



When two promising fabless startups who share the same design principle meet by chance in the era of post Moore's law

Nov. 2022

Agenda

- FADU introduction and mega trend
- FADU NVMe SSD
- RISC-V CPU
- Next step

Company overview

- FADU is a fabless company focusing on enterprise/datacenter grade SSD solution
 - 180 engineers including more than 20 Ph.Ds and seasoned engineers from Samsung and SK Hynix
- Our product is the most advanced SSD in the market
 - We focus on introducing innovative architecture for next generation PCIe NVMe SSD for datacenter and storage
 - High performance, low power, and Superior QoS are our key differentiator. We have the best SSD which is the only SSD that can beat Samsung
- We are building track records with major customers in US
 - Leader datacenters such as Microsoft, Apple, Meta (Facebook) love our product.
 - We started mass production for Meta. We expect that a few more big names will be the next customers. We are confident we will be the leader.
- We finally start delivering real business result
 - After 7 years of tough journey, we start generating profit from 2022
 - We are preparing the next leap forward.

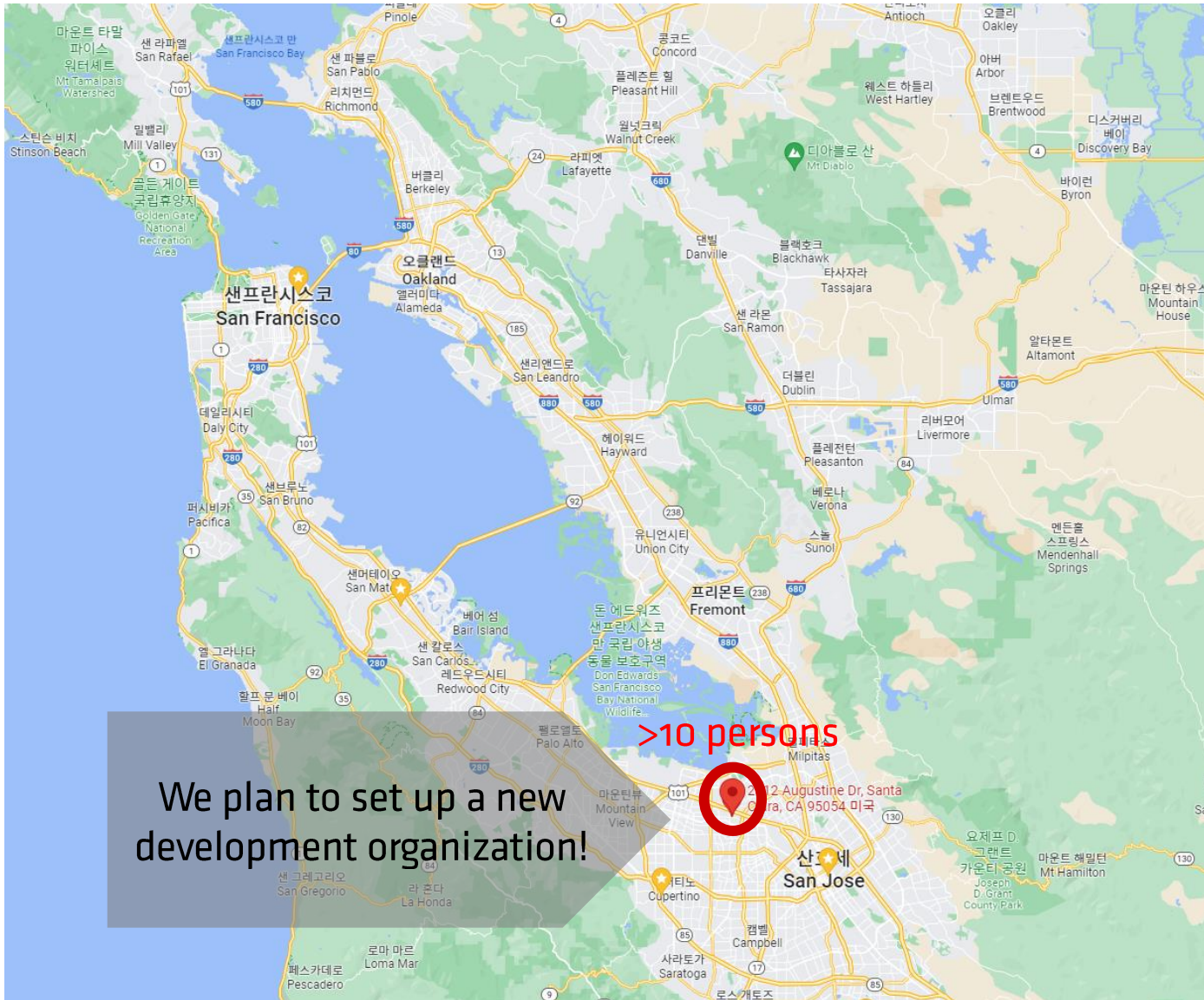


Company location – parent company

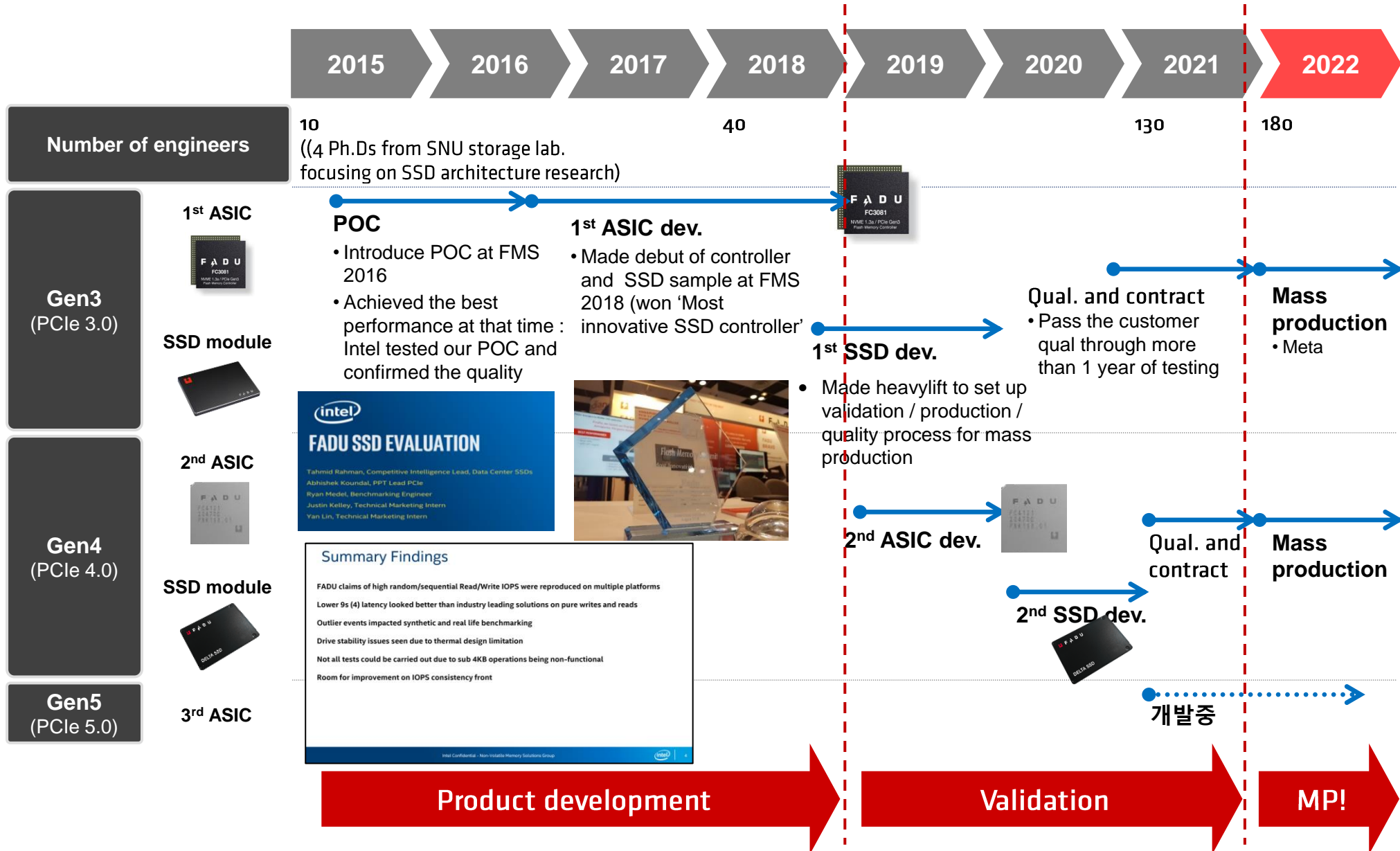
- We have always been located only in Gangnam, Lol.



Company location – subsidiary company

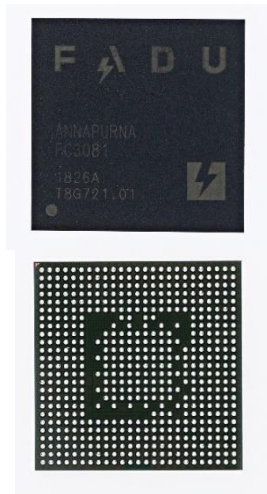


Founded in 2015, We has been gone through the entire business cycle to deliver the result



