Moore's Law and Big Data

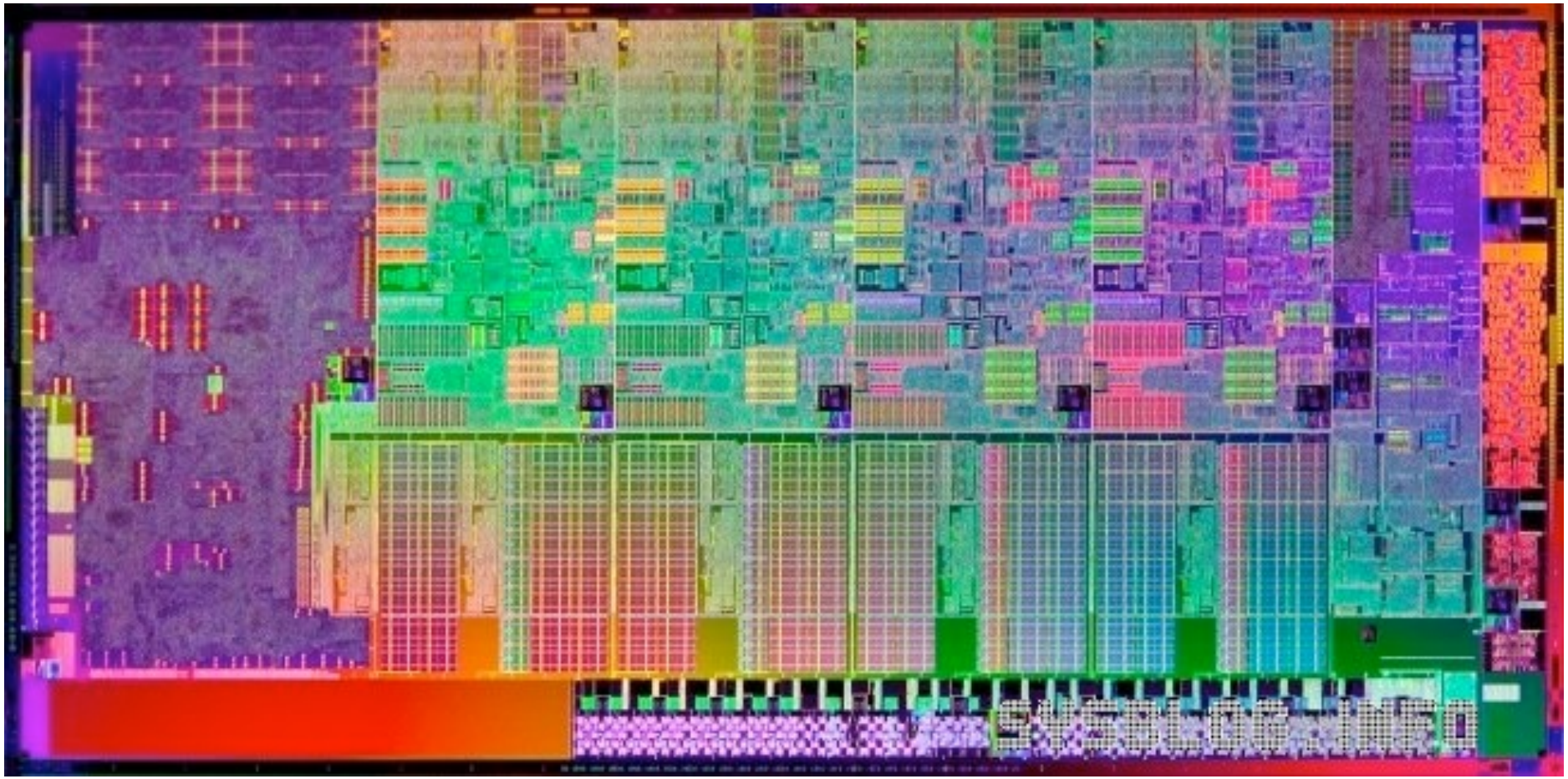
Andreas Bechtolsheim

Chairman Arista Networks
Co-Founder DSSD

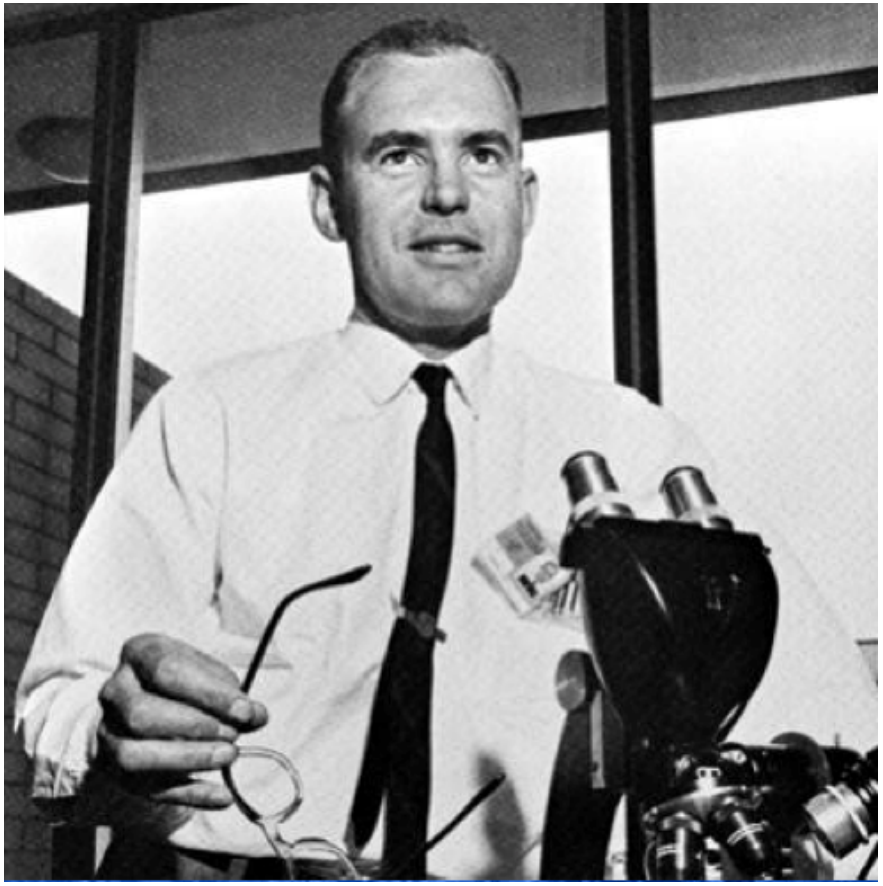
Agenda

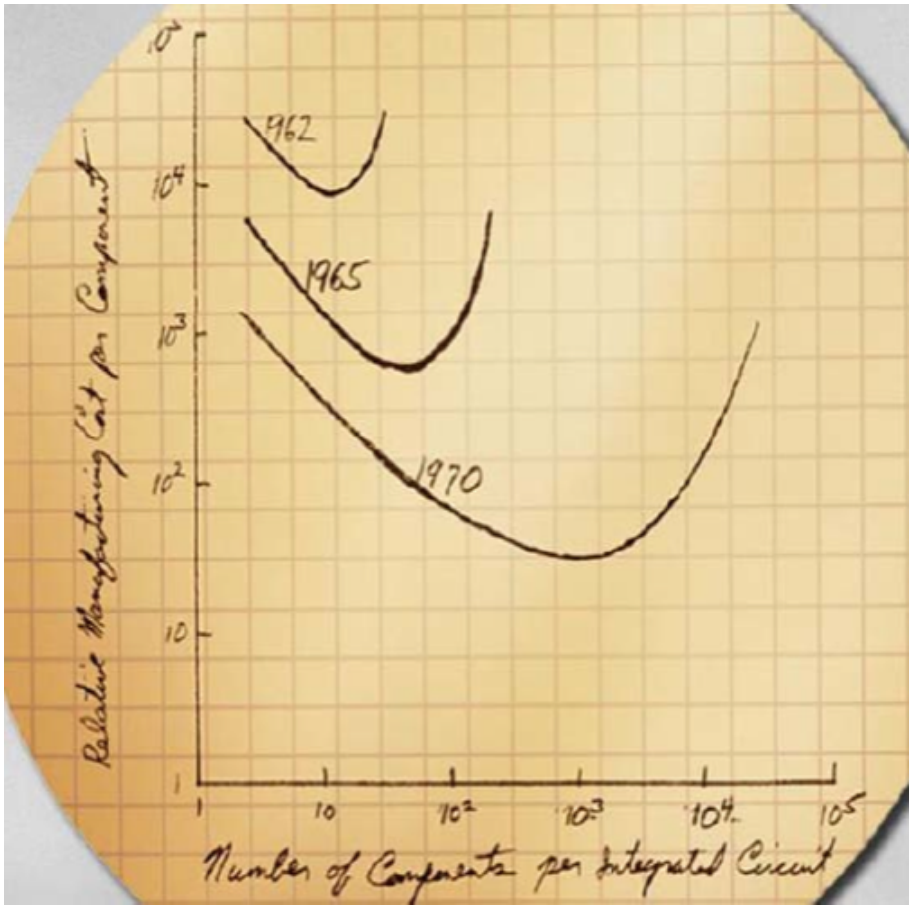
- Moore's Law
- Server Technology
- Storage Technology
- Big Data Architecture
- Use Cases

