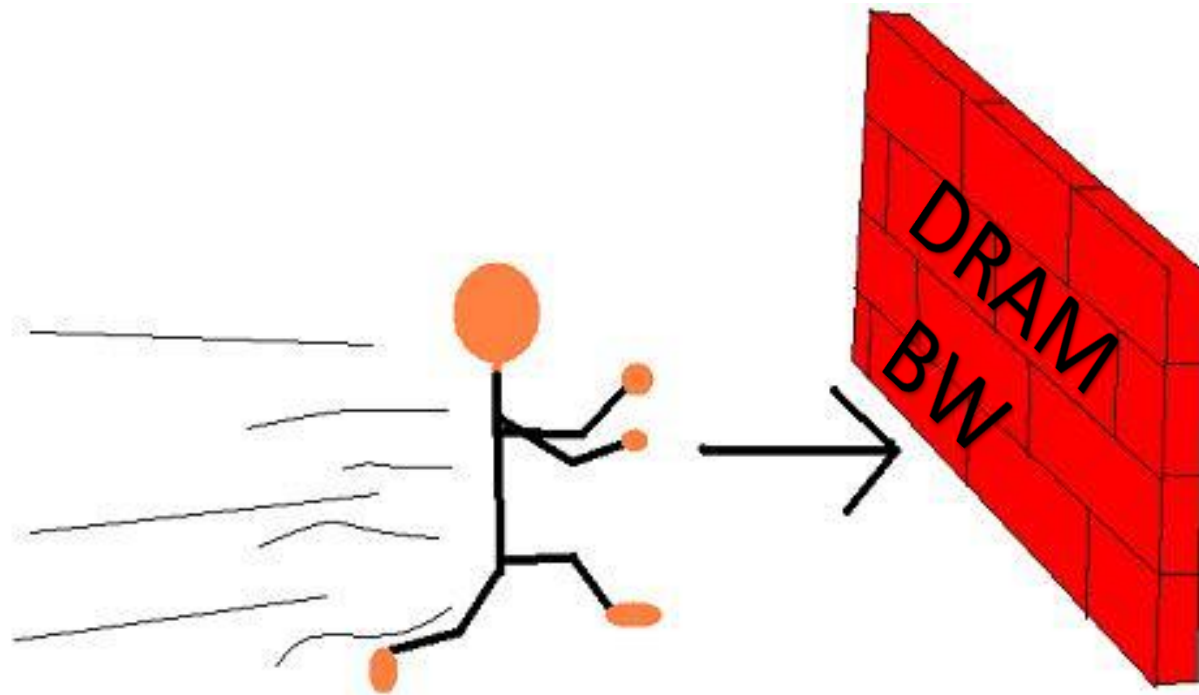
- Same data as last slide, but for the Log-impaired
- Storage Bandwidth is not literally infinite
- But the *ratio* of Network and Storage to CPU throughput is widening very quickly







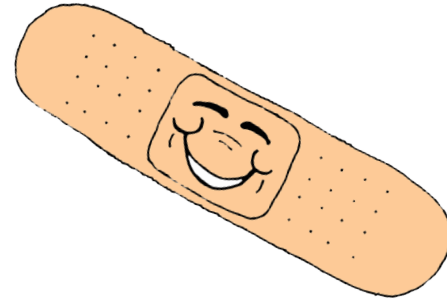
# What happens as we get closer to the limit?





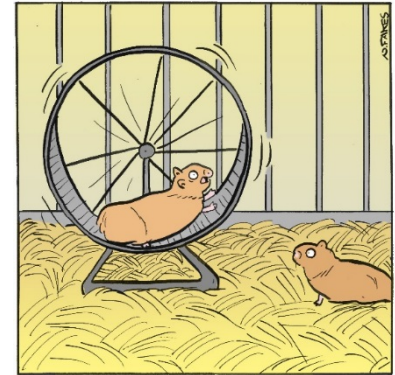
# Let's Get Small!

- New Denser Server Form Factors
  - Blades
  - Sleds
- Good short term solutions



# Effects Of The CPU/DRAM Bottleneck

- Storage Cost = Media + Access + Management
- Shared nothing architecture conflates access and management
- Storage costs will become dominated by Management cost
- Storage costs become CPU/DRAM costs



*"Sometimes I  
just feel like I'm not getting anywhere."*

# Embracing The CPU/DRAM Bottleneck

- Move management to upper layers where CPU can be right-sized by client
- What kind of media access do I want?
  - Simple enough functionality to be done directly in drive hardware – NO CPU
  - Allow direct access throughout the compute cluster over a network
  - Just enough machinery to enable coarse-grained sharing
- In short, you really want a SAN !
  - Or more technically, Fabric Connected Storage

what has been done will be again  
there is nothing new under the sun

EVERYTHING OLD  
IS NEW AGAIN

"IT'S LIKE  
DEJA VU  
ALL OVER  
AGAIN."  
YOKI REEDA

# Not Your Father's SAN

- Three problems with current SAN
  - Fibre channel transport
  - SCSI access protocol
  - Drive oriented storage allocation
- All of these want to be updated
  - Fibre channel is brittle and costly
  - SCSI initiators have long code paths catering to seldom used configurations
  - Robust sub-drive storage allocation