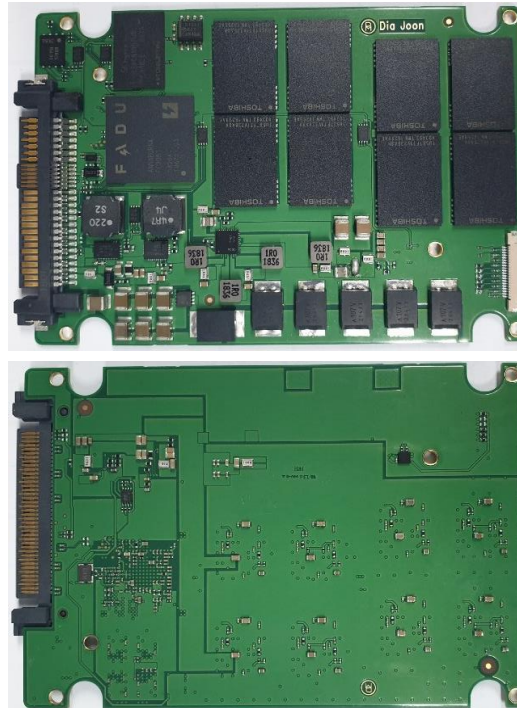
Gen3 : NVMe SSD controller (Annapurna) and product (Bravo)

Annapurna NVMe SSD controller



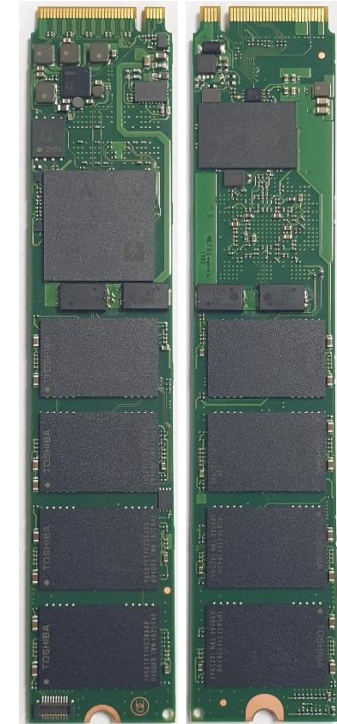
- PCIe 3.0 X 4 (2 X 2 dual port)
- 3.5 GB/s SR, 3.5GB/s SW
- 750 RR, 300 RW
- RISC-V processor
- All enterprise features supported
- 17 X 17mm packages

U.2 Dual Port 7mm NVMe Drive



- 1/2/4 TB
- Full enterprise ready
- Very low power / High performance

M.2 22x110 Drive



- 1/2/4 TB
- Full enterprise ready
- Very low power / High performance (remove thermal issue)

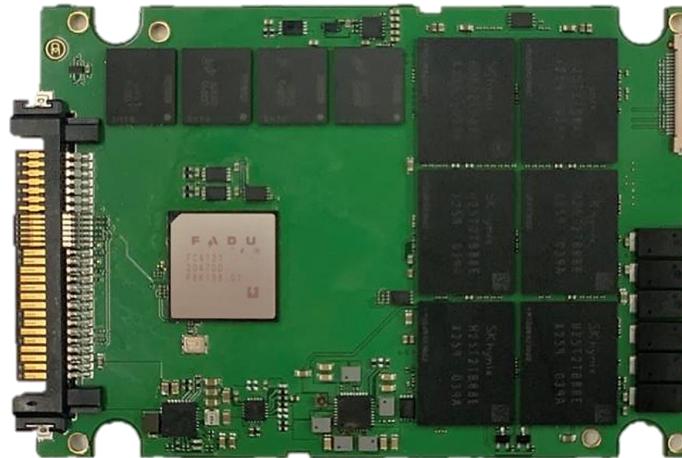
Gen4 : NVMe SSD controller (Everest) and product (Delta)

Everest NVMe SSD controller



- PCIe 4.0 X 4 (2 x 2 dual port)
- 6.7 GB/s SR, 5.0GB/s SW
- 1500 RR, 400 RW
- RISC-V processor
- All enterprise features supported
- 17 X 17mm packages