Moore's Law is Alive and Well



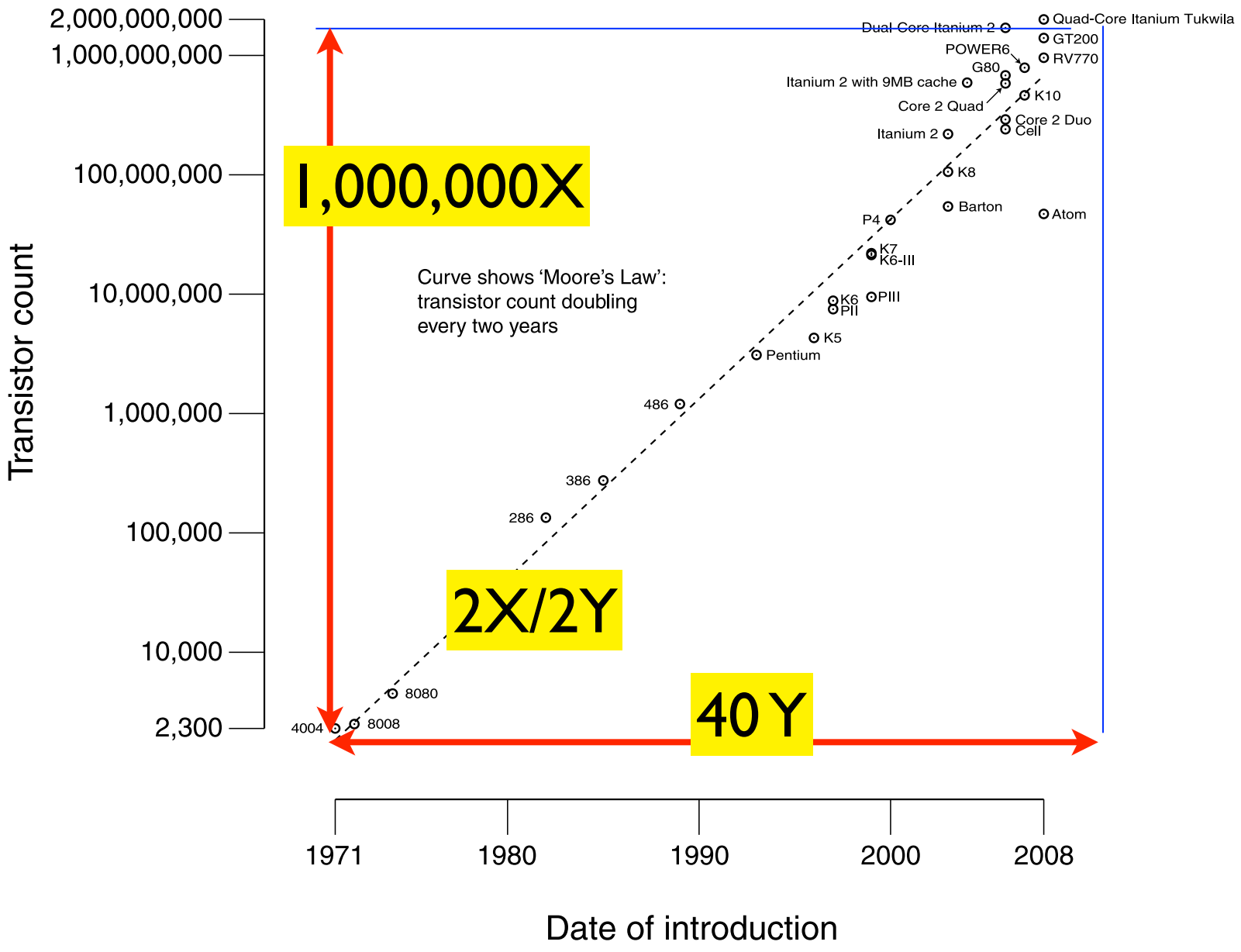
Dr. Moore's Prediction 1965



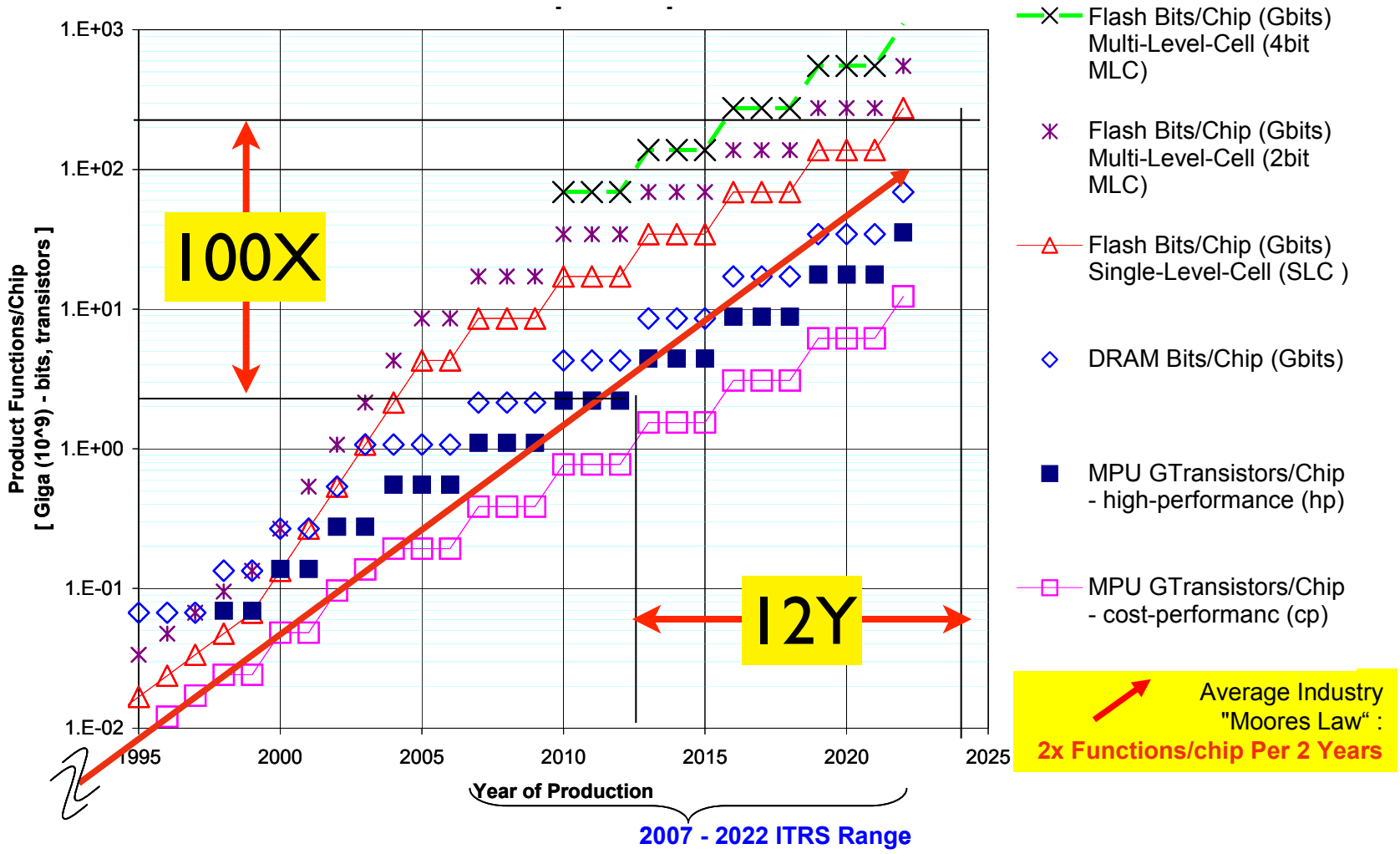


1965 Prediction: Density would double every year

Moore's Law 1971-2011: 2X/2Y



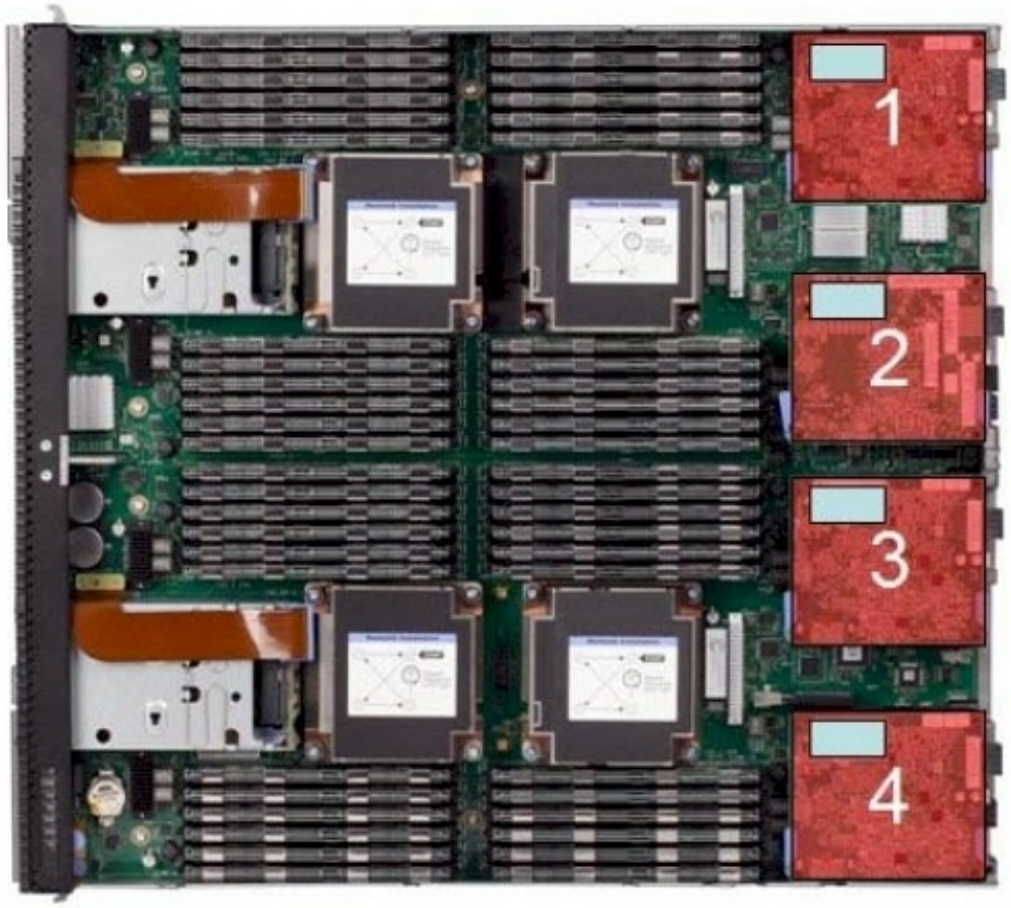
Silicon Roadmap 2012 - 2024



What can one do with 100 Billion Transistors per chip?

- **Lot's of CPU Cores**
  - 100s of high-speed cores per CPU chip
- **Lots of Memory Bits**
  - PB Flash, Multi-TB DRAM Systems
- **Lots of I/O and Network Bandwidth**
  - 10s of Gigabytes/sec per CPU

Intel 4S Server Board (2015)



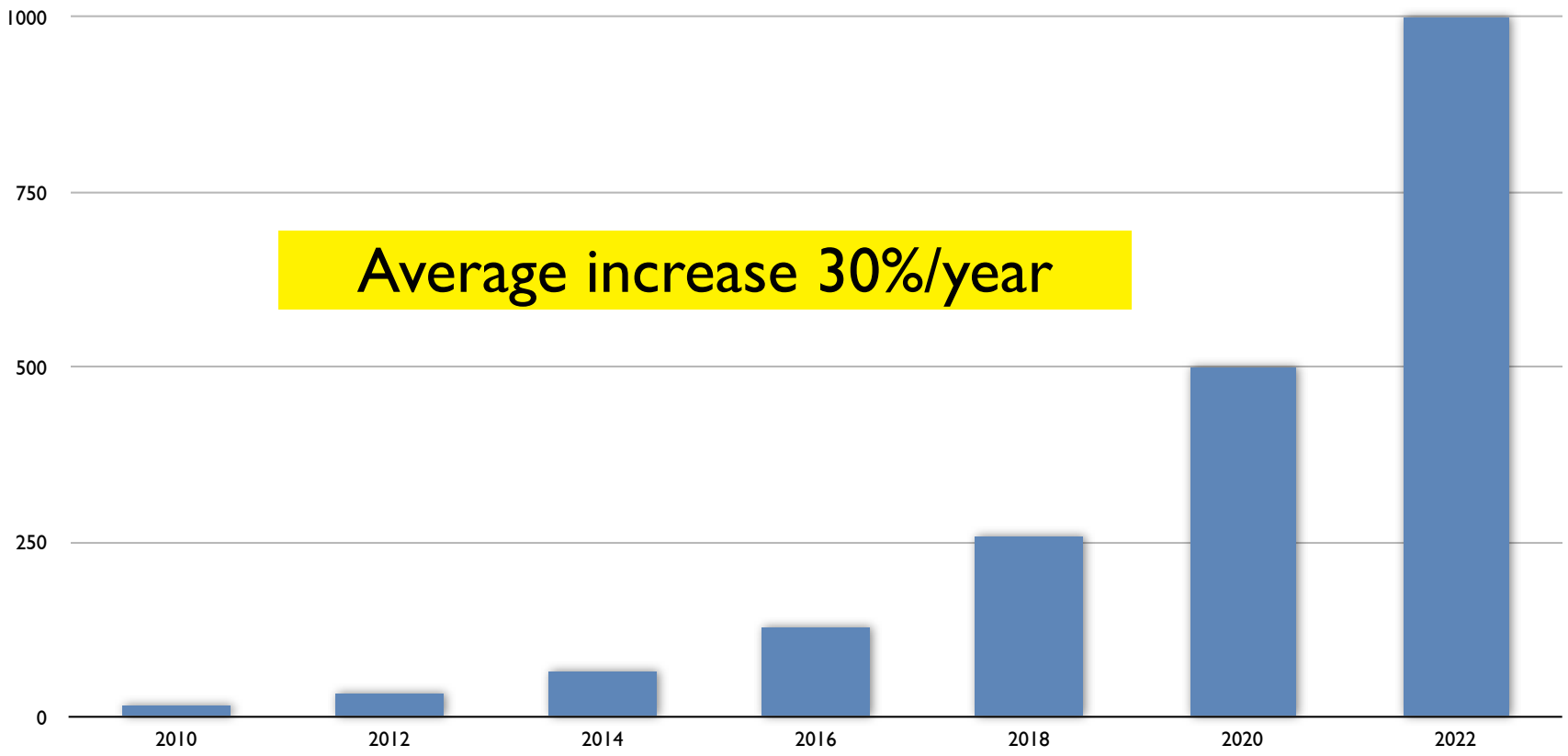
- 4 CPU Sockets
- 72 Cores
- 3+ GHz
- 48 DIMMs
- 3 TByte DRAM
- 1400W Power

SunFire E25K (2005)

- 72 SPARC Cores
- 1.5 GHz
- 1.15 TByte DRAM
- 76”H, 33”W, 64”D
- 2900 lbs
- 28.7 KWatt

