# CS6453

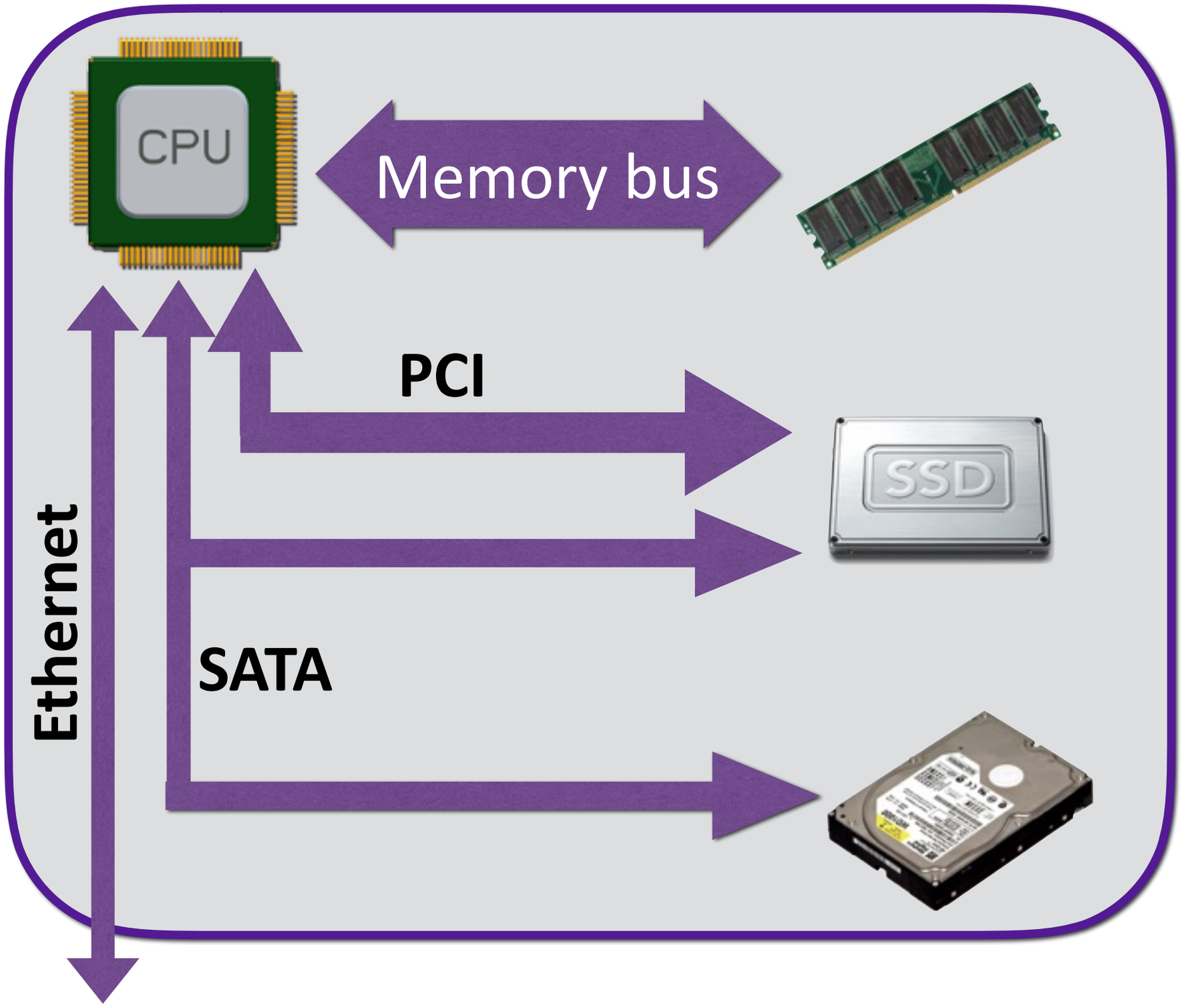
## Data-Intensive Systems:

## Technology trends, Emerging challenges & opportunities

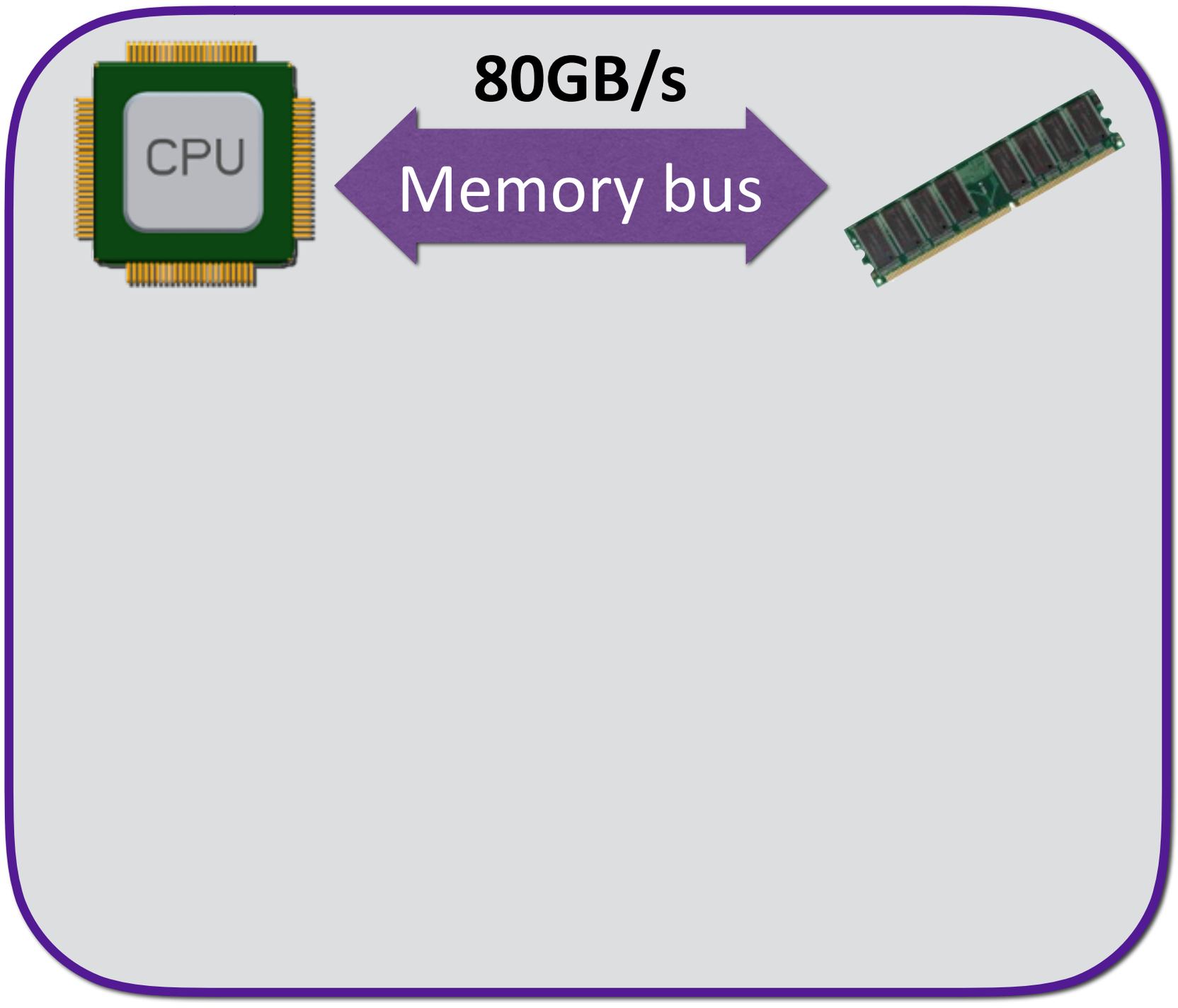
Rachit Agarwal

Slides based on:
many many discussions with Ion Stoica, his class, and many industry folks