U.2 Dual Port 7mm NVMe Drive



- 2/4/8 TB
- Full enterprise ready
- Very low power / High performance

E1.s Drive

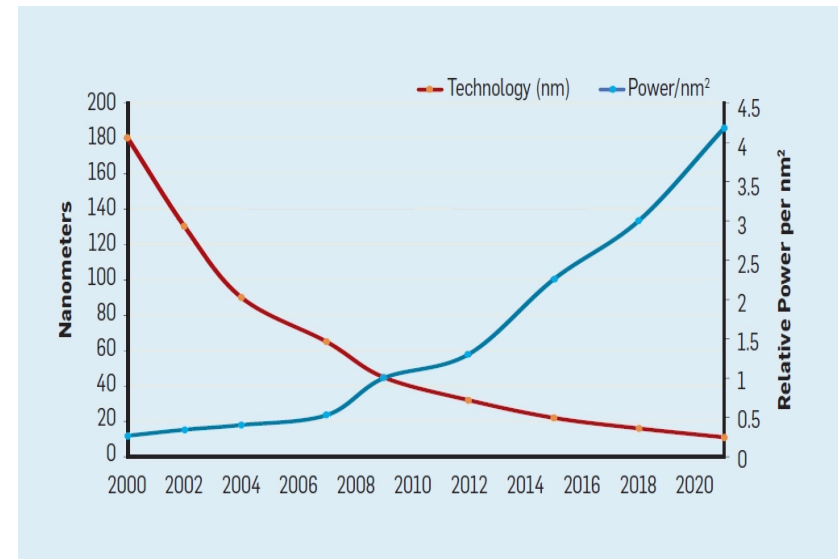
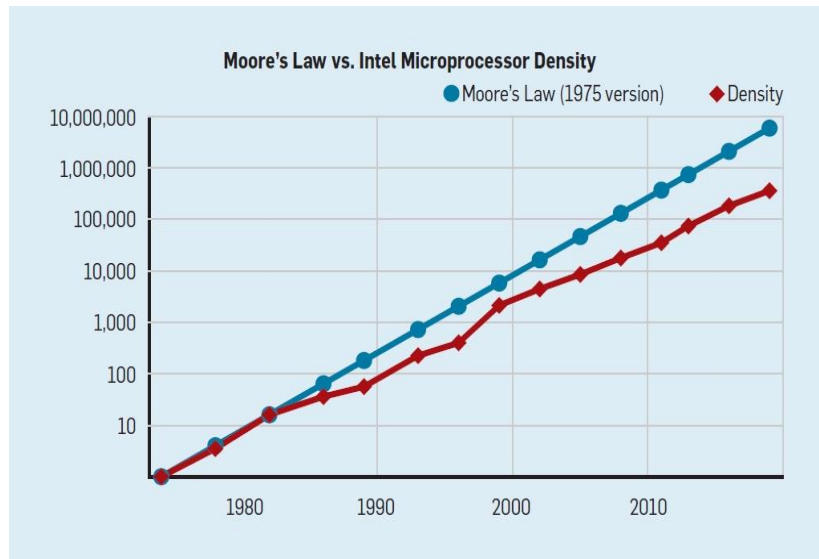


- 2/4/8 TB
- Full enterprise ready
- Very low power / High performance (remove thermal issue)

End of Moore's law

- The end of Dennard scaling and Moore's Law and the deceleration of performance gains for standard microprocessors are not problems that must be solved but facts that, recognized, offer breathtaking opportunities.
- Innovations like domain-specific hardware, enhanced security, open instruction sets, and agile chip development will lead the way

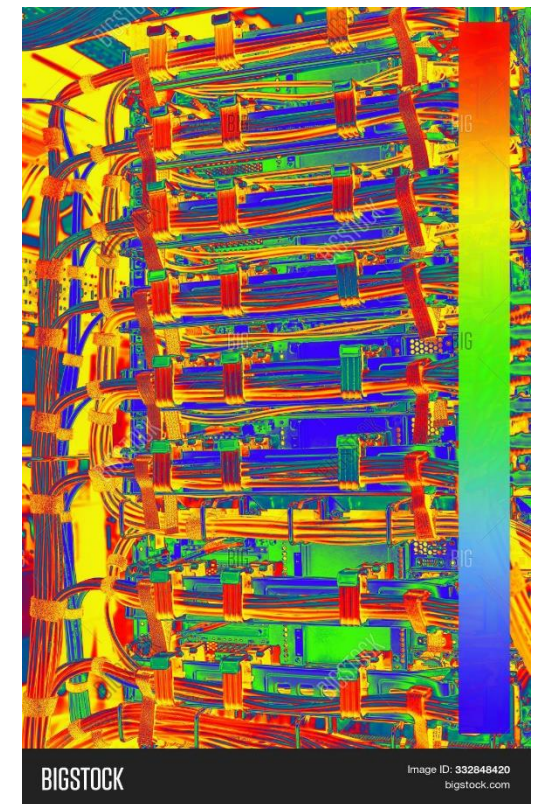
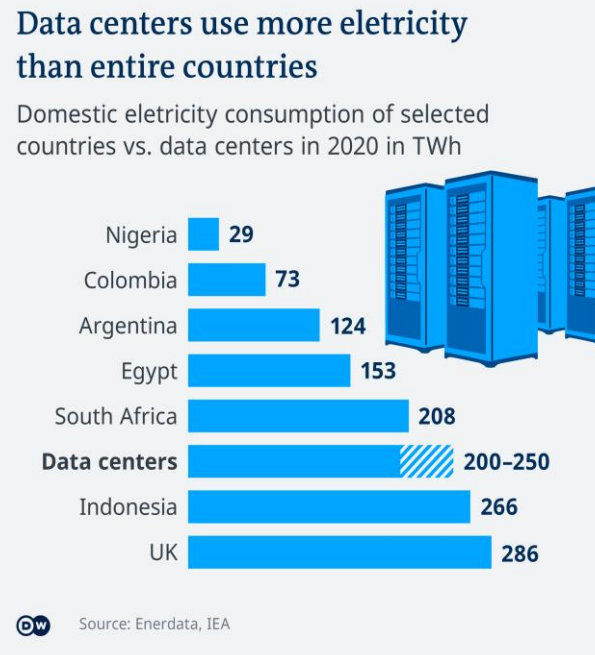
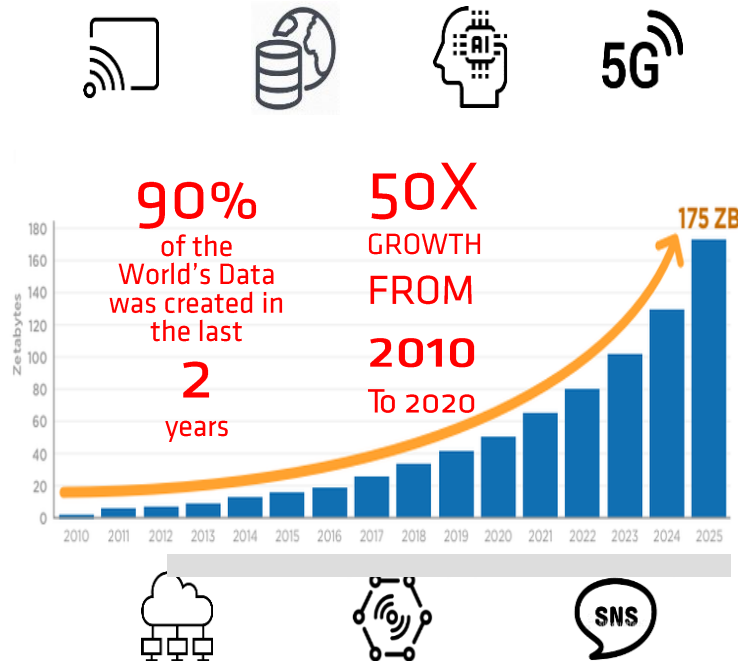
Implication: Without relying on the node shrinking, need to achieve performance and power requirement through architectural innovation