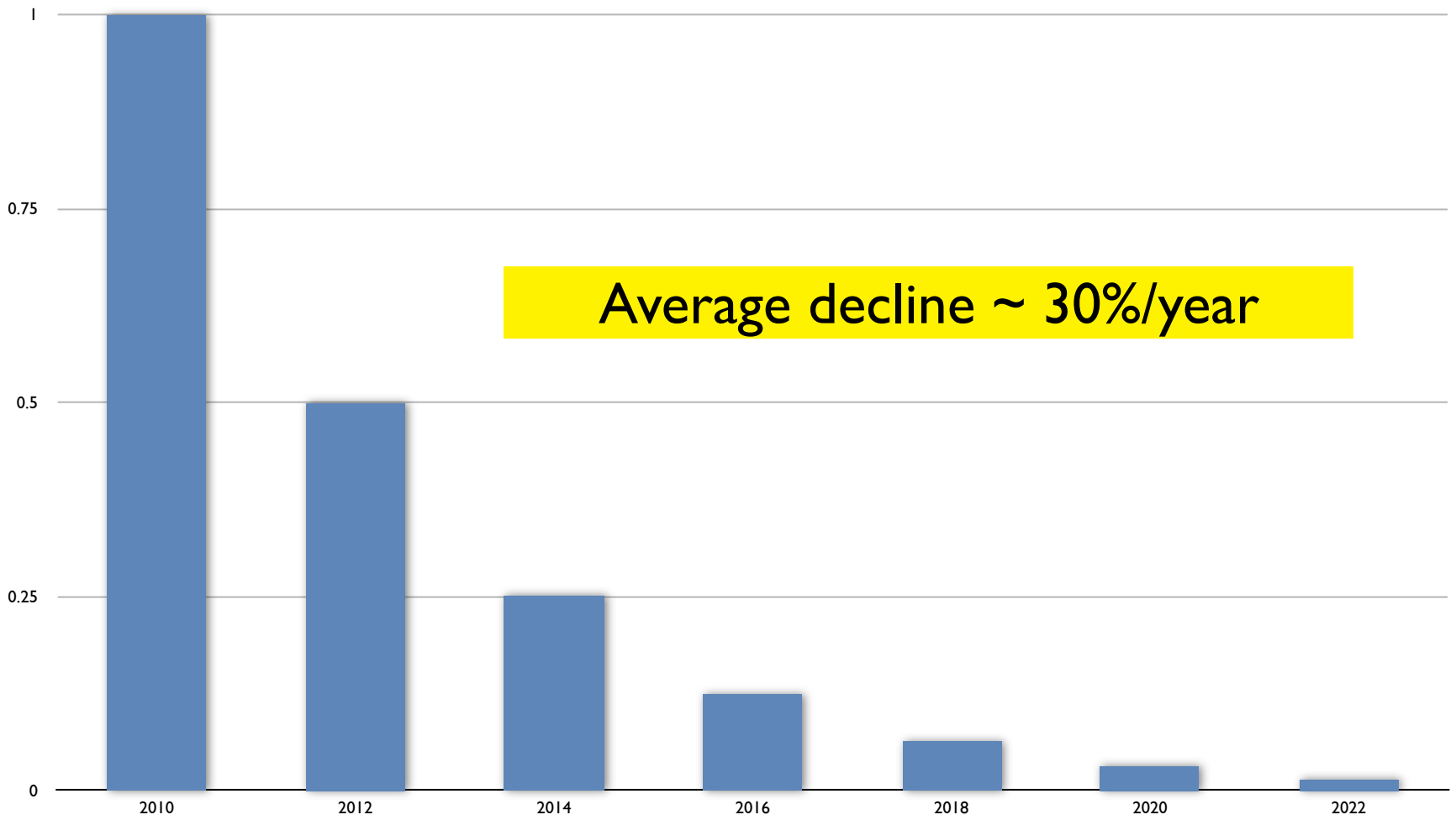
Less throughput than today’s single board 4S X86 Server

64-bit CPU Cores over Time



100X Performance Gain from 2012 to 2024

Declining Cost of Computing



Moore's Law Summary

- **Moore's Law 1971-2024**
  - Density has increased 2X every 2 Years
  - Million-fold improvement 1971-2011
  - Another 100X expected 2012-2024
  - 100 Million Fold Increase 1971-2024
- **Moore's Law is not ending, but slowing down**
  - Economic and Physics limits on the horizon

Moore's Law Challenges

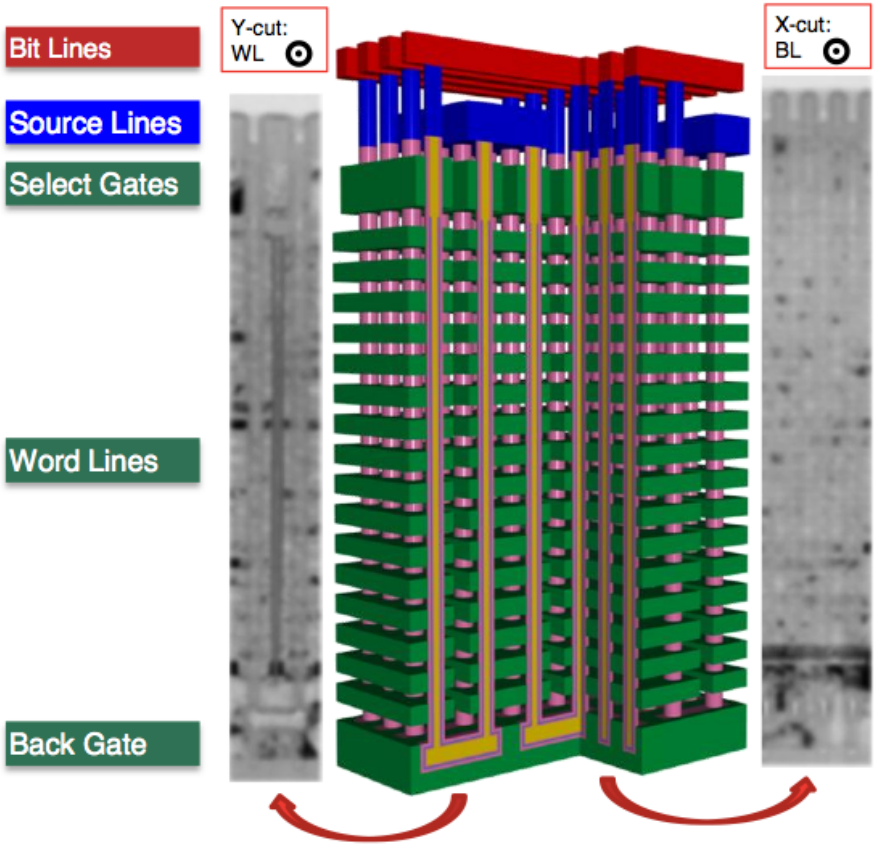
- **Economics are getting more Challenging**
  - Scaling is not "Free"
- **Geometric Scaling is Hitting Physics Limits**
  - 10nm, 7nm, 5nm, ???
- **Certain Things Don't Scale Well**
  - For example Flash Bits

However there are always creative solutions to problems

Flash Scaling Issue

- **Flash does not scale well below 15 nm**
  - Not enough electrons left to store bits
  - Charge loss due to shallow trapped electrons
  - Reduced write and read endurance
- **Solution: 3D Stacking**
  - Older process technology with many layers
  - Enables 256 Gbit (MLC), 512 Gbit (TLC)
  - More electrons per bit improves endurance

Toshiba BiCS 3D-NAND Stack



Note: Diagram not to scale

BiCS delivers smallest chip area of any published 3D-NAND

BiCS U-shaped NAND string enables maximum array efficiency

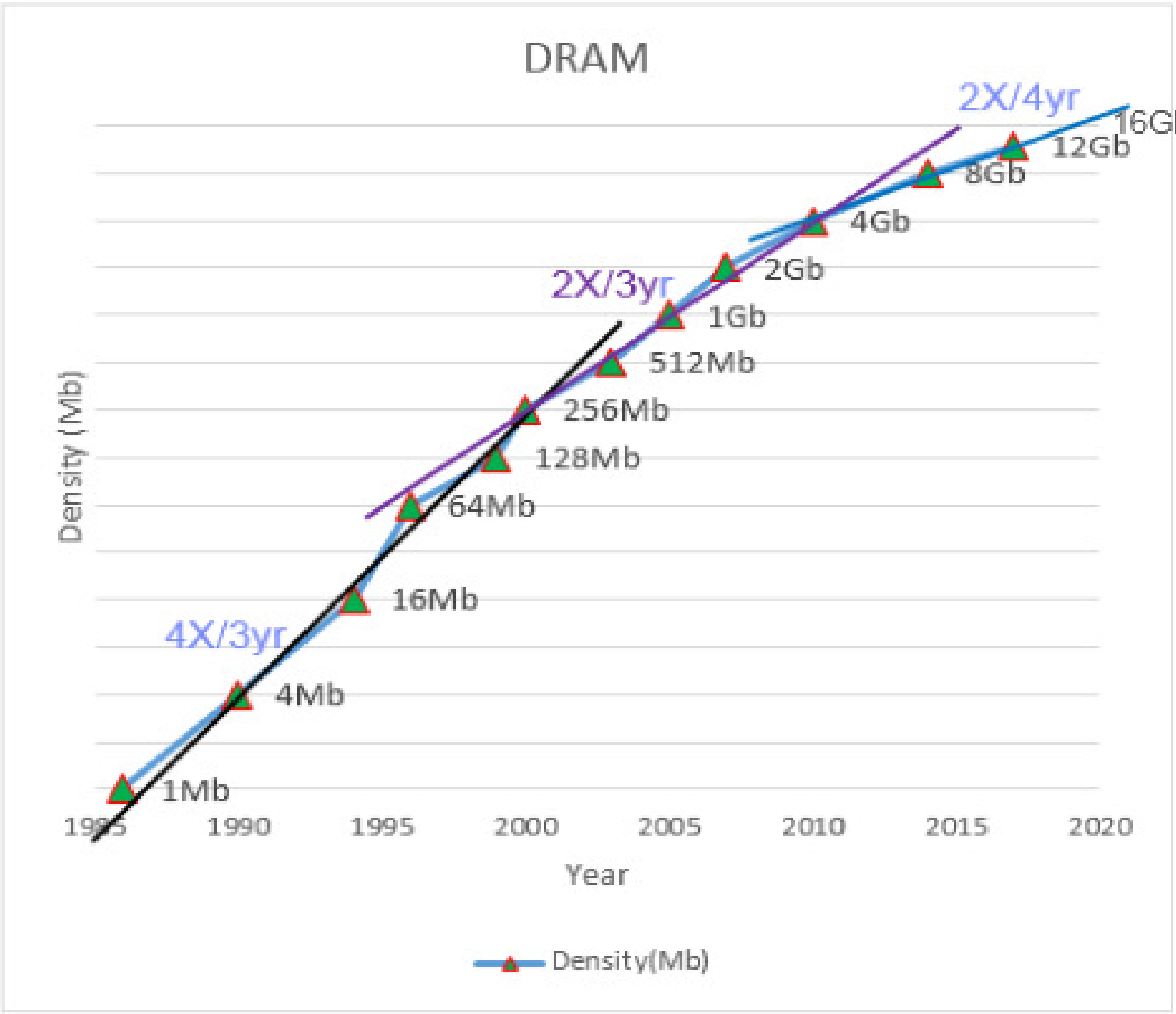
– Leverages existing NAND Fab infrastructure. Does not need EUV.
– Scaling achieved by increasing number of layers

Good progress in BiCS development

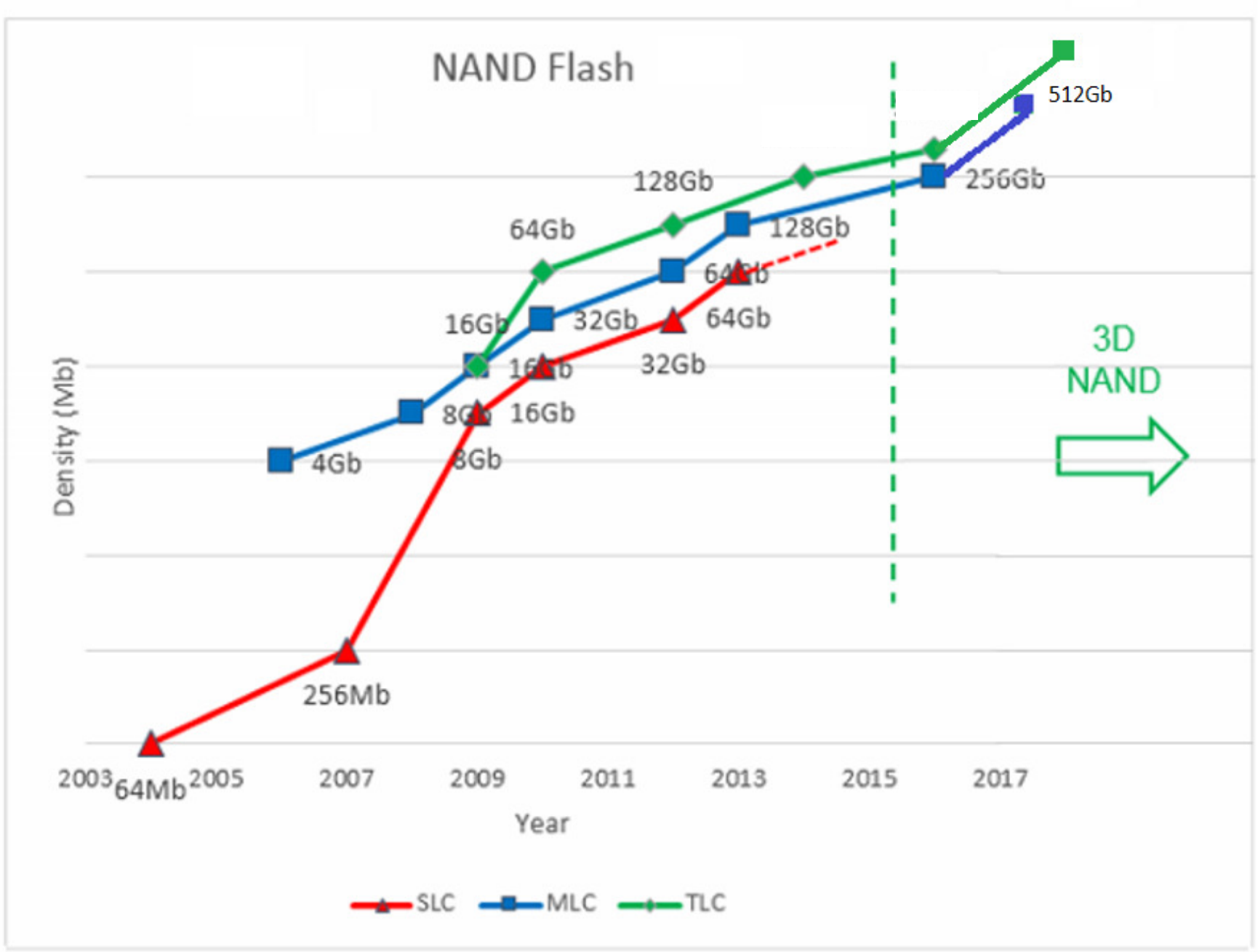
Challenges for all 3D-NAND manufacturing