# Servers — Typical node

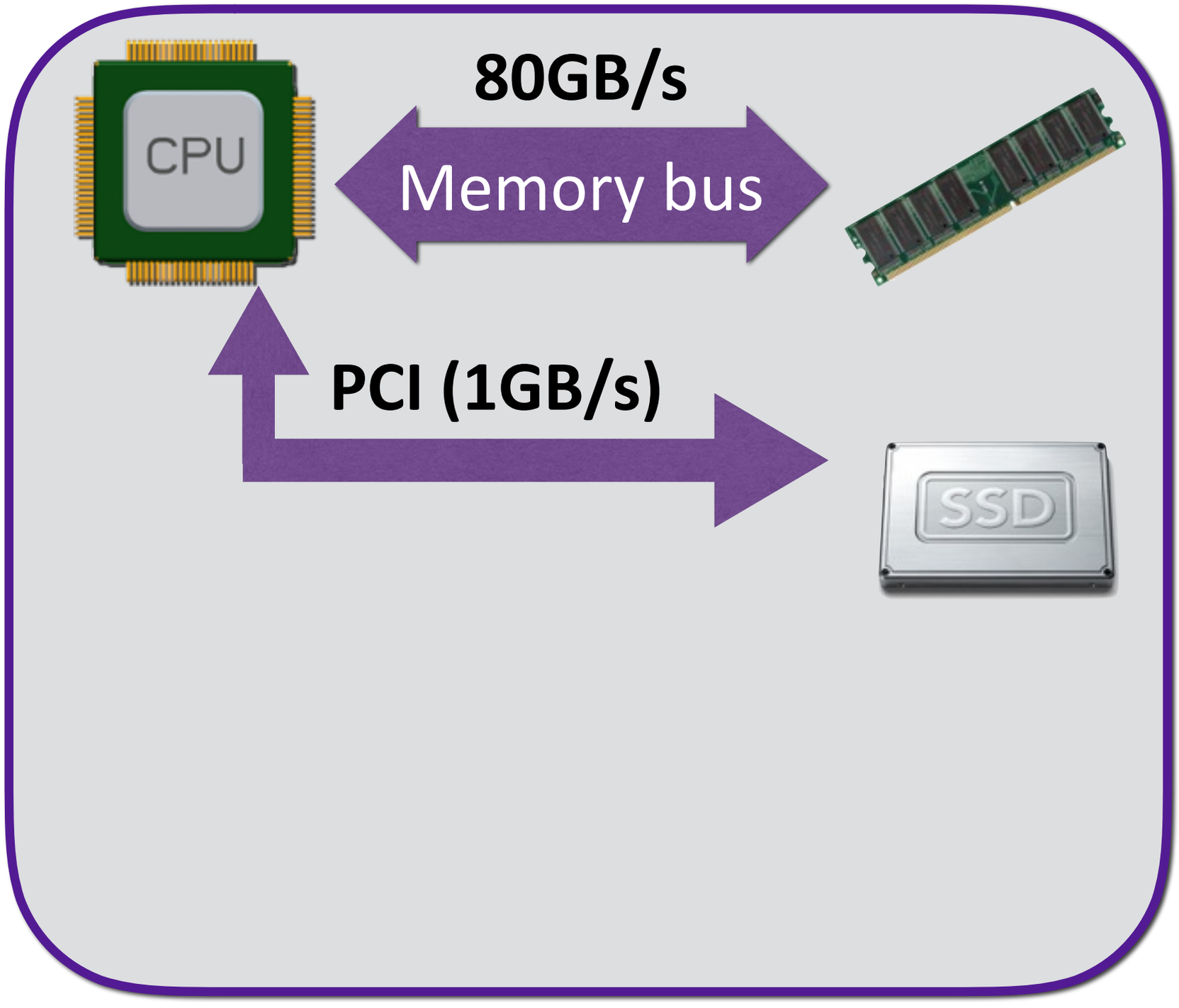


# Servers — Typical node



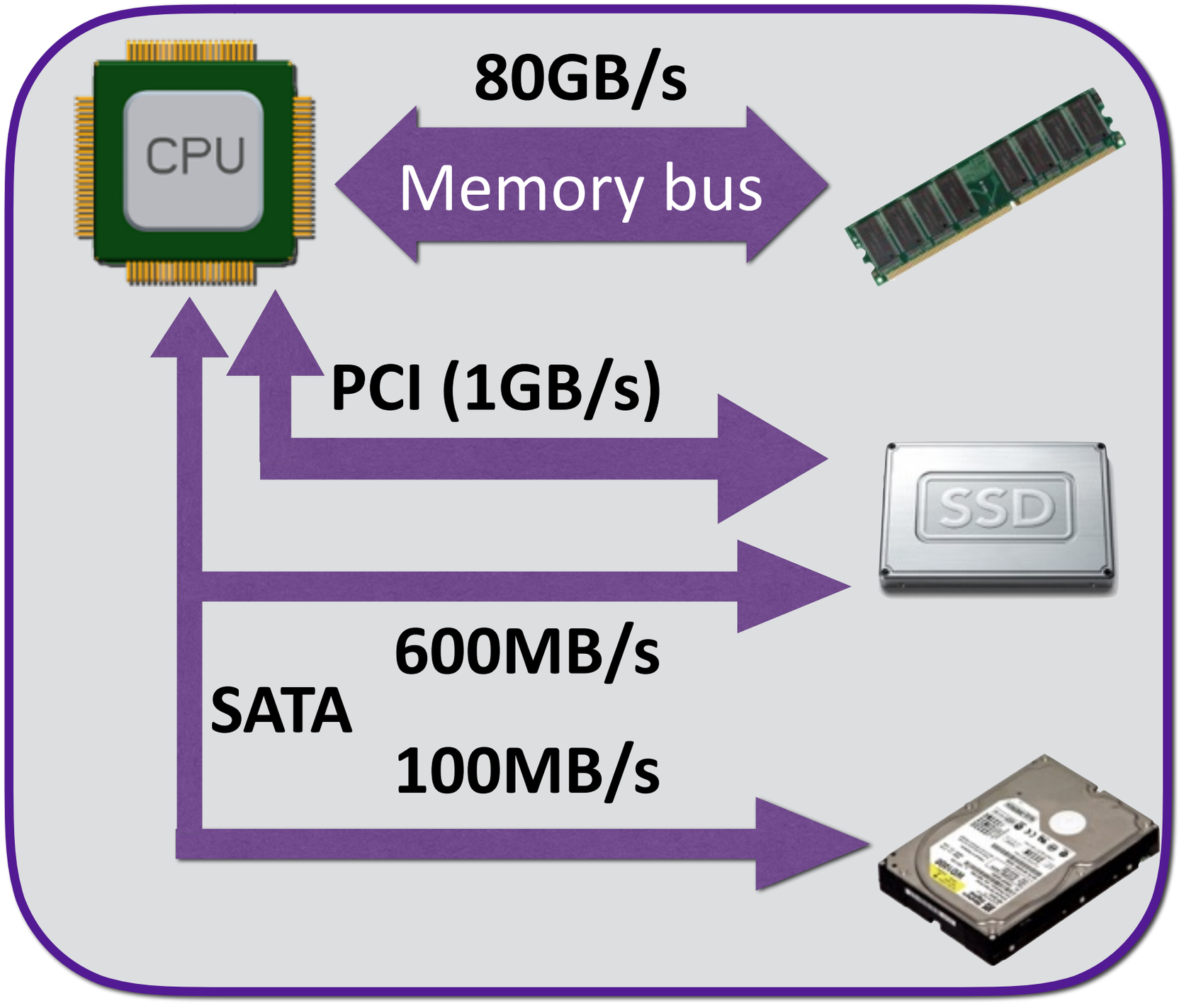

| | Capacity | Time to read |
|---|---|---|
| | 100s GB | 10s sec |

# Servers — Typical node




| | Capacity | Time to read |
|---|---|---|
| | 100s GB | 10s sec |
| | 100s GB | 10s min |