Source: "A New Golden Age for Computer Architecture"
From Turing lecture by John L. Hennessy and David A. Patterson in June 2018

Explosion of data demanding data-centric architecture, which in turn incurs serious power consumption and thermal problem

Implication: in order to minimize total cost of ownership (TCO), less power consumption and heat dissipation without sacrificing the performance is becoming more and more important

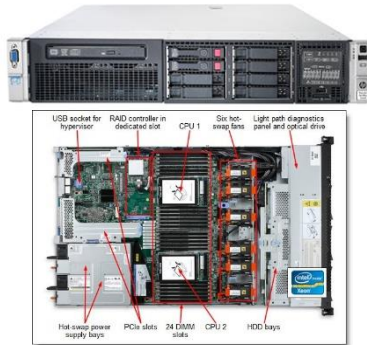


Source: Data age 2025, sponsored by Seagate with data from IDC global datasphere, Nov 2018,
Toward a memory-centric architecture, Keynote by Wester digital at flashmemorysummit 2017

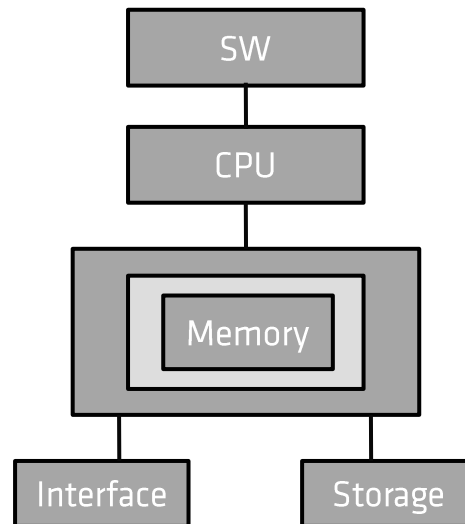
Computer System at scale: observation and implication

- Look and scope is different, but all of computer systems essentially can be abstracted as 1) compute, 2) memory, 3) storage, 4) network, 5) interconnect
- Even though Moore's law is ended over, the goal of every computer system is to achieve best service/application level performance with minimum TCO