– NAND poly TFT devices, a first in volume manufacturing
– High aspect ratio etching of large number of layers and its control
– High volume manufacturing requires new etching equipment and techniques for scaling to high number of layers

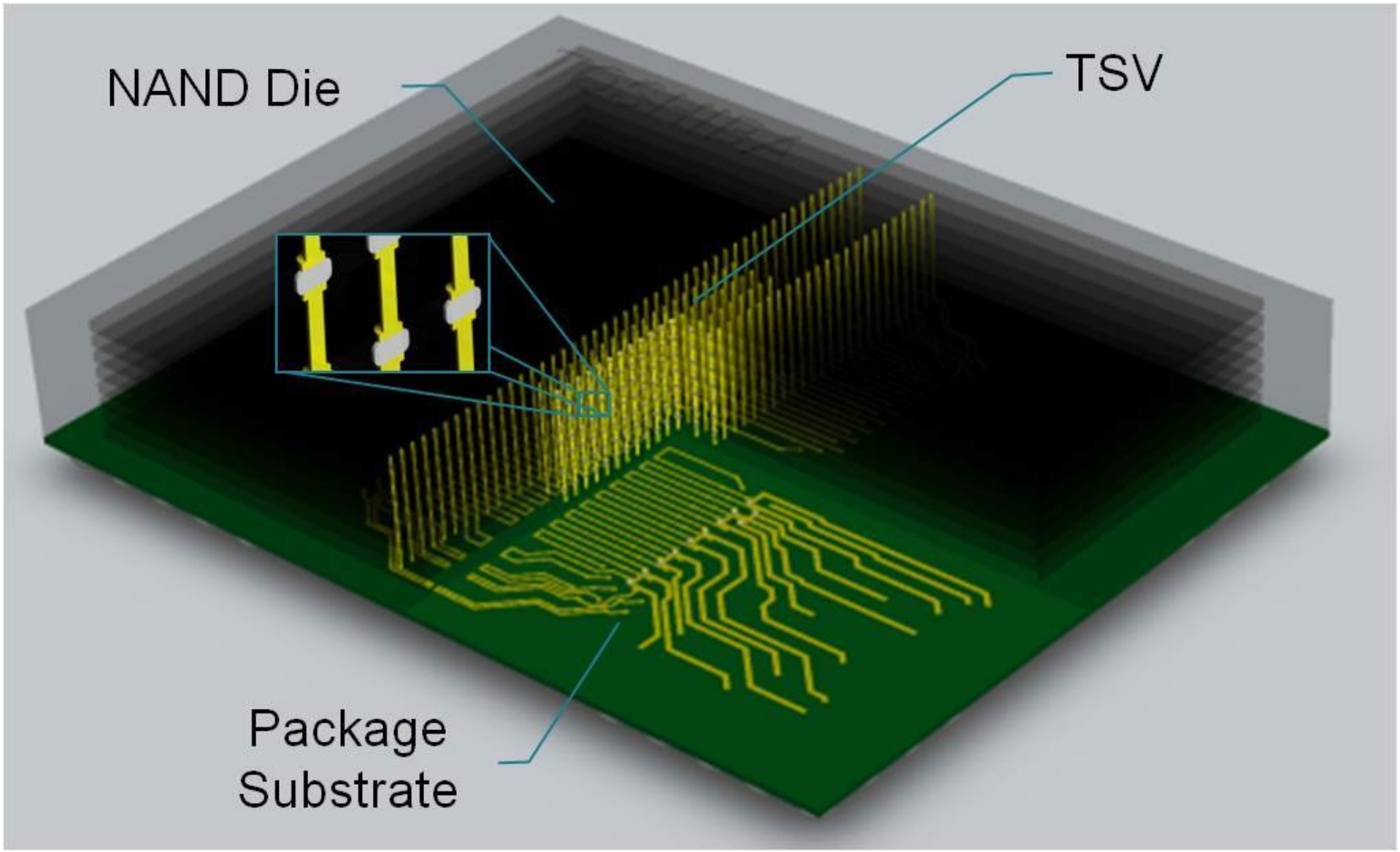
DRAM Memory Scaling



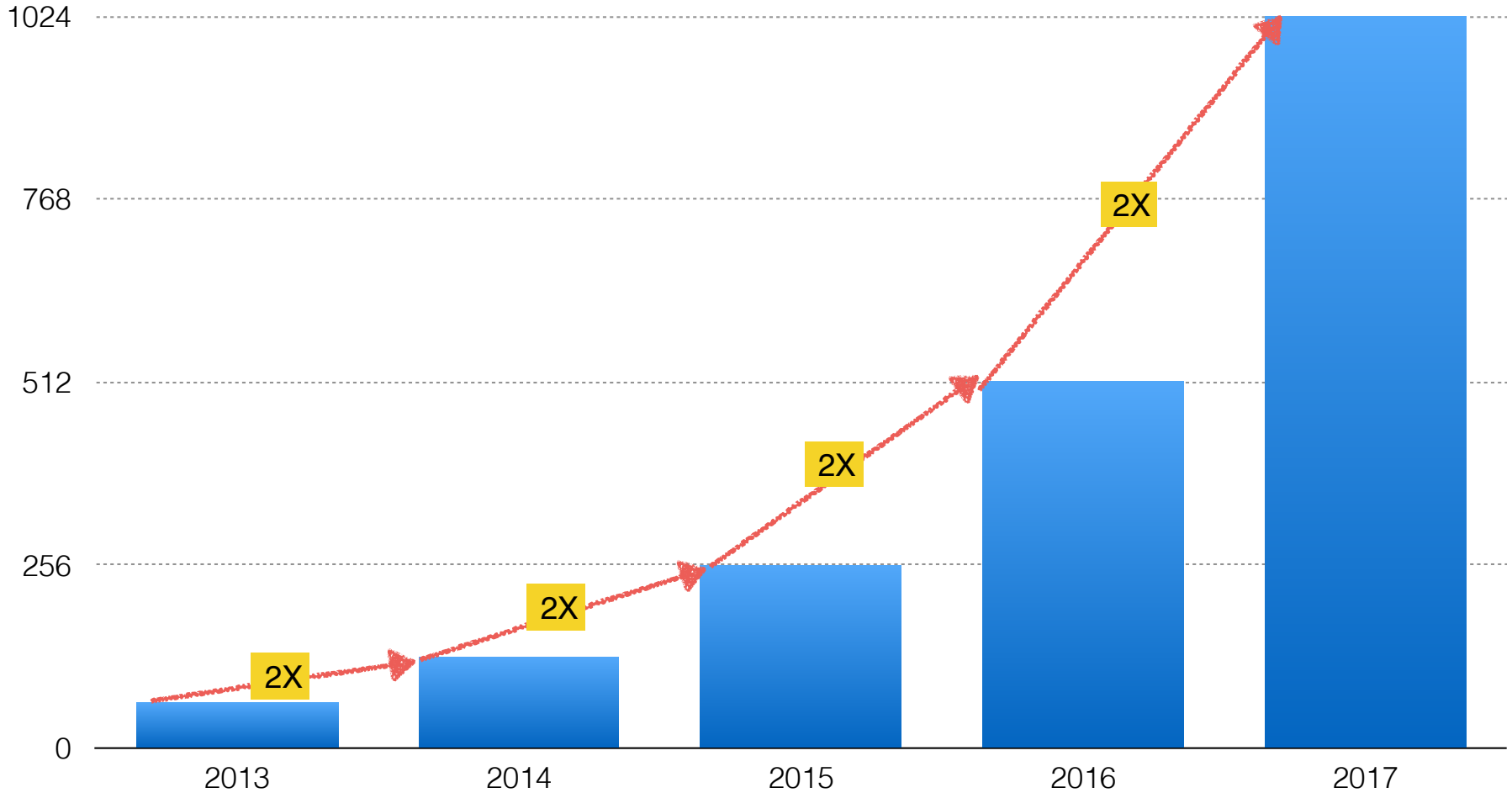
DRAM Memory Scaling



Up to 16 3D NAND Die/Package

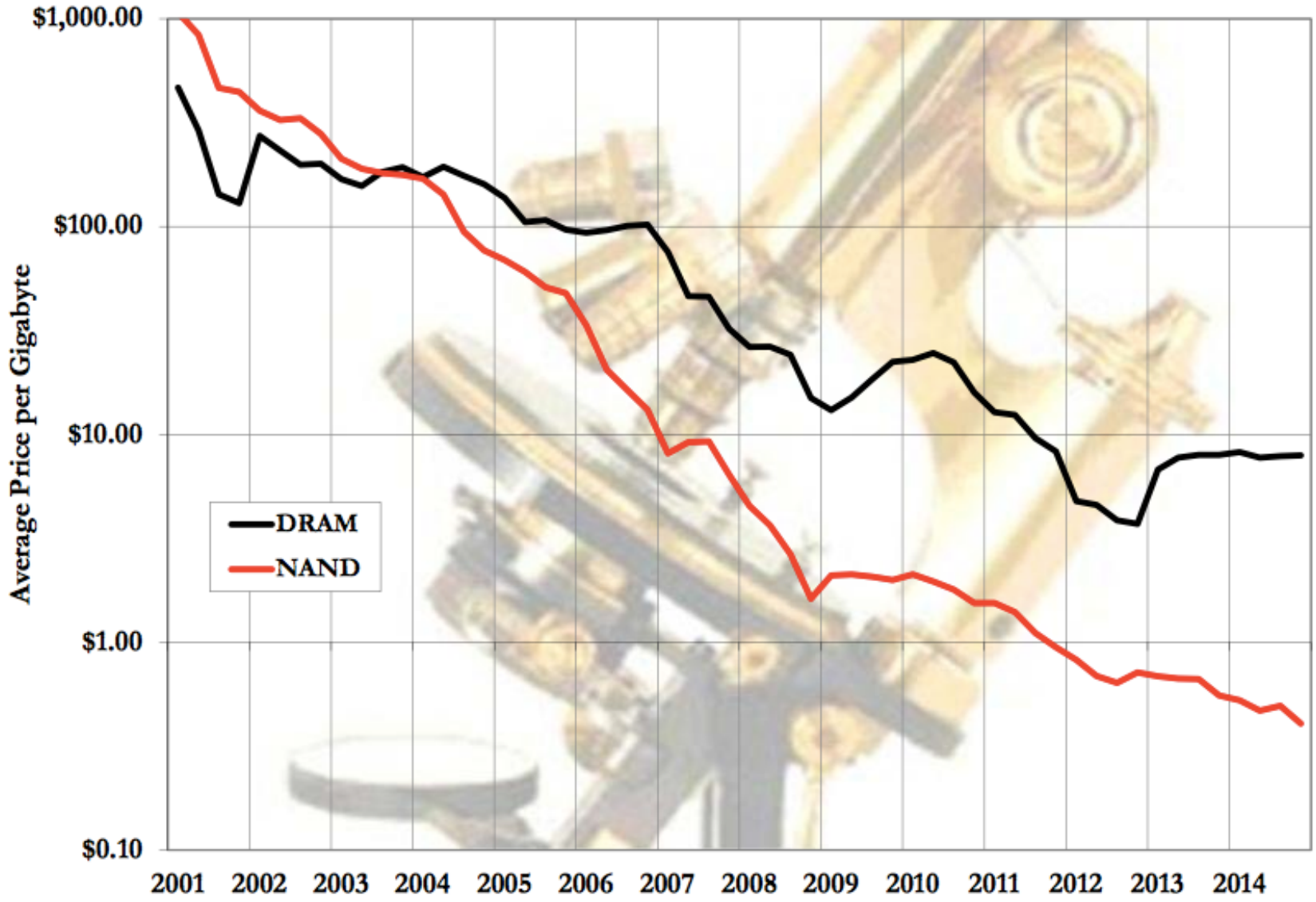


NAND Density/Package (GB)