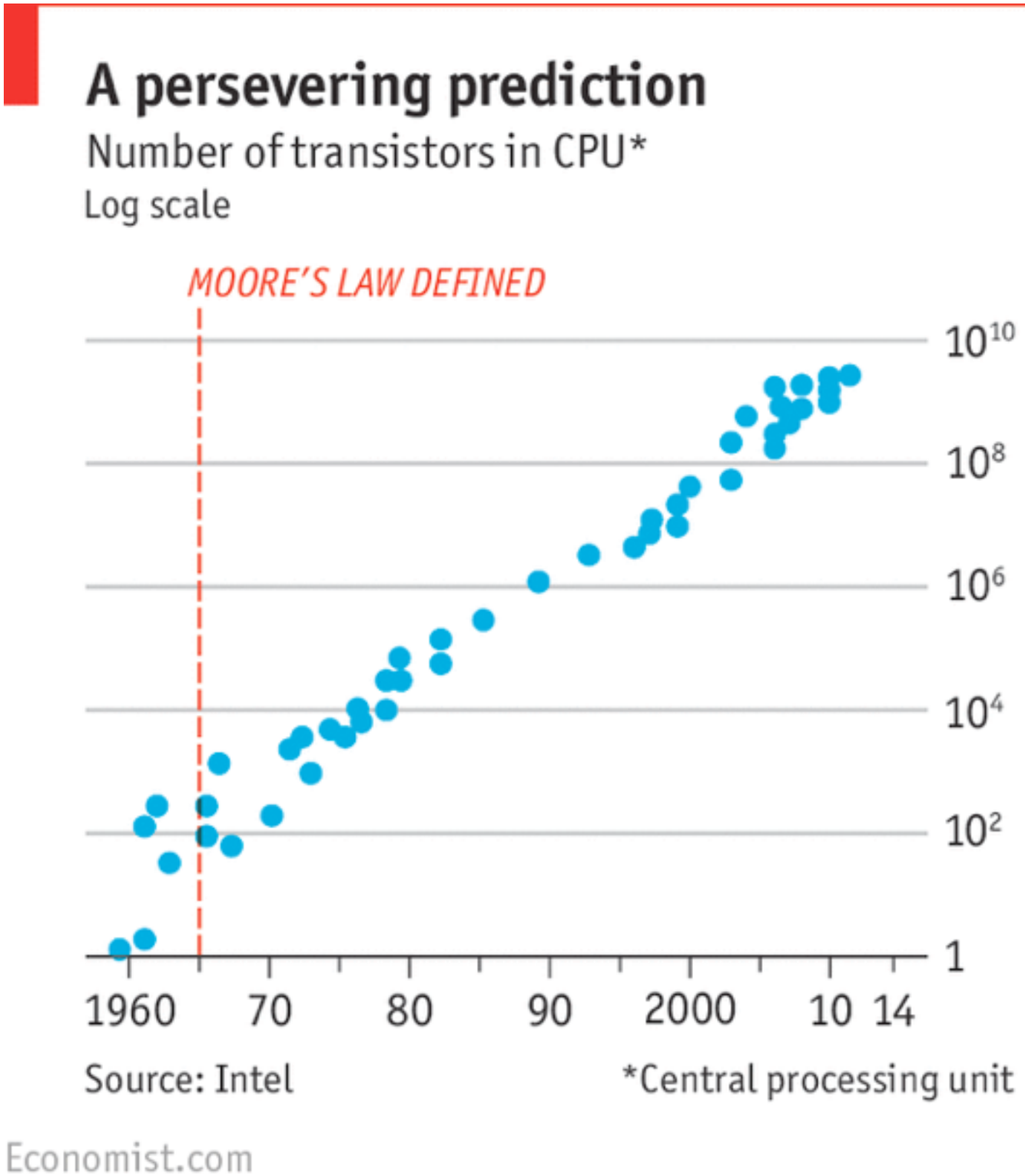
# Servers — Typical node



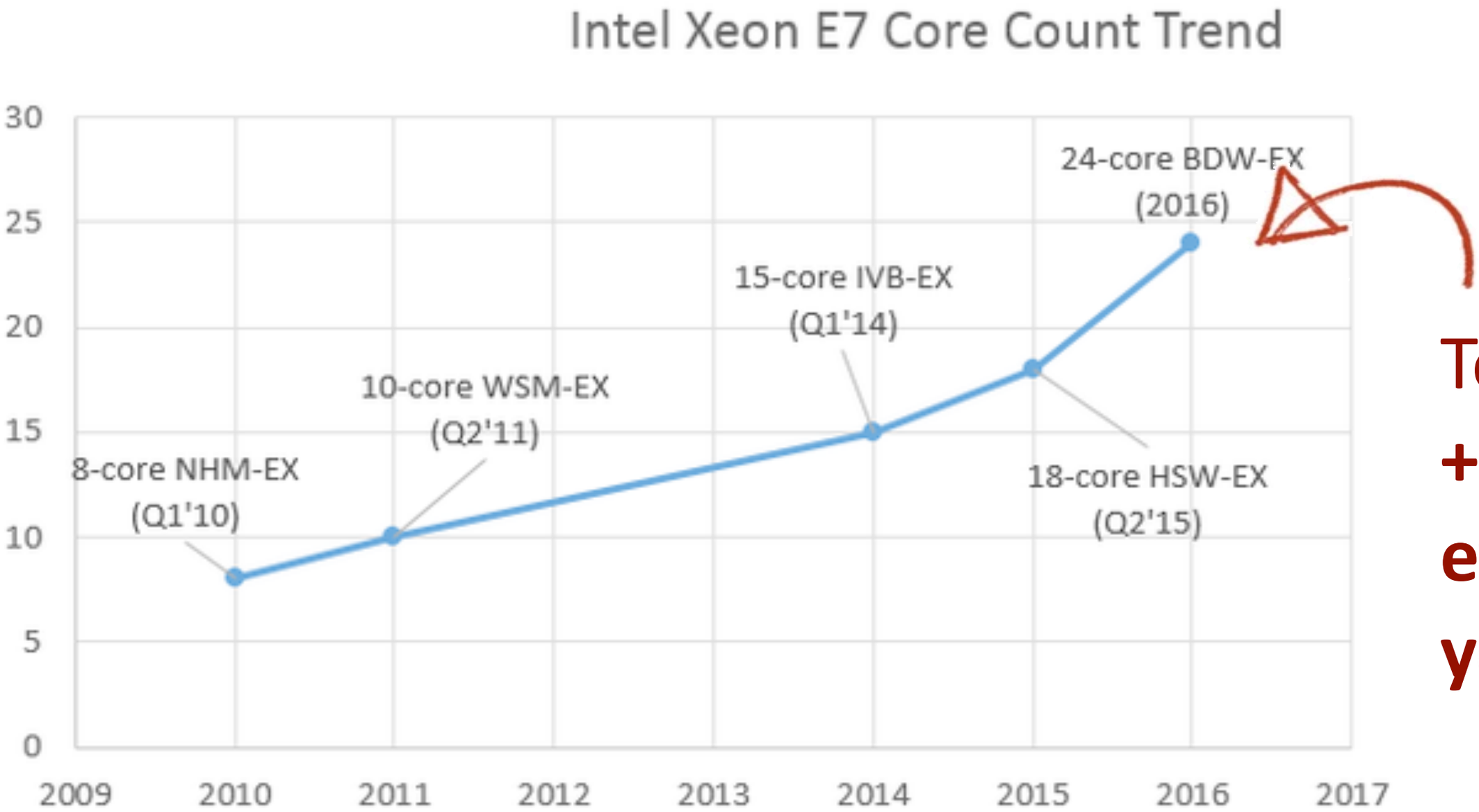

| | Capacity | Time to read |
|---|---|---|
| Memory bus | 100s GB | 10s sec |
| PCI (1GB/s) | 100s GB | 10s min |
| SSD | | 10s min |
| SATA 600MB/s | 1s TB | hours |
| SATA 100MB/s | | |

# Trends — Moore's law slowing down?

- **Stated 50 years ago by Gordon Moore**
- **Number of transistors on microchip double ~2 years**
  - Why interesting for systems people?
- Bryan Krzanich — Today, closer to 2.5 years