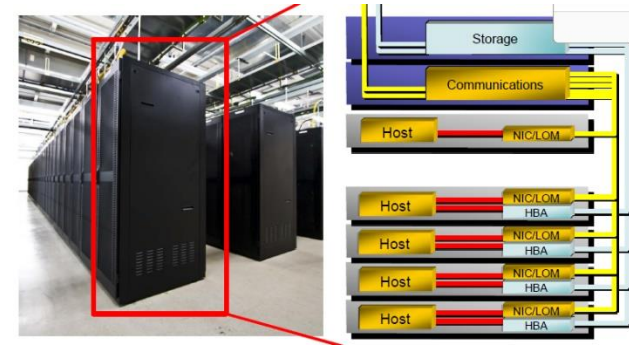
Server Scale



Abstracted computer system



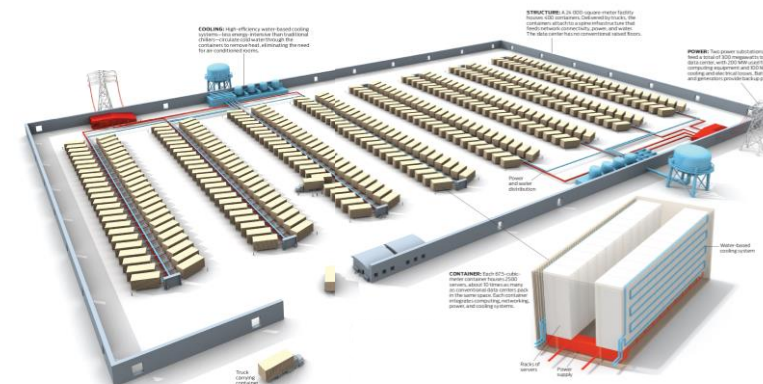
Rack Scale



Silicon & Device Scale



Data Center Scale

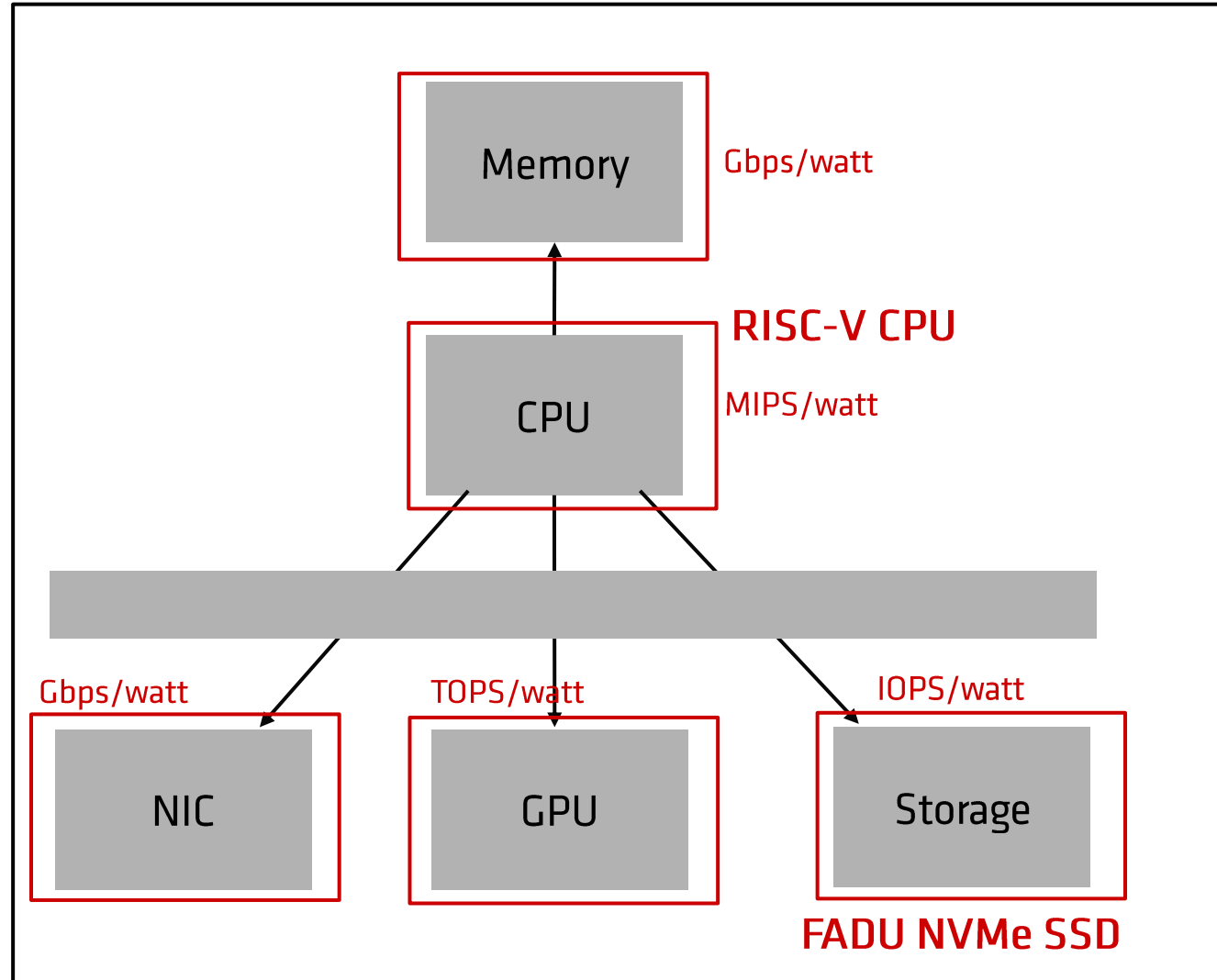


Optimization techniques for the computer system

Constraints & Requirements

- Cost increase should be minimized
- Power and thermal have become essential constraints
- Performance should be as high as possible