# SAN 2.0

Newer  
than New

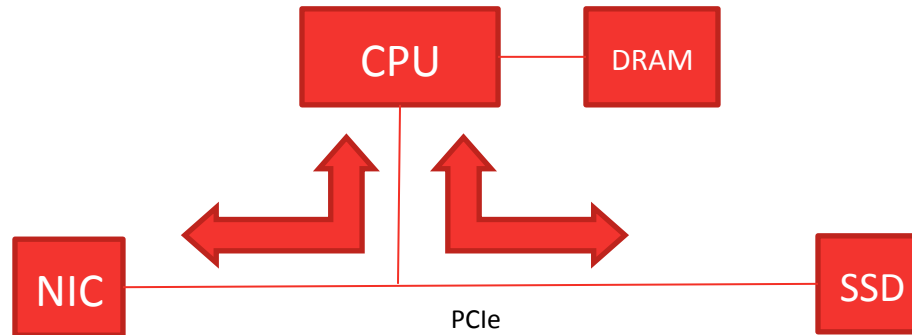
- NVMe over Fabrics
- 1.0 Spec is out for review, hopefully done in May
- Simple enough for direct hardware execution of data path ops
- Minimal initiator code path lengths improve performance
- Namespaces allow sub-drive allocations
- Not mature enough for enterprise deployment – yet

# SAN 2.0

- What storage network?
  - Current candidates are FC, Infiniband and Ethernet
- Ethernet has best economics – if you can make it work
- RoCE is easy on the edge, but hard on the interior
  - Only controlled environments have shown multi-switch scalability
  - General scalability in a multi-vendor environment likely to be difficult
  - Wonderful for intra-rack storage networking
- iWarp is hard on the edge, but easy on the interior
  - Scarcity of implementations inhibits deployment
- *Storage over IP will see limited cross rack deployment until this is resolved*

# First Generation Of SAN 2.0

- Implementations using OTS stuff are in progress
- Server side implementations look pretty conventional too
- 4-5 MIOPS have been shown
- Seems like 10 MIOPS isn't unreasonable to expect



# Second Generation SAN 2.0

- Soon, NICs will forward NVMe operations to local PCIe devices
- CPU removed from the *software* part of the data path
- CPU is still needed for the *hardware* part of the data path
- IOPS improve, BW is unchanged
- Significant CPU freed for application processing
- *Getting closer to the wall!*



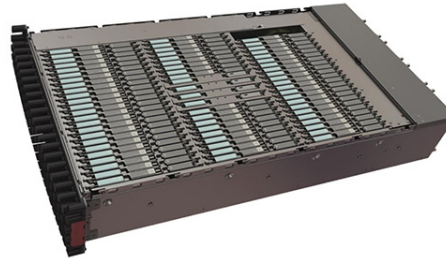
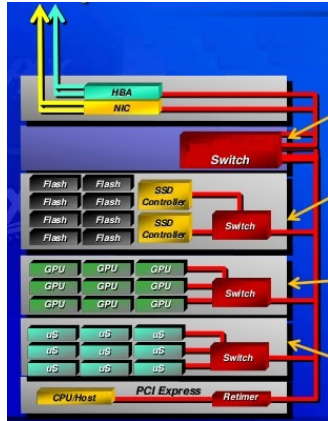
# Third Generation SAN 2.0, Imagined

- New generation of combined SSD controller and NIC
  - Rethink of interfaces eliminates DRAM buffering
- Network goes right into the drive
- No CPU to be found
- Works well with rack scale architecture



# Let's Get Really Small

- Disaggregated / Rack Scale Architecture
  - Fabric connected
  - Independently scale compute, networking and storage



# Call To Action

- Fabric-connected storage isn't well managed by existing FOSS
- Lots of upper layer management software is available
  - OpenStack, Ceph, Gluster, Cassandra, MongoDB, SheepDog, etc.
- Lower layer cluster management still primitive



# What's It All Mean?

- New form factors are in everybody's future
- The coming avalanche of storage bandwidth wants to be free
  - Not imprisoned by a CPU
- Rack Scale Architecture allows new Storage/Compute configs
- Storage will be increasingly “Software Defined” as the HW evolves

The SanDisk logo is positioned in the top-left corner of the image. It features the word "SanDisk" in a bold, red, sans-serif font, with a registered trademark symbol (®) to the upper right of the "k". The background of the entire slide is a low-angle, upward-looking shot of a modern glass skyscraper, with the sky visible through the central opening of the building's structure.

**SanDisk®**

**EXPANDING THE  
POSSIBILITIES OF STORAGE**

©2016 SanDisk Corporation. All rights reserved. SanDisk is a trademark of SanDisk Corporation, registered in the United States and other countries. Other brands mentioned herein are for identification purposes only and may be the trademarks of their holder(s).