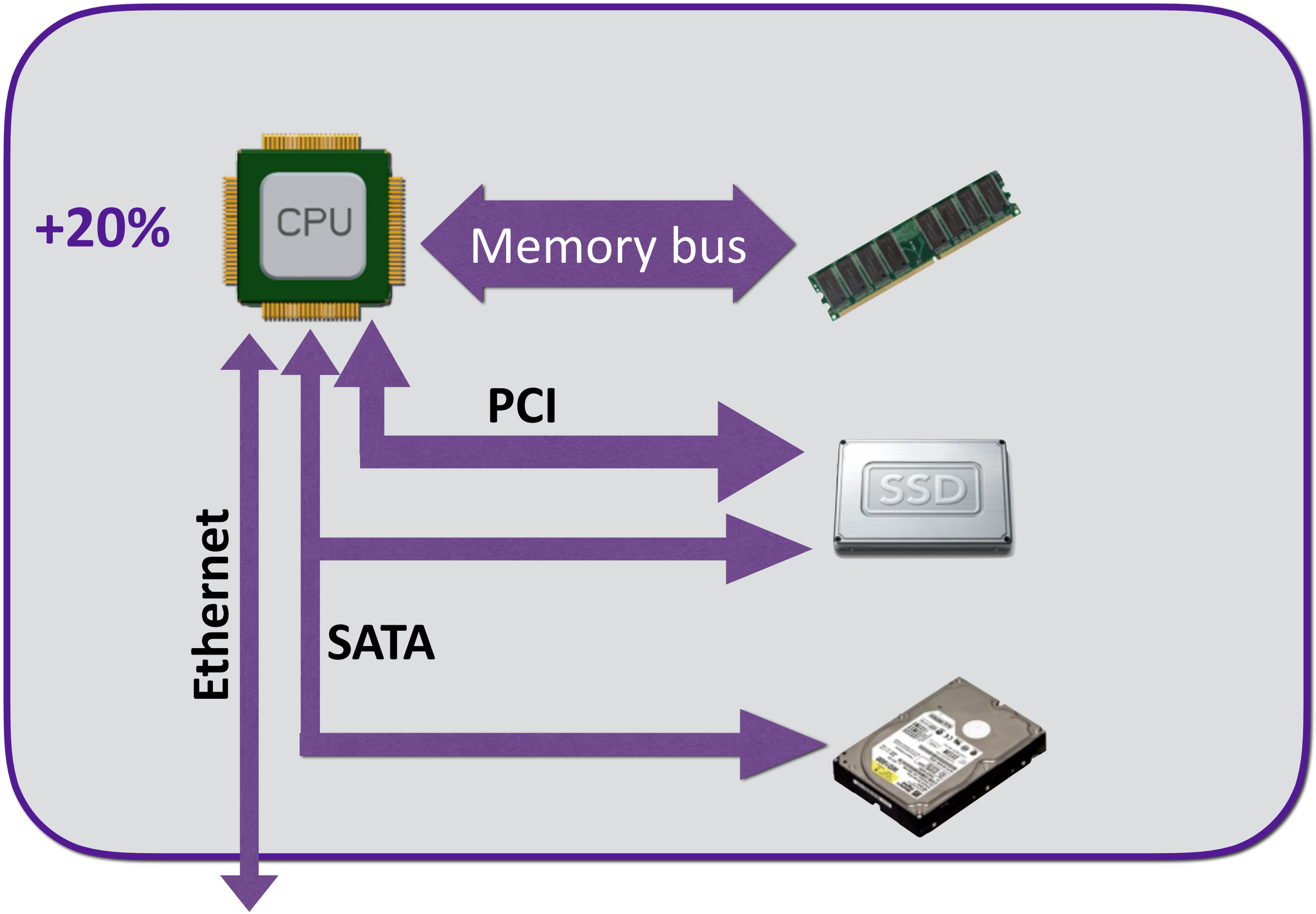


# Trends — CPU (#cores)

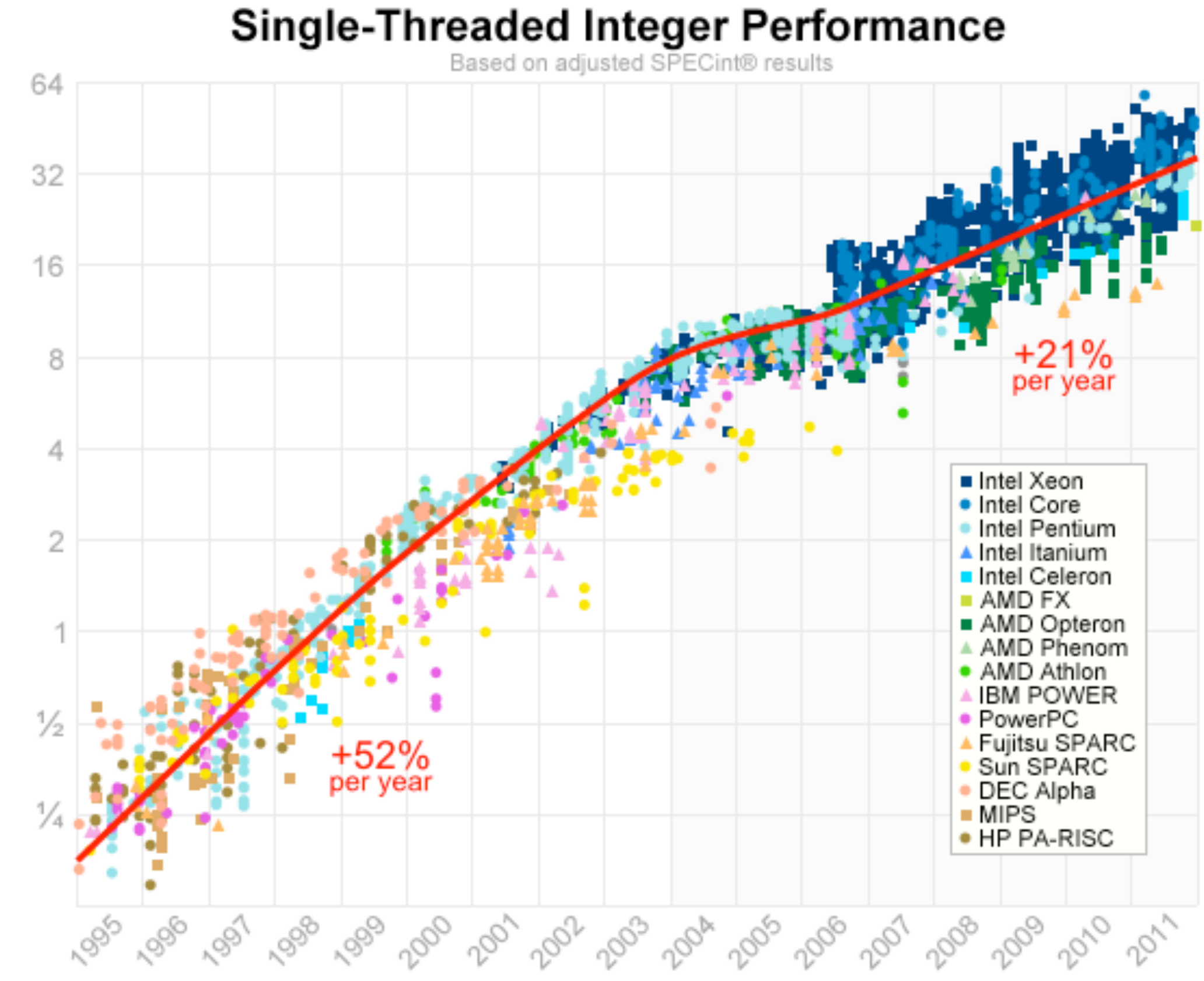



Today,
**+20%**
**every**
**year**

# Servers — Trends



# Trends — CPU (performance per core)

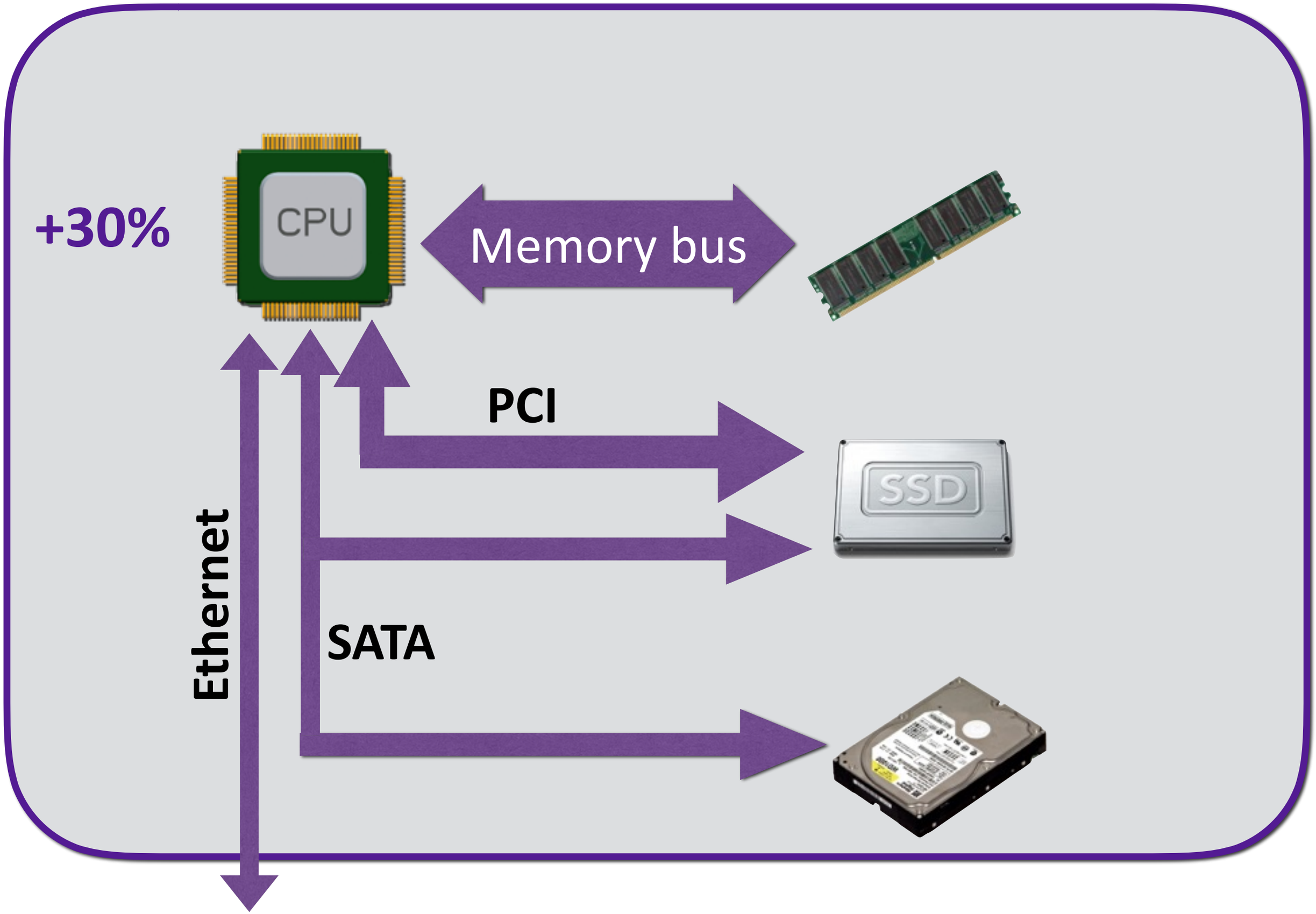



Today, **+10% every year**

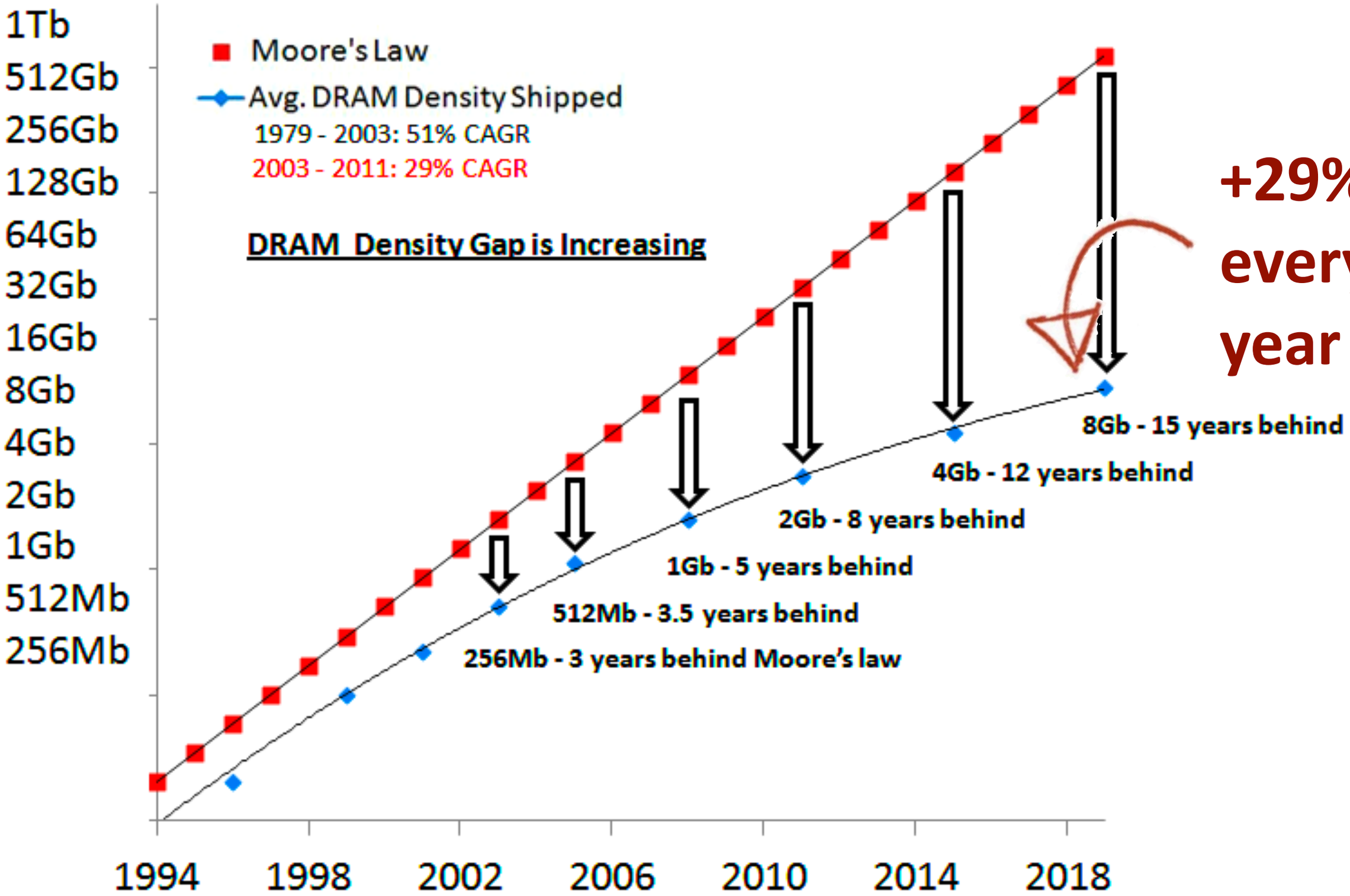
# Trends — CPU scaling

- Number of cores: +20%