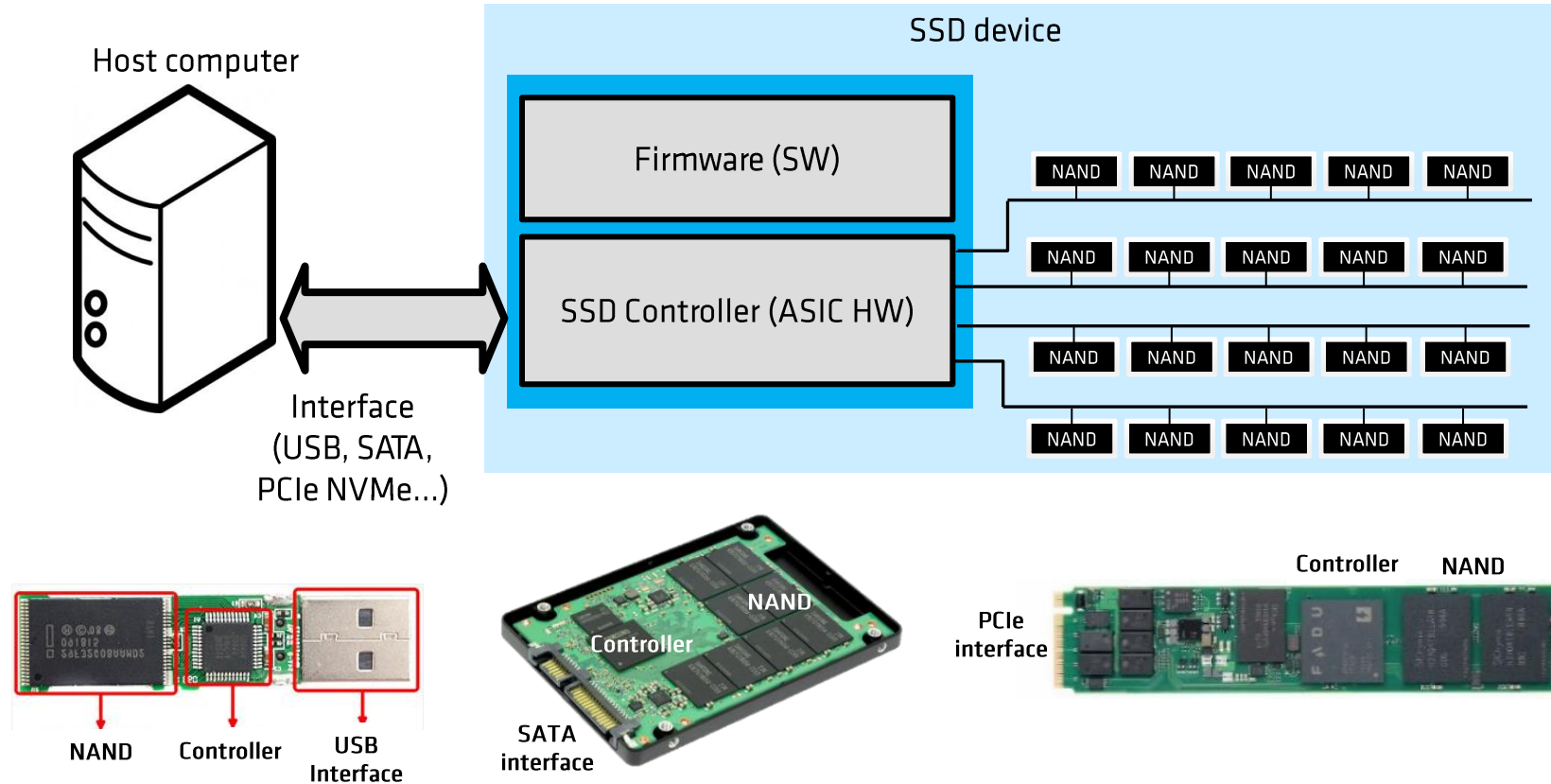
- Optimizing each component



Agenda

- FADU introduction and mega trend
- FADU NVMe SSD
- RISC-V CPU
- Next step

Understanding SSD – why controller is critical



Pros and Cons of NAND

- NAND is non-volatile memory
- NAND is very cheap! (DRAM 1GB = \$4 vs. NAND 1GB = \$0.07)
- NAND is very weak (easily worn out) and slow



NAND needs controller to be a product

(1) Boost performance through parallelizing NAND operations

(e.g. 4TB SSD has 512Gb NAND die X 64.
Write speed of one die = 50MB/s
while SSD should perform >2,000MB/s)

(2) To maintain the quality of performance throughout lifetime of SSD with dying NAND dies

Challenges => The rise of NVM express interface.



- The Interface speed is becoming incredibly faster than before
- Inherent peculiarities of Flash memory. It's getting worse
- Nobody cares about up to performance. What really matter is consistency in performance (QoS)
- At last, power and thermal becomes the most crucial hurdle in enterprise SSD design

- NVMe vs. SATA

	SATA	NVMe
IO speed	600 MB/s	16 GB/s (Gen5 case)
Maximum queue depth	Up to 1 queue; 32 commands per queue	Up to 64K queues; 64K commands per queue
Uncacheable register accesses (2000 cycles each)	6 per non-queued command; 9 per queued command	0~1 per command
MSI-X and interrupt steering	A single interrupt; no steering	2048 MSI-X interrupts
Parallelism and multiple threads	Requires synchronization lock to issue a command	No locking
Multicore Support	Limited	Yes

* IDF-2012, Danny Cobb, NVM Express and the PCI Express SSD Revolution

Challenges => NAND is becoming worse and worse

- The Interface speed is becoming incredibly faster than before
- Inherent peculiarities of Flash memory. It's getting worse
- Nobody cares about up to performance. What really matter is consistency in performance (QoS)
- At last, power and thermal becomes the most crucial hurdle in enterprise SSD design

The Good

- Low latency
- Low power consumption
- High Reliability
- Small form factor
- Massive parallelism

....