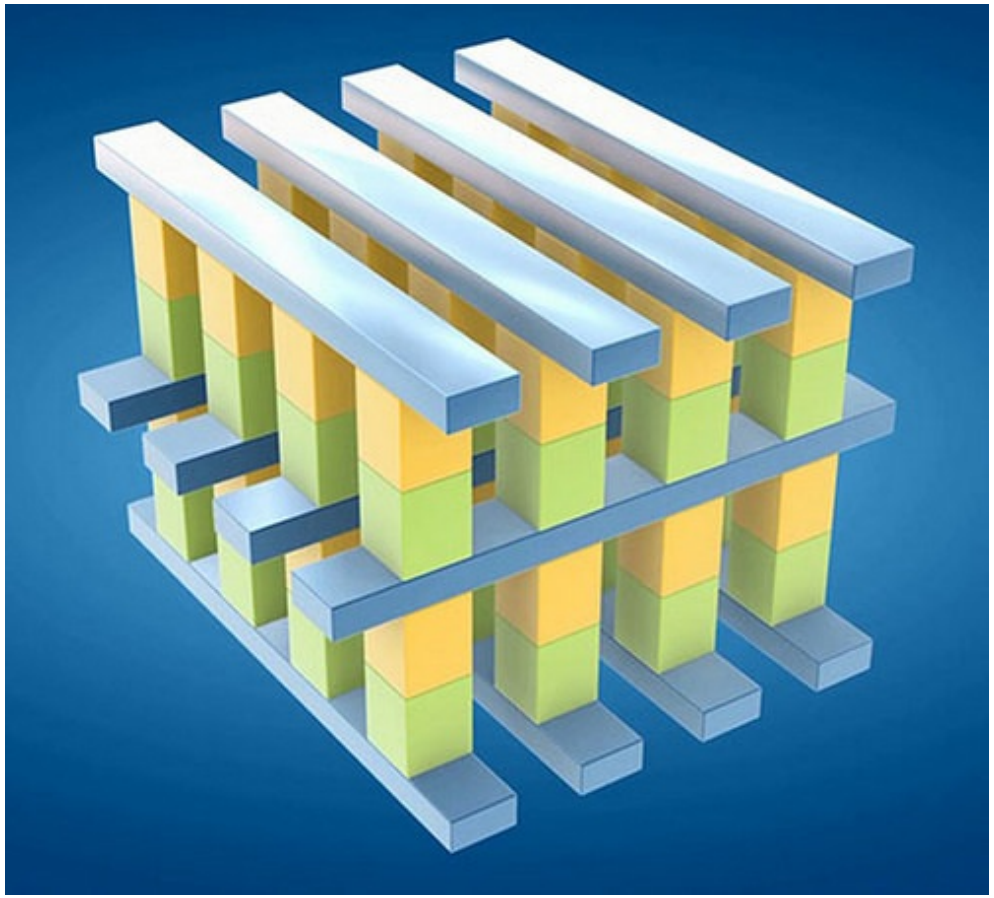
Combination of TLC, 3D NAND, 16 Stack Die/Package

DRAM vs FLASH Pricing



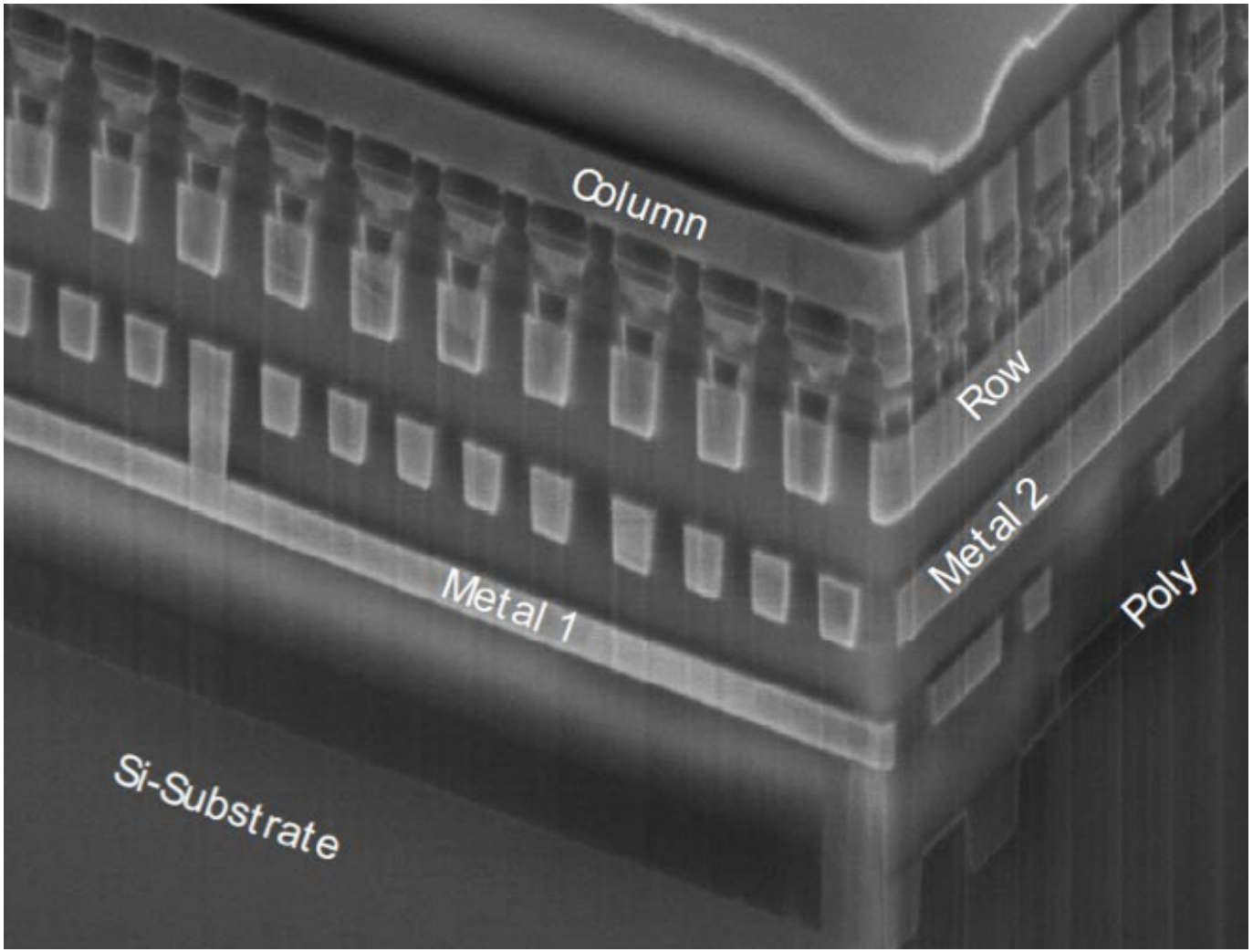
Source: ***A Close Look at the Intel/Micron 3D XPoint Memory***, Objective Analysis 2015

Intel 3D X-Point NV Memory



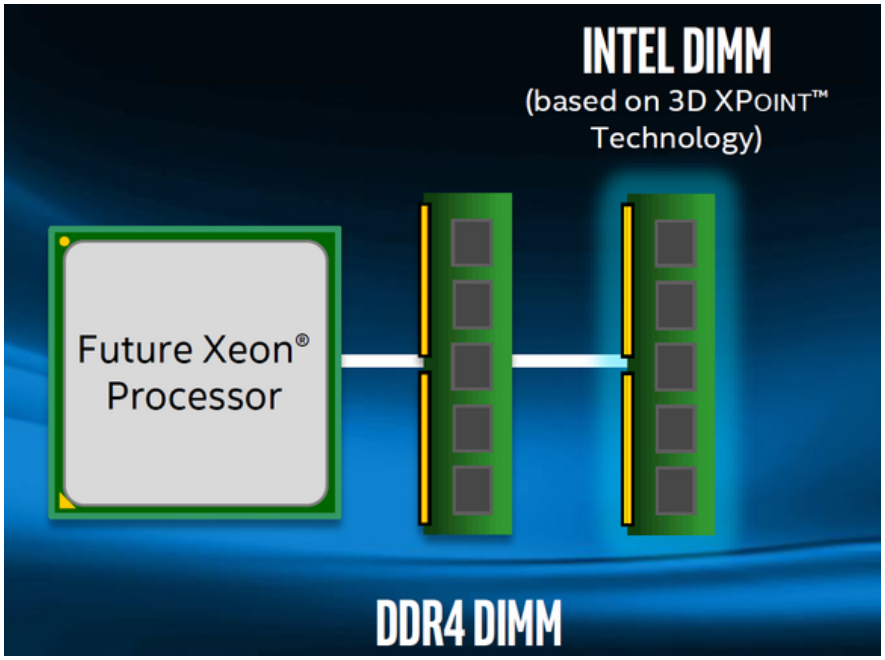
- **Read Performance**
  - Comparable to DRAM
- **Write endurance**
  - Much better than Flash but not unlimited (10^7)
- **Density**
  - Higher than DRAM but lower than Flash
- **Cost**
  - Lower than DRAM but higher than Flash

Intel 3DXP The Real Thing



Intel has researched PCM for 45 Years!

Intel 3X Xpoint Form Factors

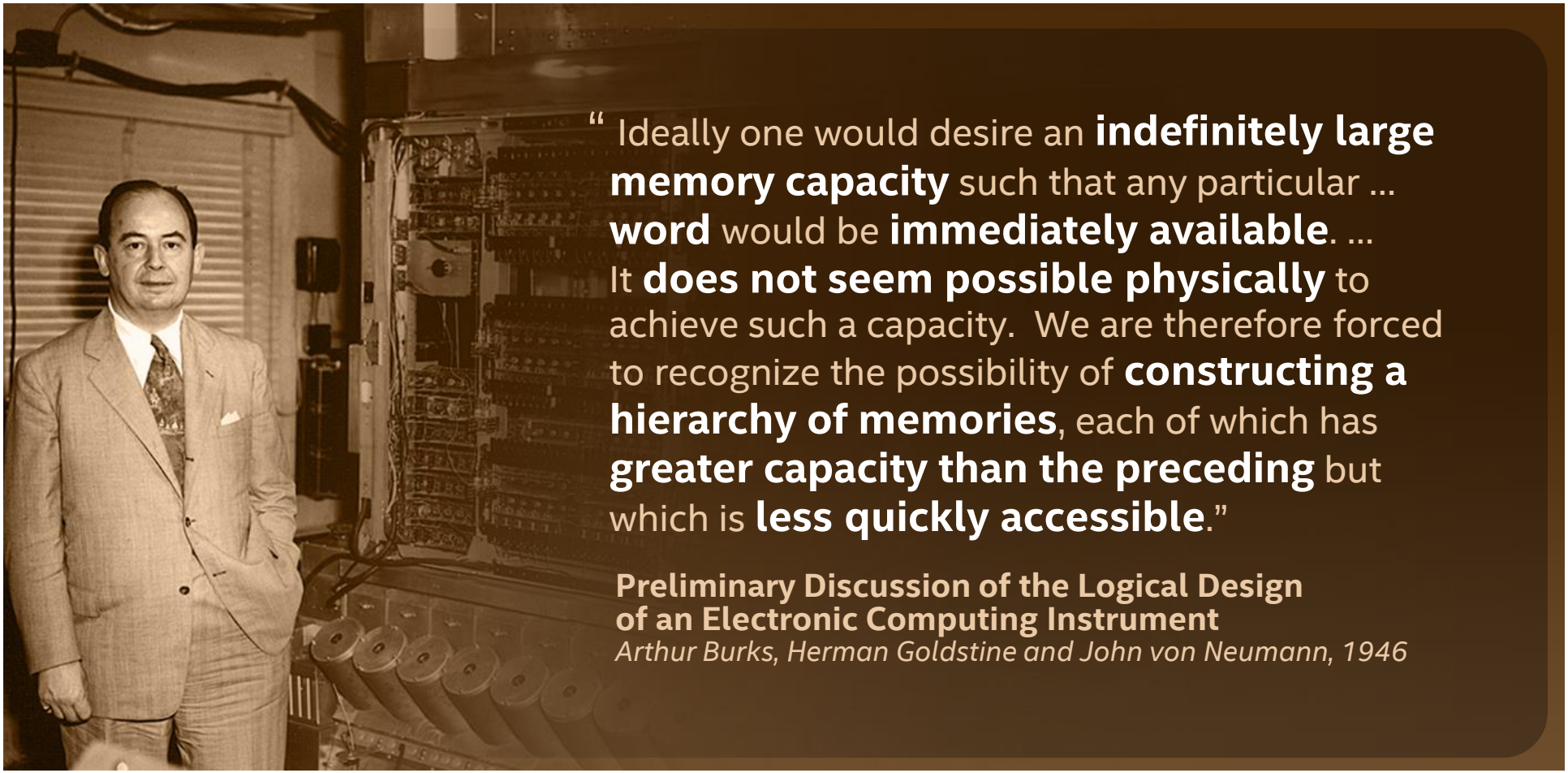


DDR4 DIMM Formfactor
Supports mixed DRAM/3DXP Memory Configuration
Up to 512 MB/DIMM