- Performance per core: +10%
- Overall: +30-32%

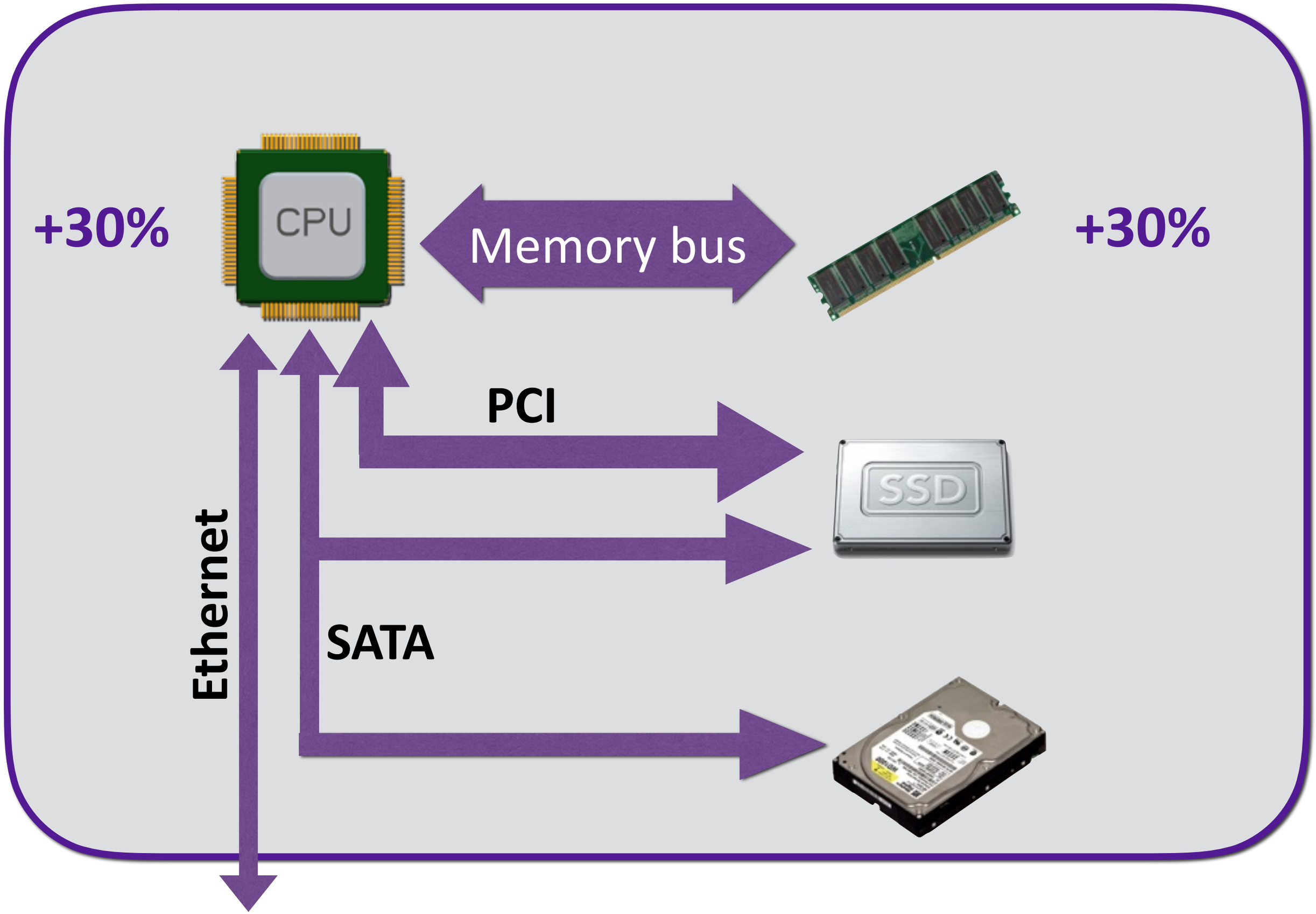
# Servers — Trends



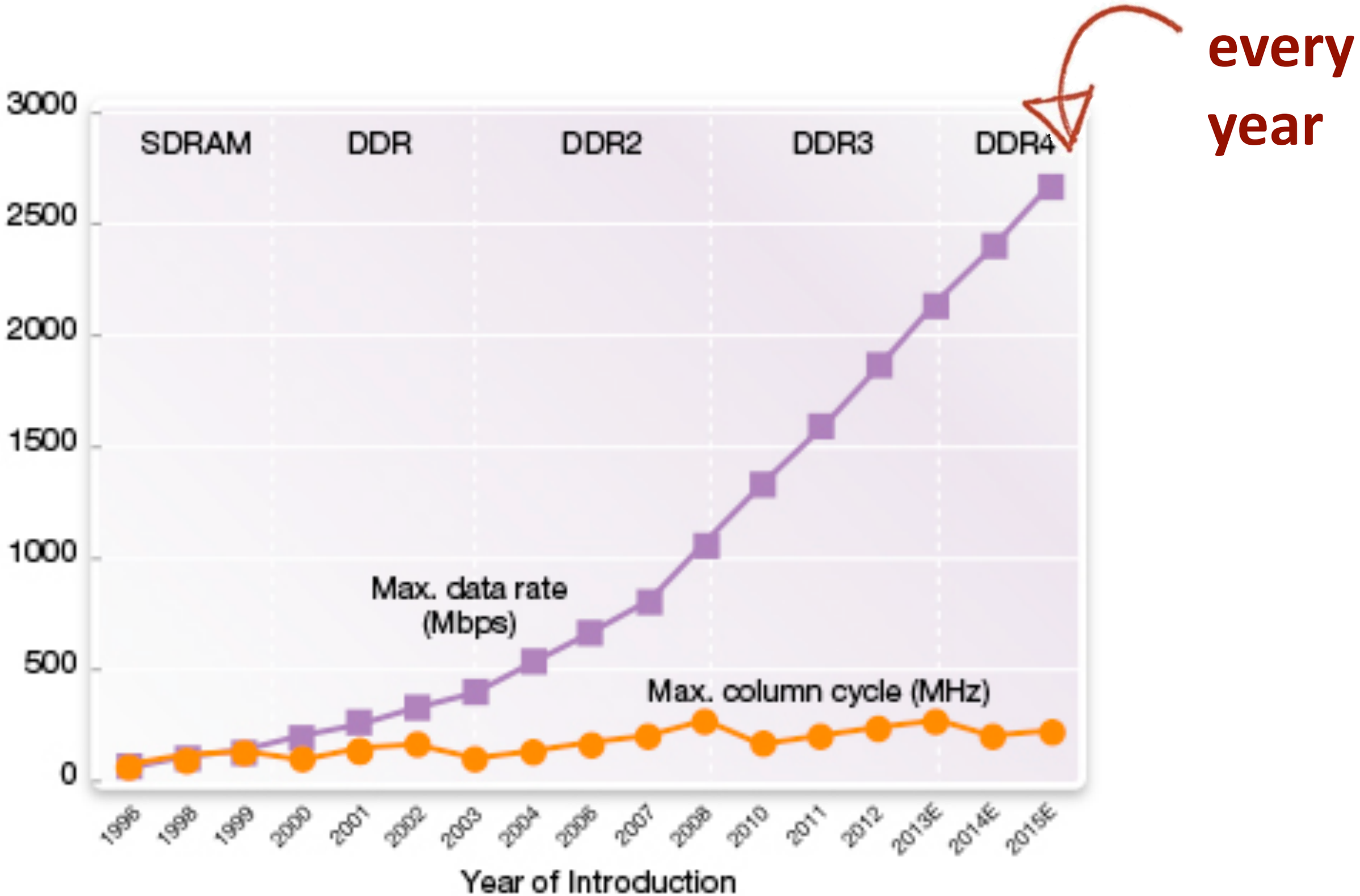
# Trends — Memory



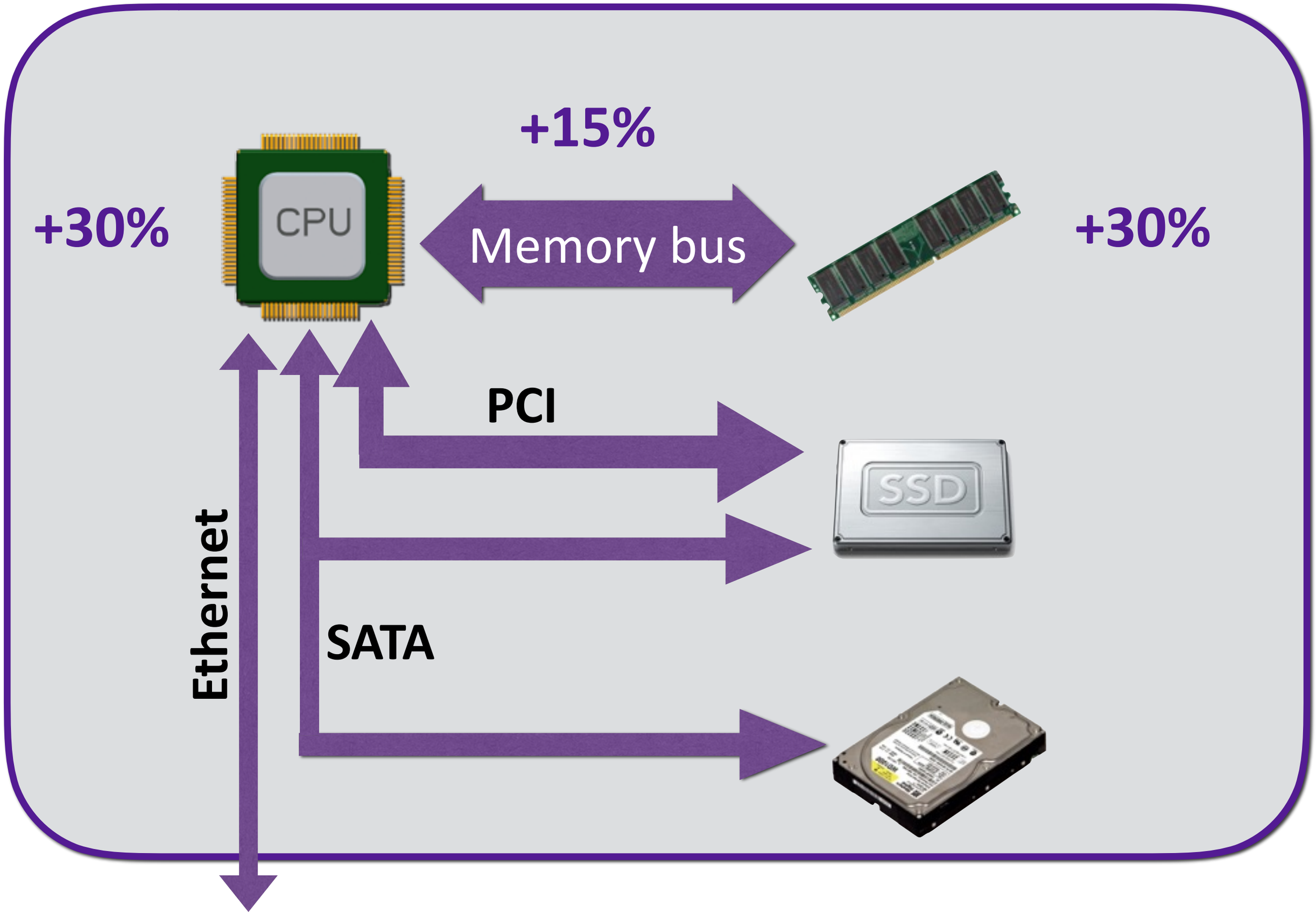
# Servers — Trends



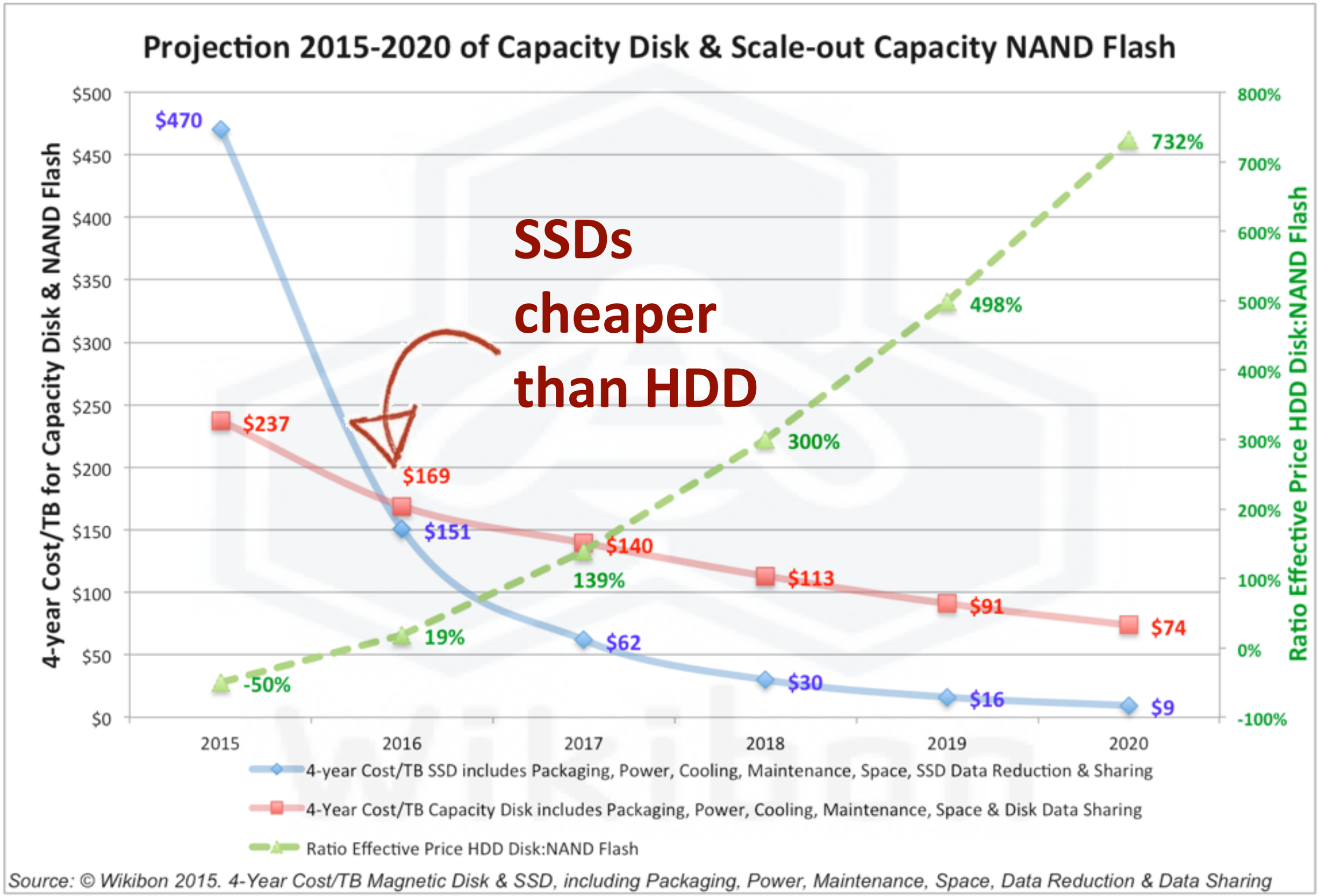
# Trends — Memory Bus



# Servers — Trends



# Trends — SSD

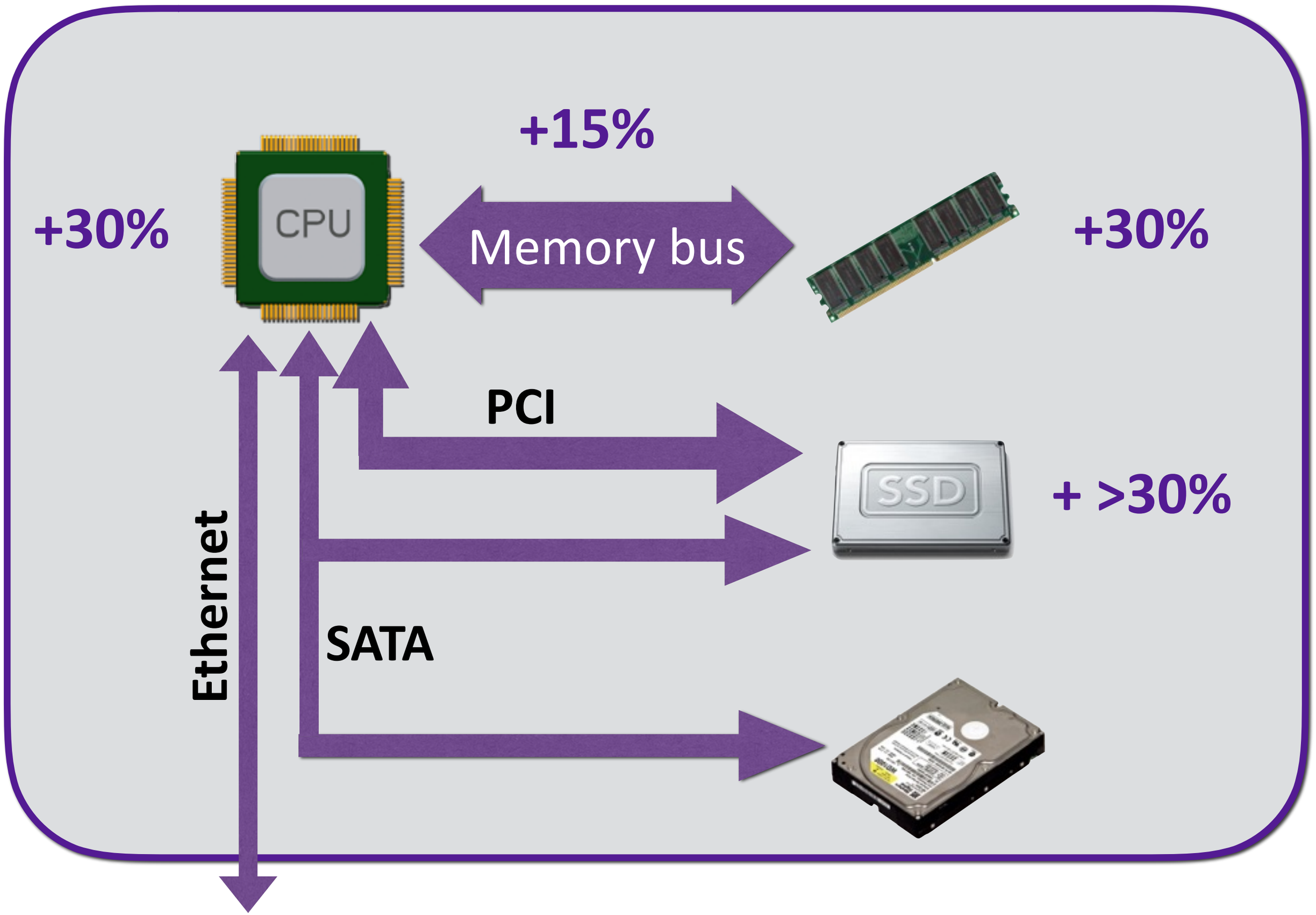