NVMe PCIe FormFactor
Supports High-Speed Storage Applications
Up to 4 TB/Container

How to Build Very High-performance Scalable Storage Compute Systems

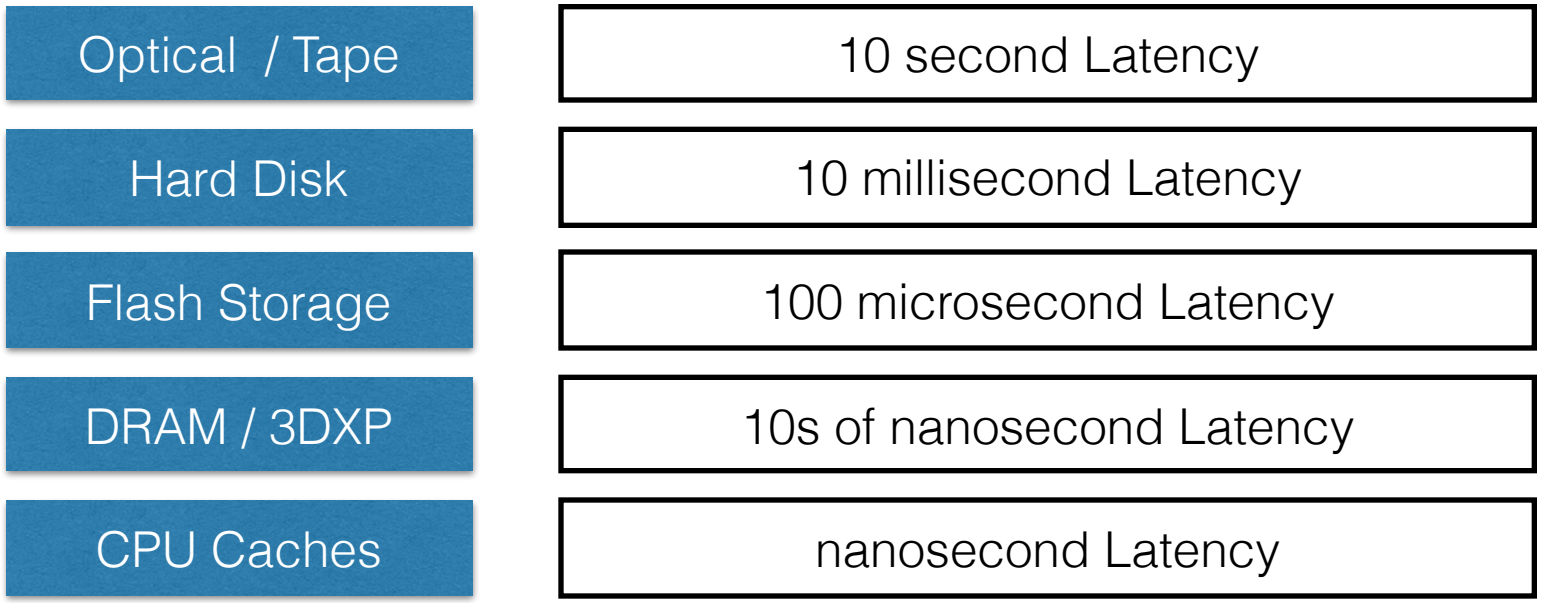
The Memory Storage Hierarchy



" Ideally one would desire an **indefinitely large memory capacity** such that any particular ... **word** would be **immediately available**. ... It **does not seem possible physically** to achieve such a capacity. We are therefore forced to recognize the possibility of **constructing a hierarchy of memories**, each of which has **greater capacity than the preceding** but which is **less quickly accessible**."

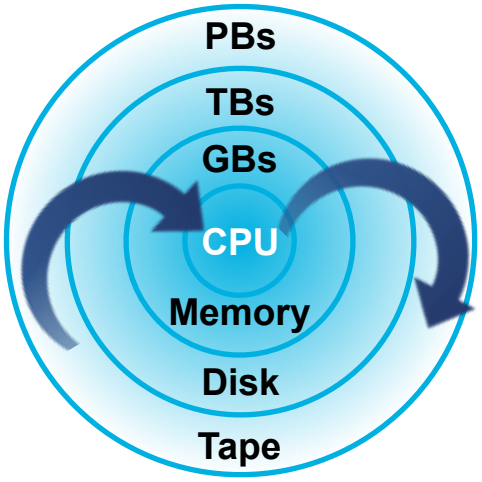
Preliminary Discussion of the Logical Design of an Electronic Computing Instrument
Arthur Burks, Herman Goldstine and John von Neumann, 1946

Today's Memory Storage Hierarchy



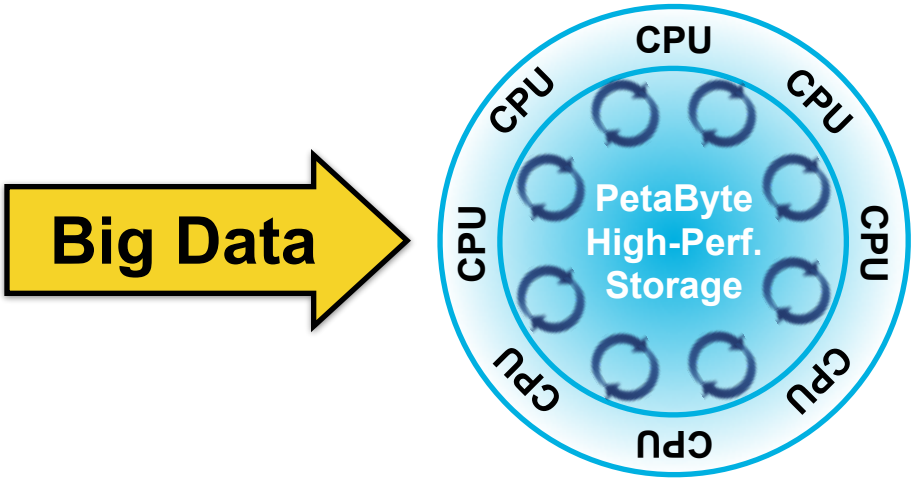
A New Data-centric Storage Model

Old Compute-Centric Model



- Data lives on Disk or Flash
- Deep Storage Hierarchy
- Apps use small % of data
- Legacy I/O Stacks (FC/SCSI)

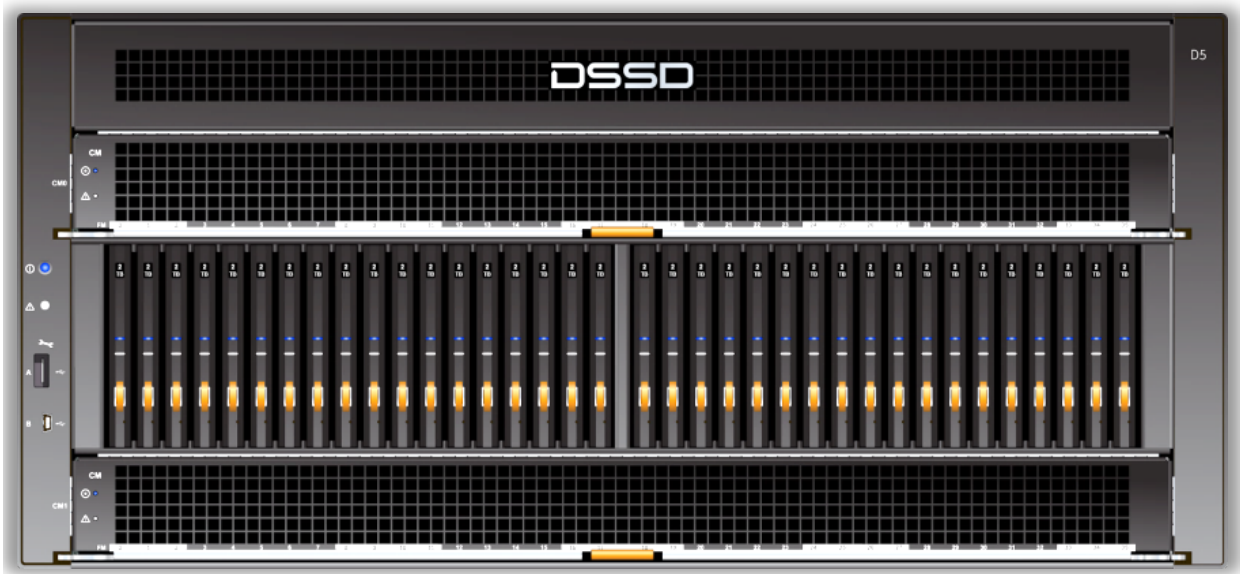
New Data-centric Model



- Data lives in persistent memory
- Many CPUs surround and use it
- Big Data Apps use entire datasets
- Need to rethink I/O Stack

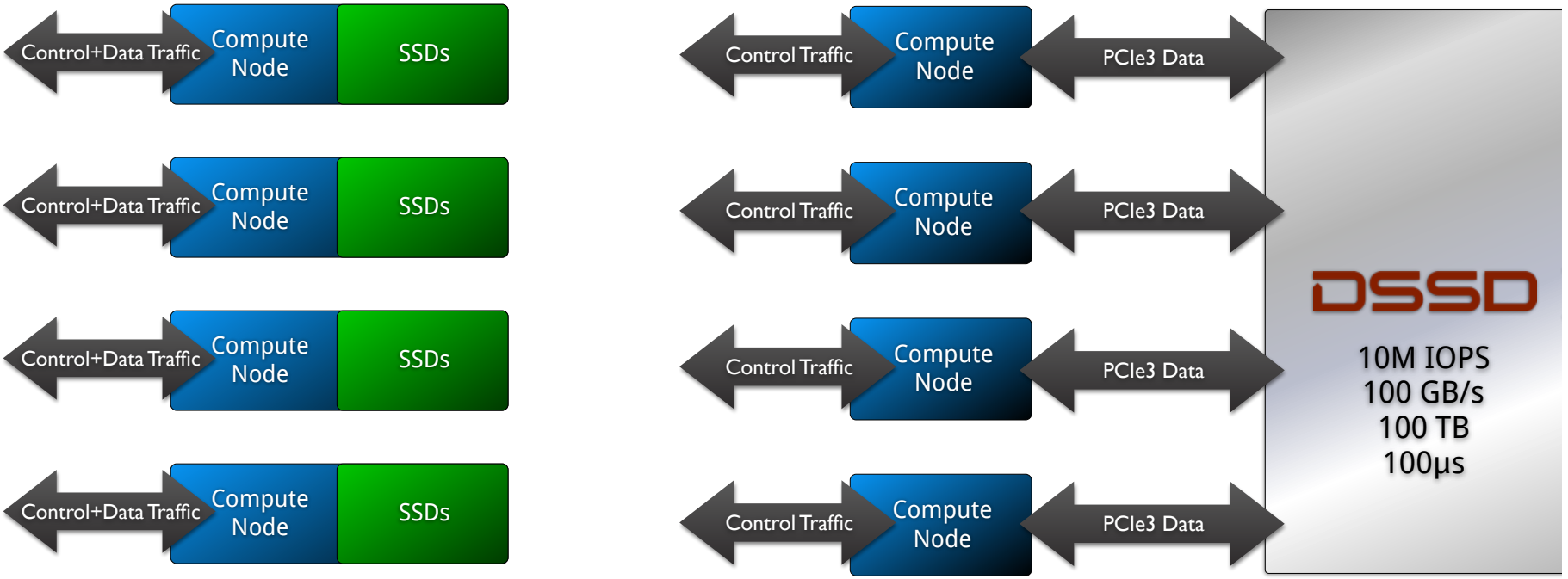
Big Data Apps require rethinking of storage architecture

Rack-scale DSSD Flash Storage



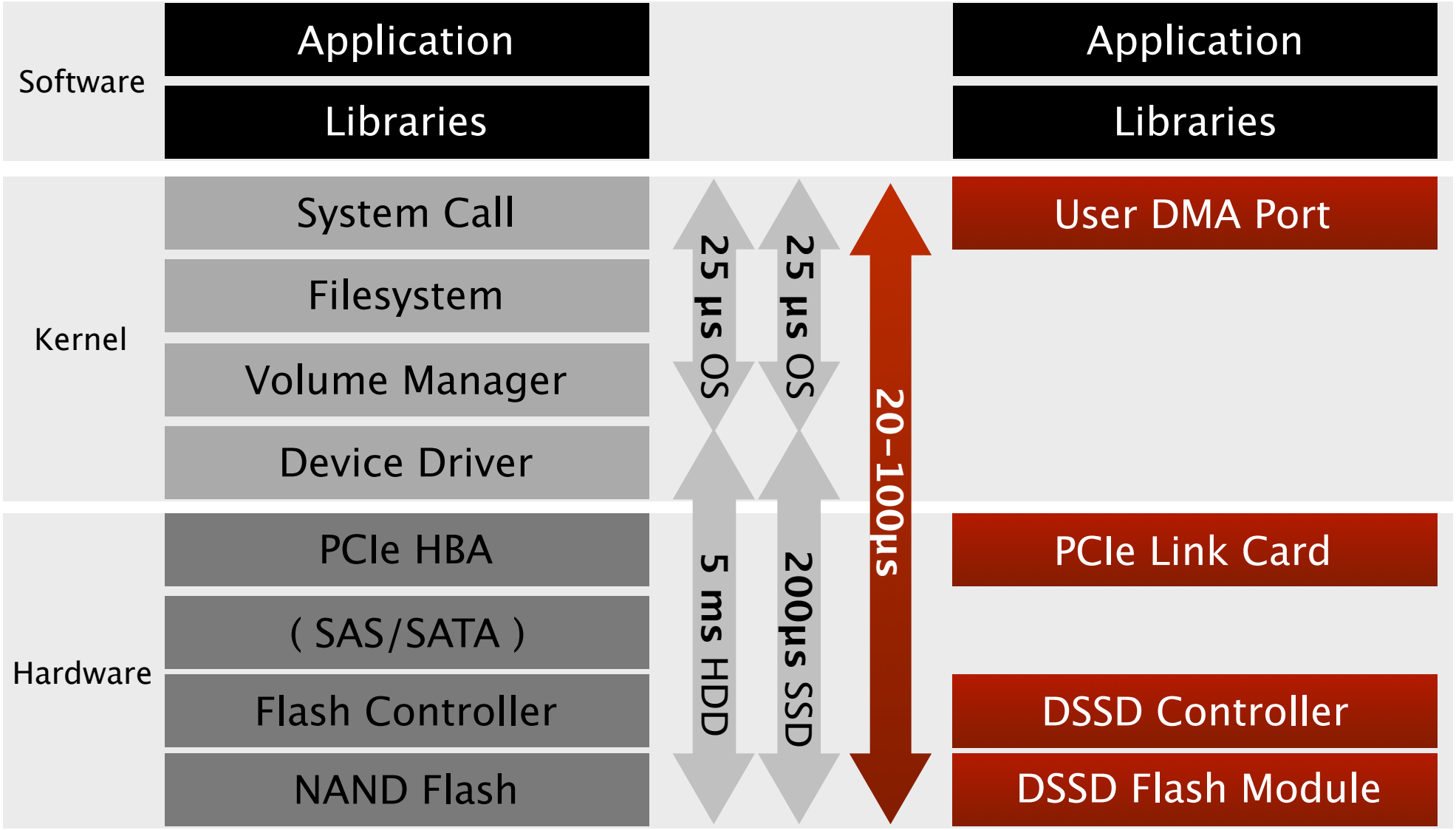
10 Million IOPS, 100 GByte/sec IO
100 usec Latency, 100 TByte Storage

Rack-Scale Shared NVMe Storage

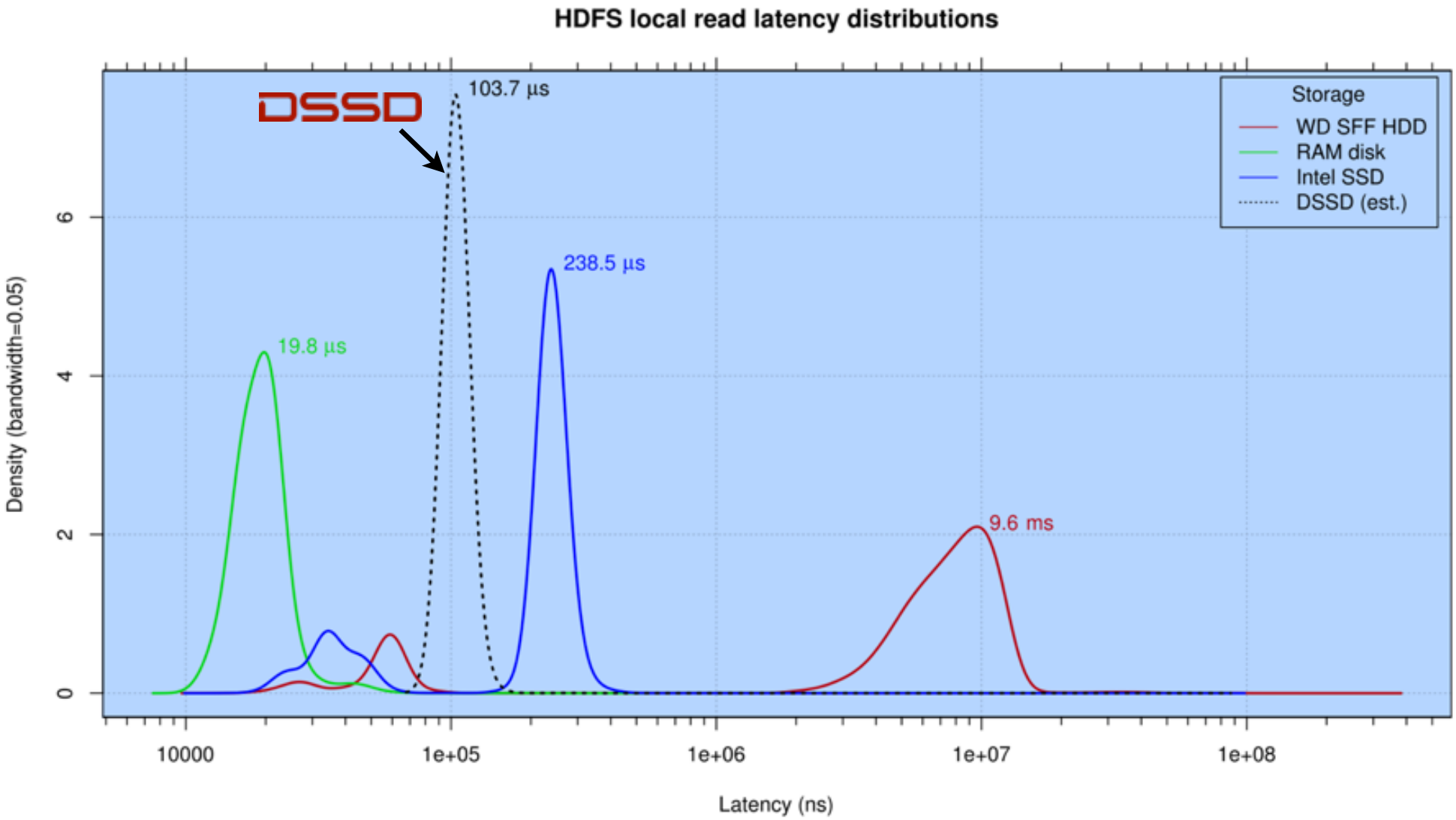


Rack-scale NVMe Storage is bigger, faster, and more reliable than local Flash Storage