*Source: © Wikibon 2015. 4-Year Cost/TB Magnetic Disk & SSD, including Packaging, Power, Maintenance, Space, Data Reduction & Data Sharing*

# Trends — SSD capacity scaling

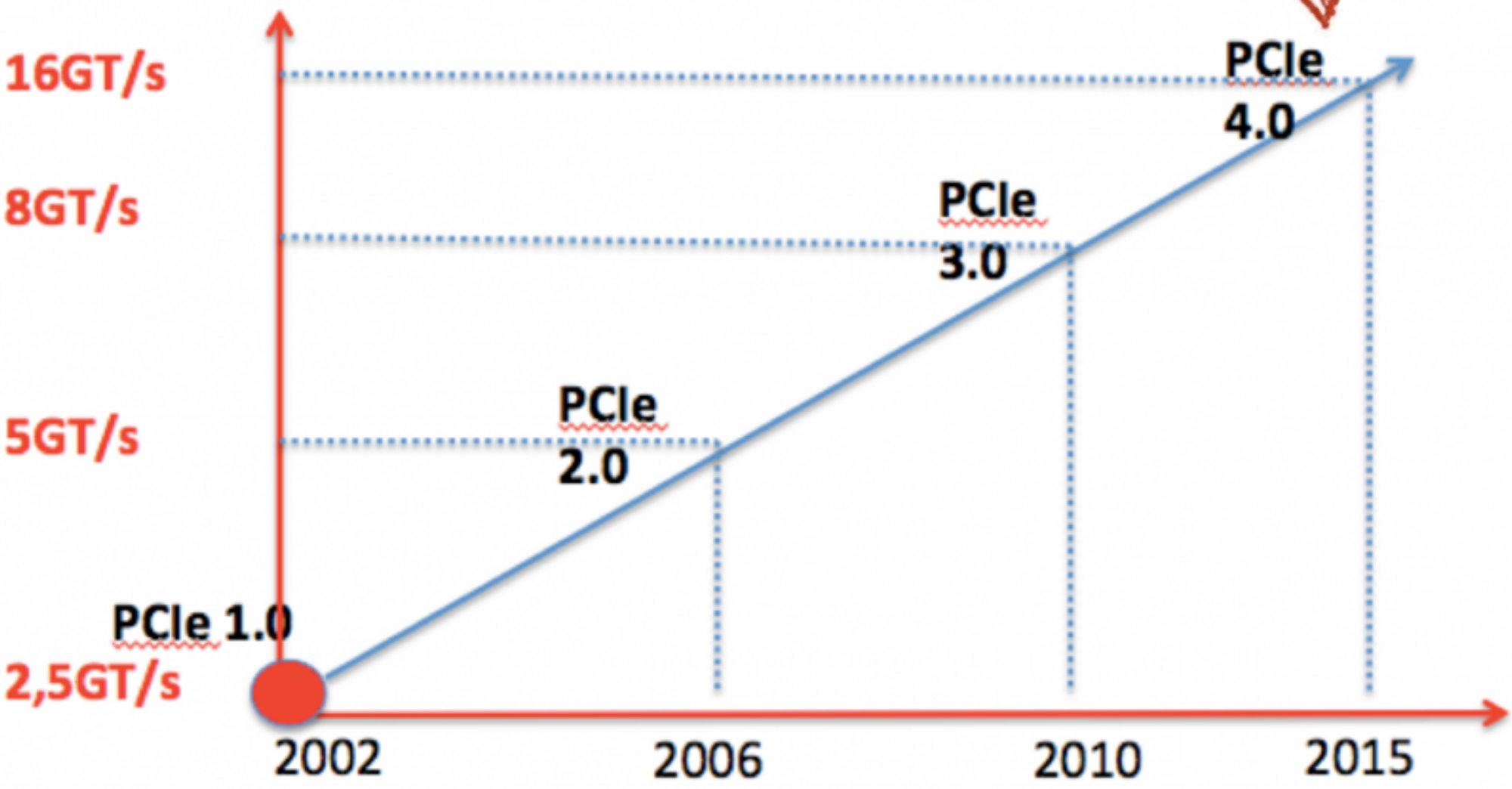
- Following Moore’s law (late start)
- 3D technologies
    - May even outpace Moore’s law