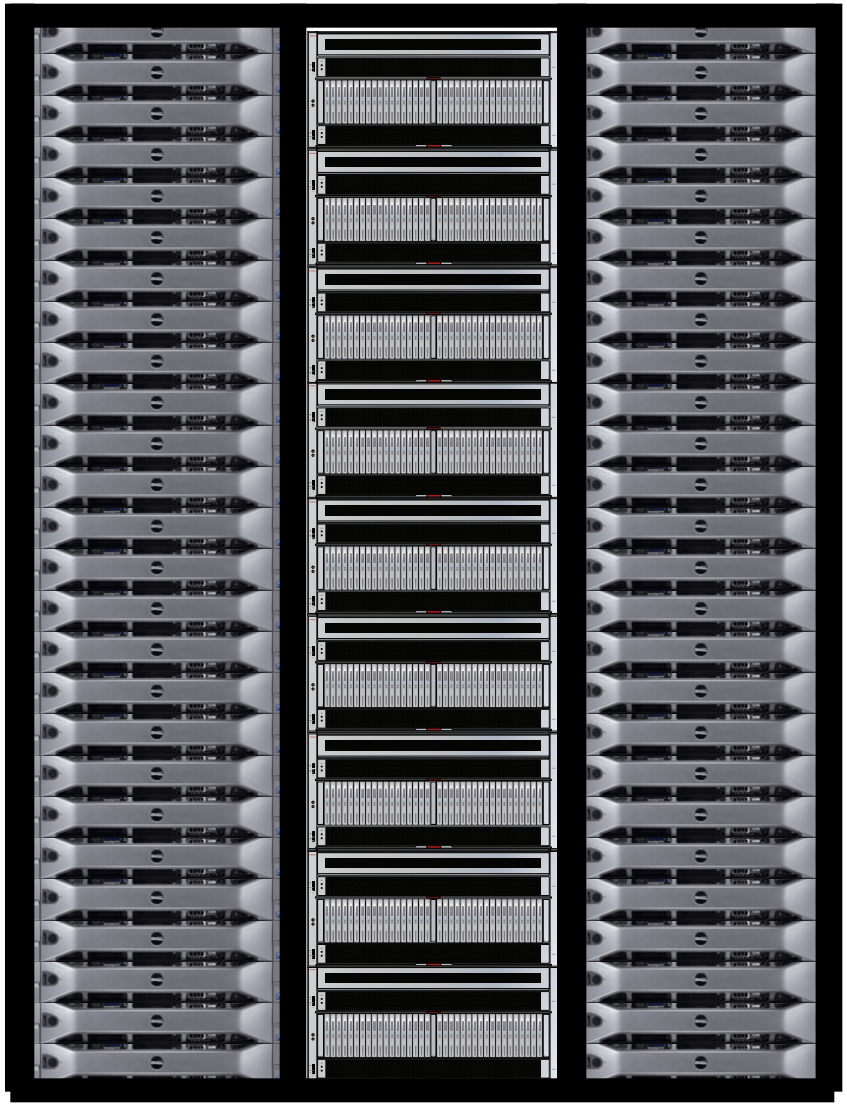
Rack-scale Direct Access Storage



100 usec Latency, All the Time



NSF/TACC DELL/DSSD Storage Cluster



Petabyte Flash Capacity

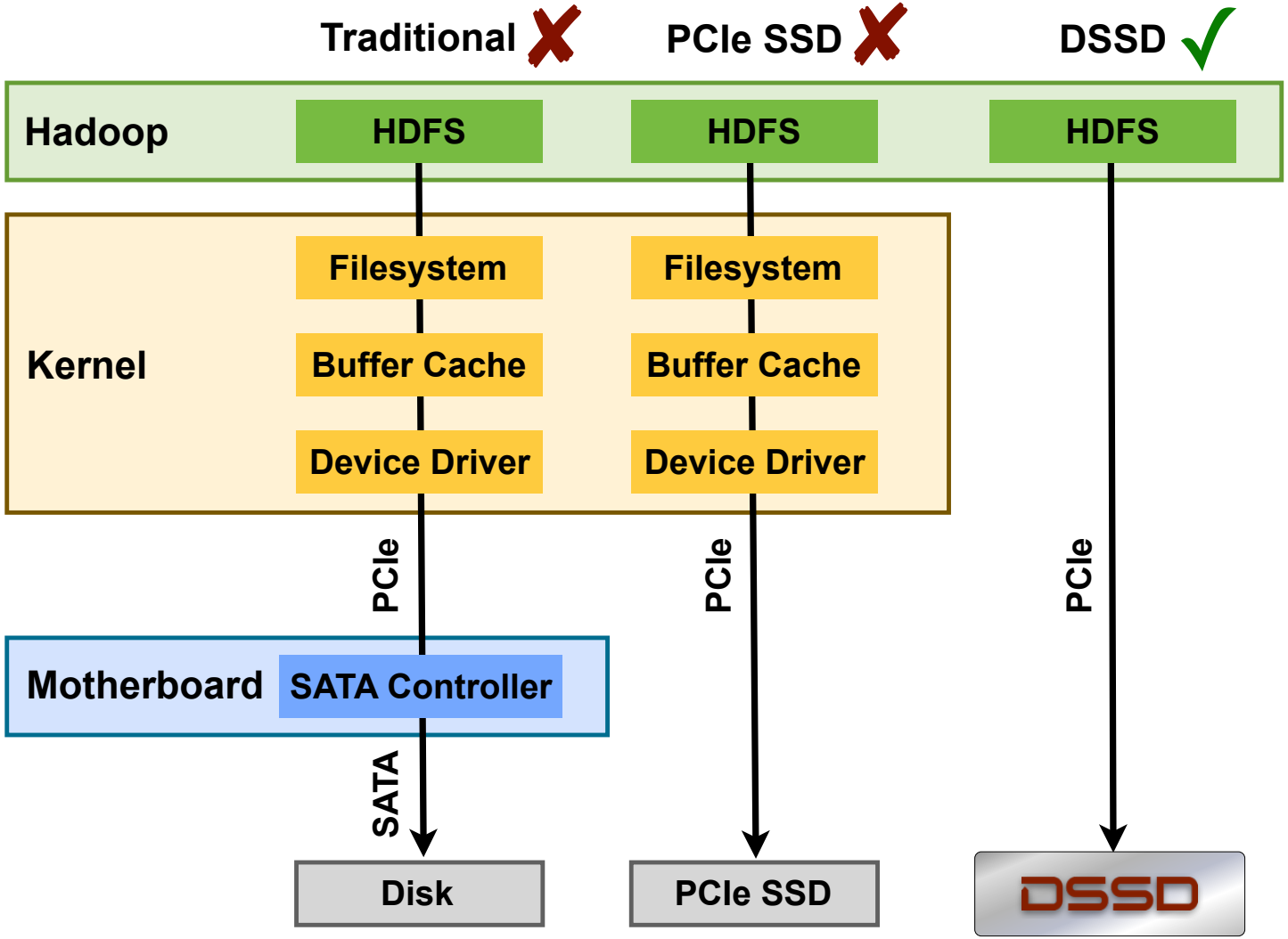
Petabyte/sec IO

Fastest Storage System

on the planet

DSSD Confidential

Direct Support for Hadoop/HDFS



Application Level Benefits

- 10X Improvements in HDFS Performance
- 3X Improvement in Oracle Benchmarks
- Consistent 100 usec Latency to Application
- Shared Data Avoids Need for Replication
- Enterprise-grade Data Integrity and Security

Use Cases

Fraud Detection

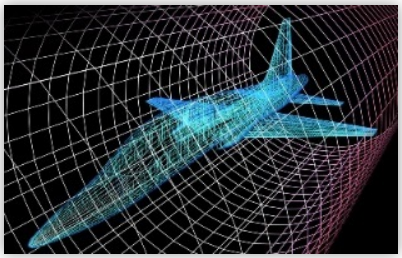
Risk Analytics

Predictive Modeling

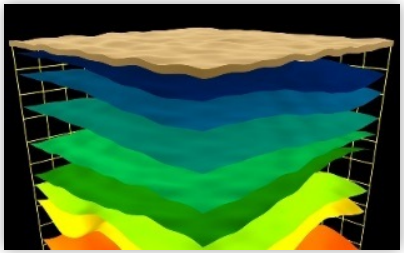
Real Time Analytics on Streaming Data

Financial: Transaction Modeling

Government: Design



Oil & Gas: Extraction Simulation Grids



Life Sciences & Research

Financial Modeling

Ingest Market Data

Analyze in near real-time

Correlate with Historical Data

Order of Magnitude Speed-up over existing solutions

Customer Quotes

“Cloudera has tested the DSSD D5 appliance in our lab, and we've seen an order of magnitude increase in performance. It’s the fastest HBase cluster we’ve ever tested.”

Mike Olson, Chief Strategy Officer at Cloudera

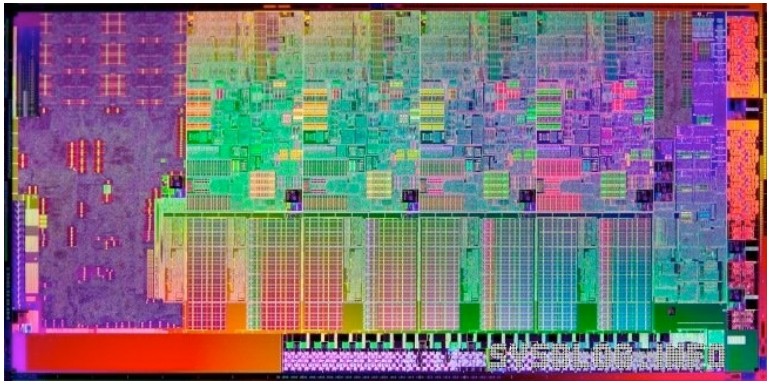
“DSSD D5 has fundamentally changed our business by eliminating unnecessary software, hardware and pre-processed batch jobs. With DSSD, we’re able to support applications and analytics at never-before-seen speeds.”

Brian Dougherty, Chief Technical Architect, CMA

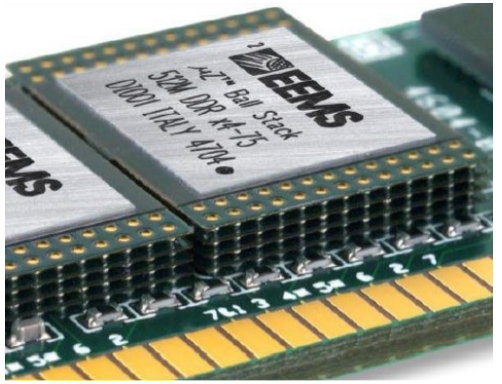
"No other supplier has a storage array equivalent in performance to the D5; it stands alone."

Chris Mellor, The Register

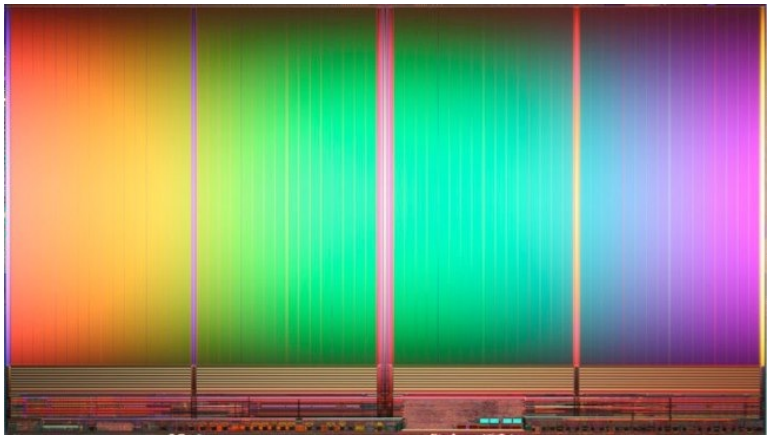
Summary: Technology Drivers



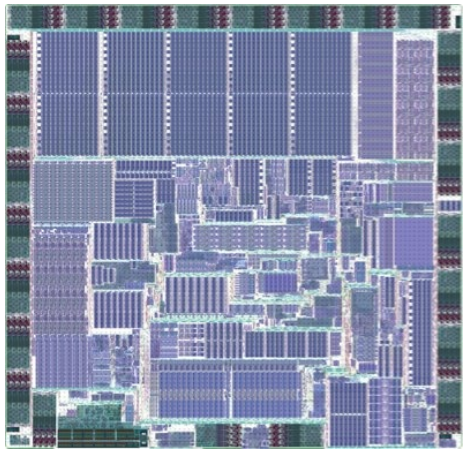
Faster CPUs



Denser Memory



Bigger Flash Memory

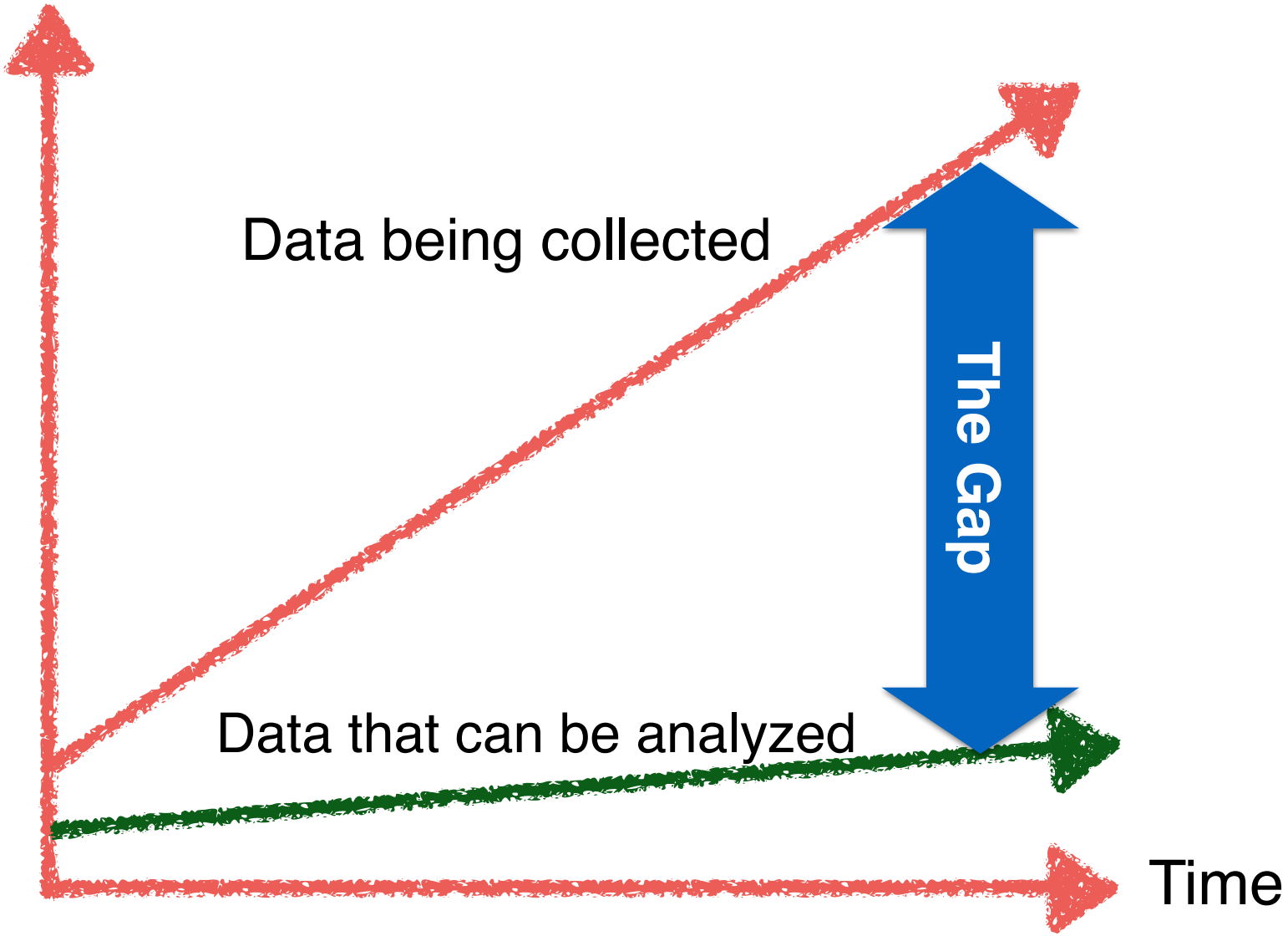


Faster Switch Chips

Data Centric System Architecture

Combination of NVMe Flash, 3D XPoint, High-performance Interconnect and new Software APIs solve the Data Deluge

Narrow the Data Analysis Gap



Conclusions

1. Rapid Growth in Real-time and Historical Data
2. Many Opportunities to Monetize such Data
3. Reducing Latency to Applications is Key
4. Rack-scale Storage provides 10X Speedup
5. Expect Further Performance Improvements