# Servers — Trends

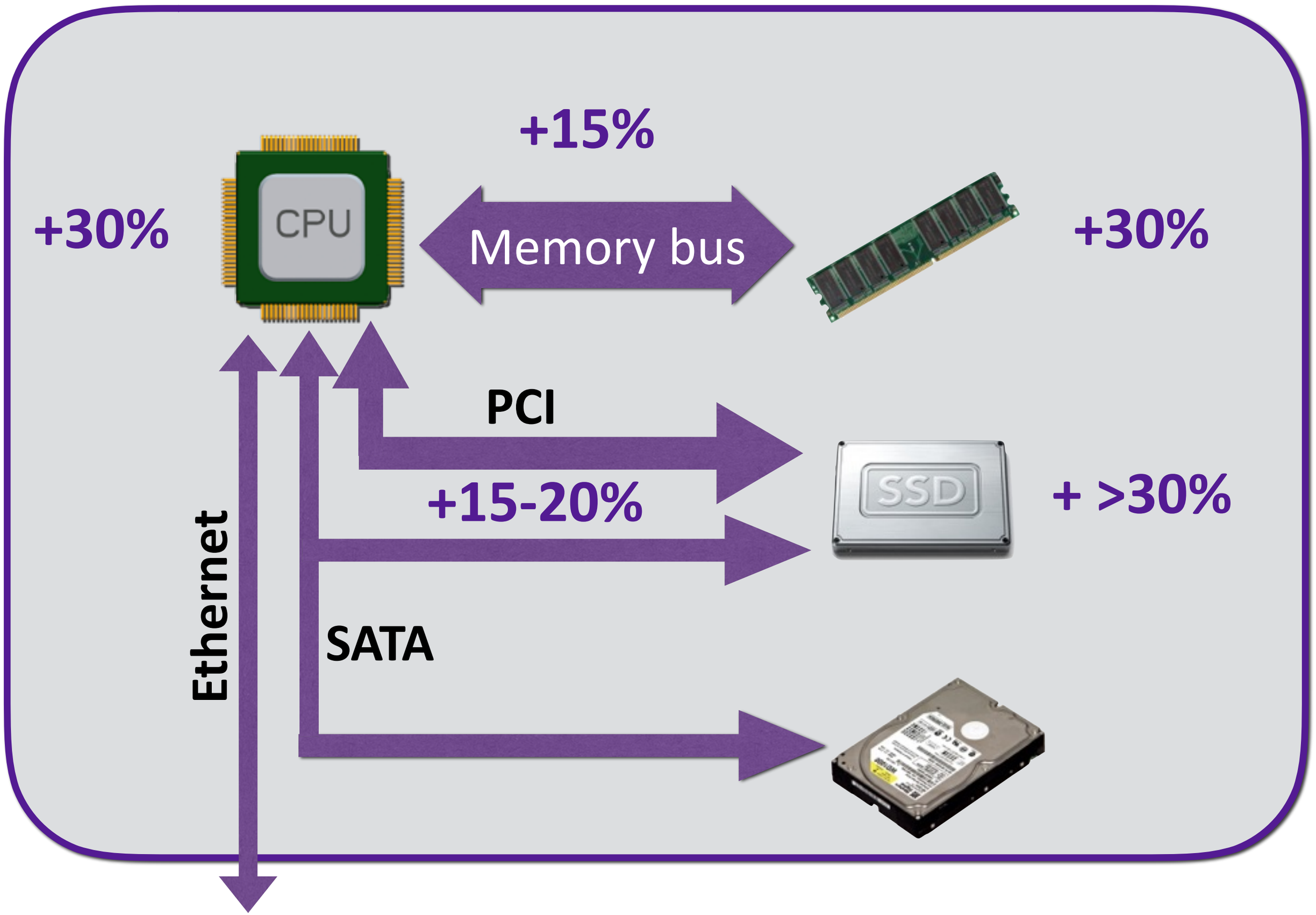


# Trends — PCI bandwidth (and ~SATA)

**+15-20% every year**

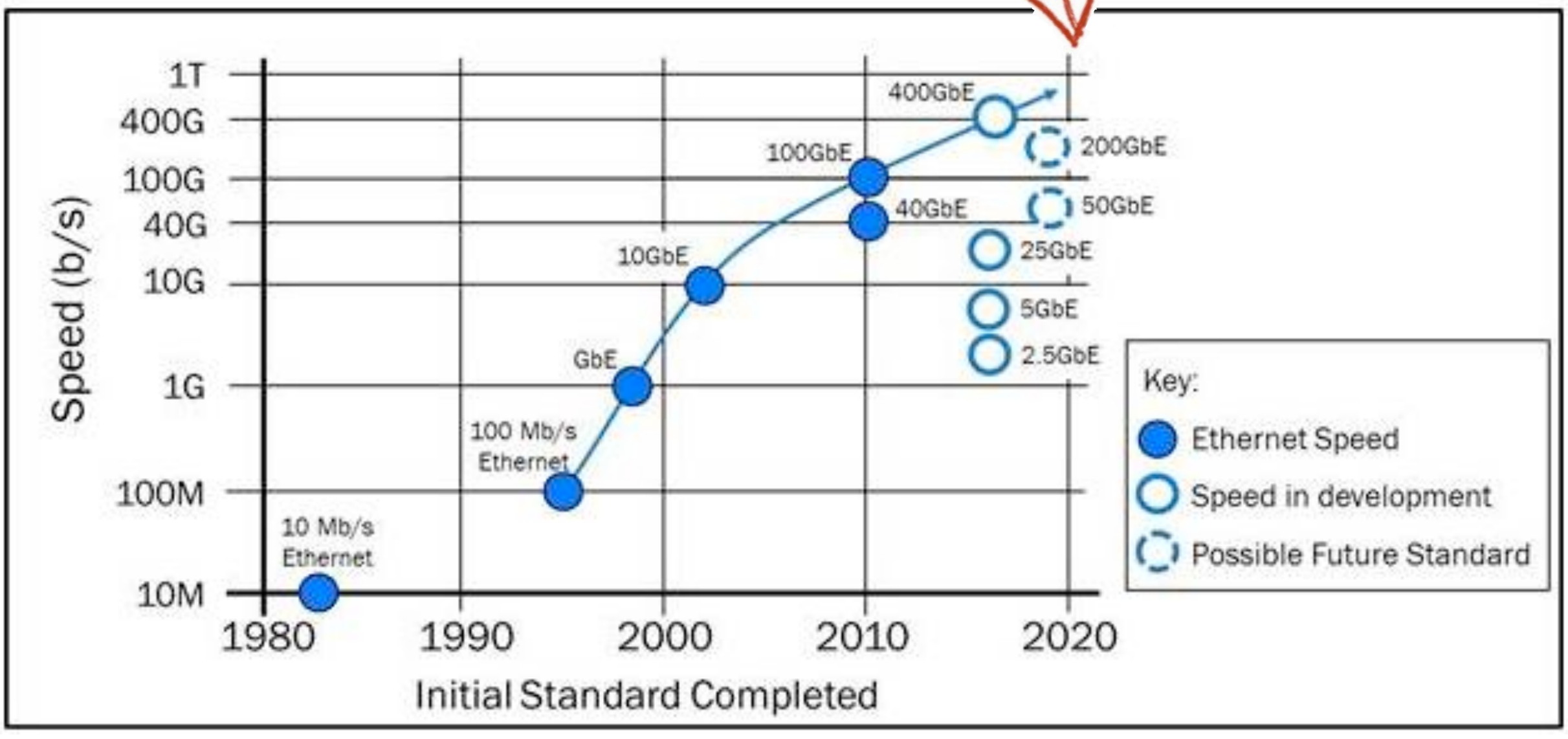


# Servers — Trends

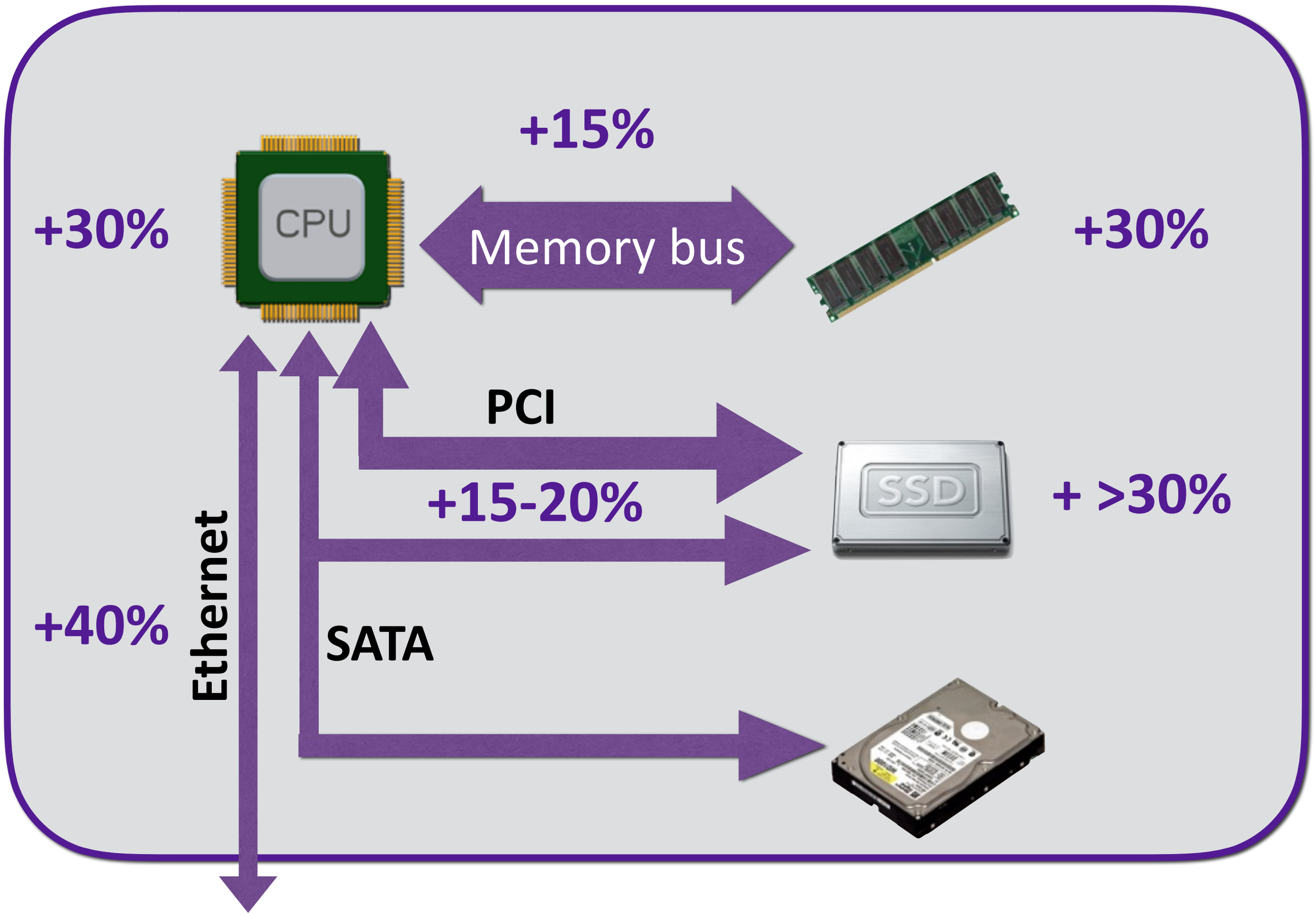


# Trends — Ethernet bandwidth

**+33-40% every year**



# Servers — Trends



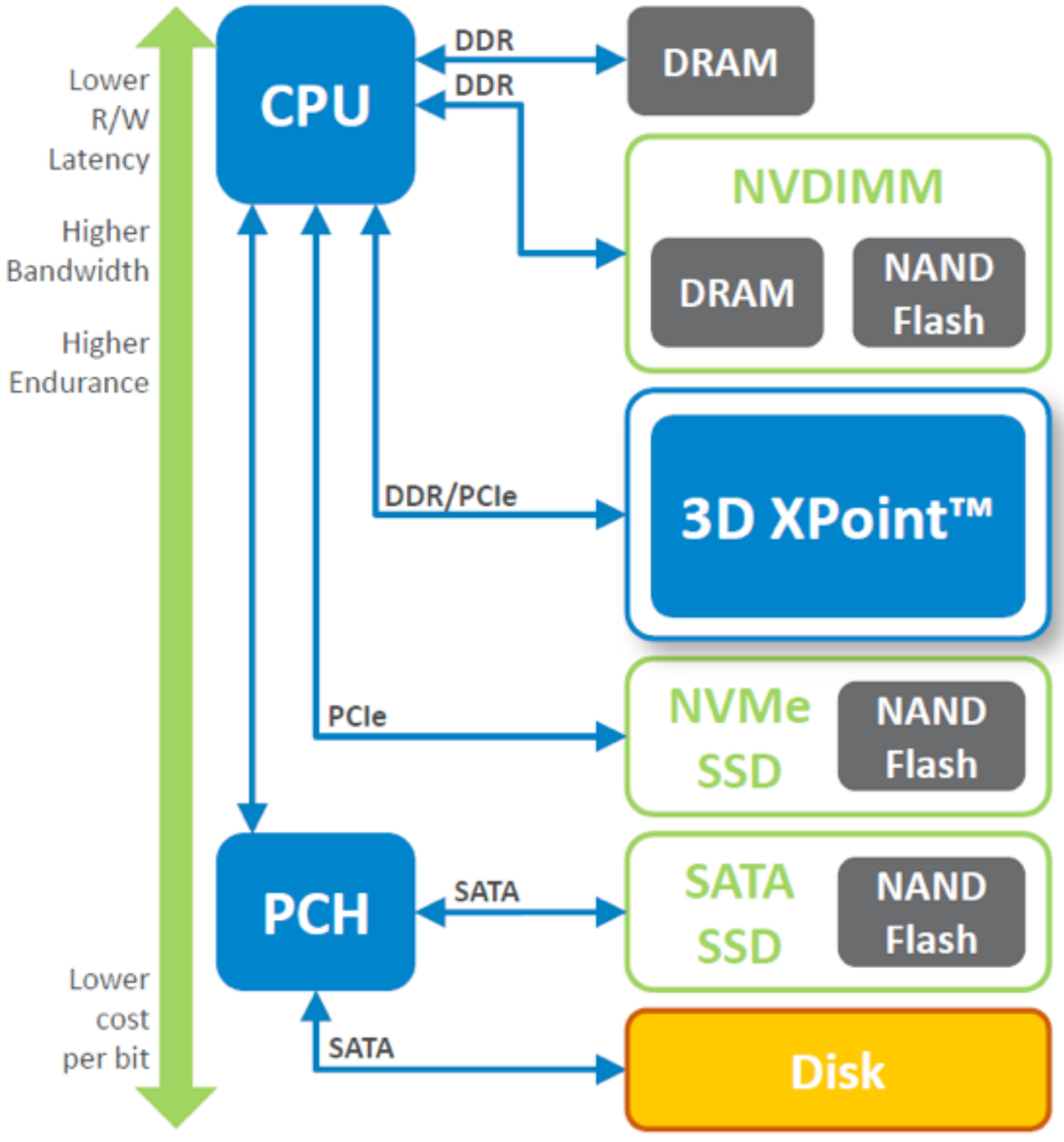
# Trends — Implications?

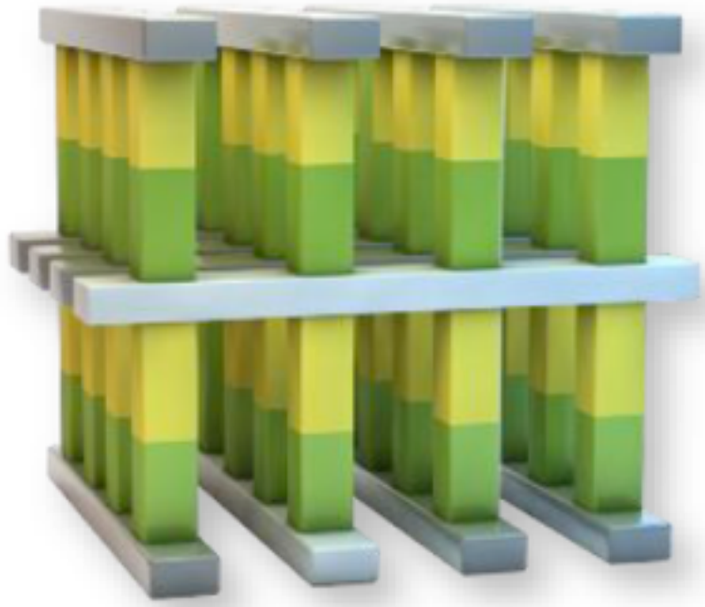
- **Intra-server** Bandwidth an increasing bottleneck
- How could we overcome this?
  - Reduce the size of the data?
    - What does that mean for applications?
  - Prefer remote over local?
    - Challenges?
    - Non-intuitive; we always prefer locality

# Trends — Emergence of new technologies

- **Non-volatile memory**
- 8-10x density of DRAM (close to SSD)
- 2-4x higher latency
- But who cares? Bandwidth is the bottleneck…

# Trends — Emergence of new technologies



- 3D XPoint technology provides the benefit in the middle
- It is considerably faster than NAND Flash
- Performance can be realized on PCIe or DDR buses
- Lower cost per bit than DRAM while being considerably more dense

# Trends — Emergence of new technologies

https://www.youtube.com/watch?v=IWsjbqbkqh8

# Trends — & Implications

- **HDD** is new tape
- **SSD/NVRAM** is the new persistent storage
    - But, increasing gap between capacity and b/w concerning …
- **Deeper storage hierarchy** (L1, L2, L3, DRAM, NVRAM, SSD, HDD)
    - Do CPU caches even matter?
    - How do design software stack to work with deeper hierarchy?
- **CPU-storage "disaggregation"** is going to be a norm
    - Easier to overcome bandwidth bottlenecks
    - Google and Microsoft have already realized
    - What happens to locality?
    - Re-think software design?

# Paper 1 — Memory-centric design

- **SSD/NVRAM** is the new persistent storage (+archival)
    - Not just the persistent storage, THE storage
    - +(private memory), deep storage hierarchy
- **CPU-storage "disaggregation"**
    - NVRAM shared across CPUs
- **Challenges?**
    - How to manage/share resources?
        - NVM: accelerators and controllers
        - Addressing? Flat virtual address space?
        - NVM sharing in multi-tenant scenarios?
        - NVM+CPU+Network: software-controlled?
        - Storage vs compute heavy workloads?

# Paper 1 — Memory-centric design

- **New failure modes? [very interesting direction!!]**
  - CPU-storage can fail independently
  - Very different from today's "servers"
  - Good? Bad?
  - Transparent failure mitigation…?
- **How about the OS?**
  - Where should the OS sit?
  - What functionalities should be implemented within the OS?
- **Application-level semantics**
  - ?

# Paper 2 — Nanostores (An alternative view)

- **DRAM** is dead
- **SSD/NVRAM** is the new persistent storage (+archival)
    - Not just the persistent storage, THE storage
    - No storage hierarchy
- **CPU-storage "convergence"** is going to be a norm
    - CPU-storage hyper-convergence
    - Berkeley IRAM project (late 90s)
- **Challenges?**
    - Network? (topology, intra-nanostore latency, throughput)
    - How does this bypass the trends discussed earlier?

# Trends — The missing piece?

- **Data volume** increasing significantly **faster than Moore's law**
  - 56x increase in Google indexed data in 7 years
  - 173% increase in enterprise data
  - Uber, Airbnb, Orbitz, Hotels, …
- **Data types**
  - Images, audio, videos, logs, logs, logs, genetics, astronomy, ….
  - YouTube: ~50TB of data **every day**

# Trends — Discussion

- **Other missing pieces?**
  - Software overheads
  - Application workloads
  - Specialization vs. generalization?