FROM *THE DARK NIGHT*

The Bad

- Power failures
- Bad blocks
(Program/Erase Errors)
- Program disturbance
- Read disturbance

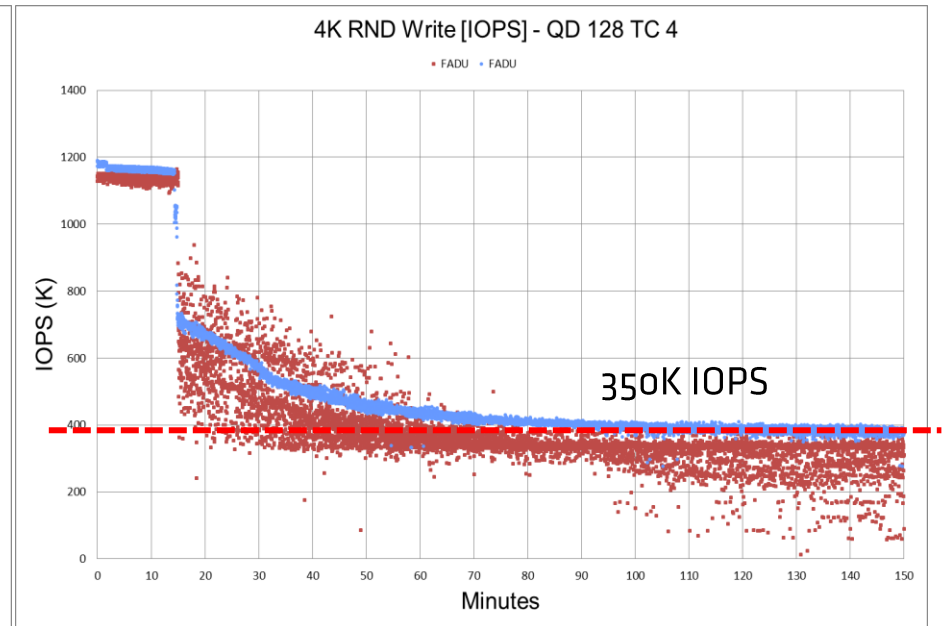
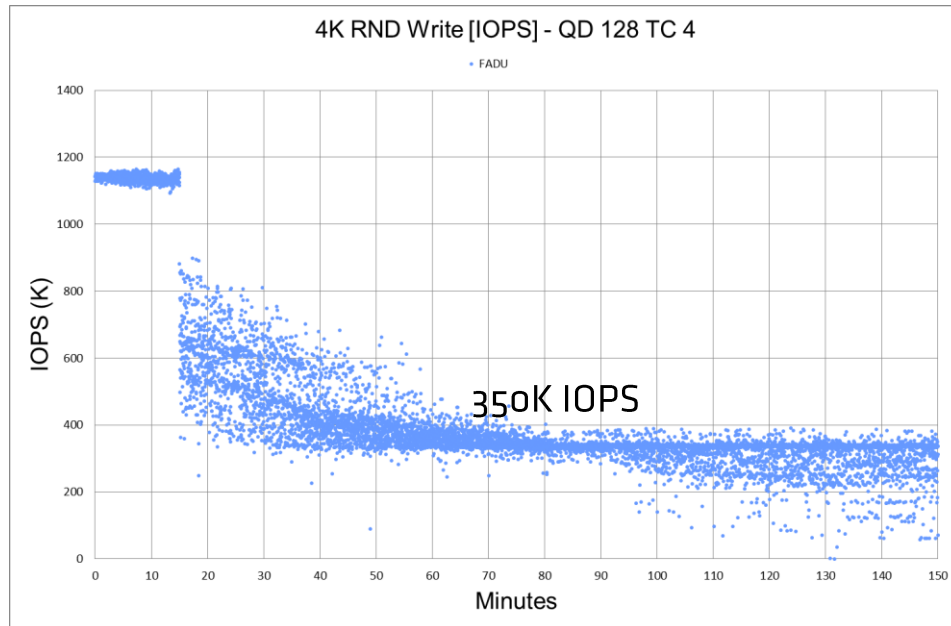
The Ugly

- Limited Endurance
- Retention errors

Node shrink, 3D stacking, and MLC technology further challenge the development of SSD controller

Challenges => Performance quality becomes matter

- The Interface speed is becoming incredibly faster than before
- Inherent peculiarities of Flash memory. It's getting worse
- Nobody cares about up to performance. What really matter is consistency in performance (QoS)
- At last, power and thermal becomes the most crucial hurdle in enterprise SSD design

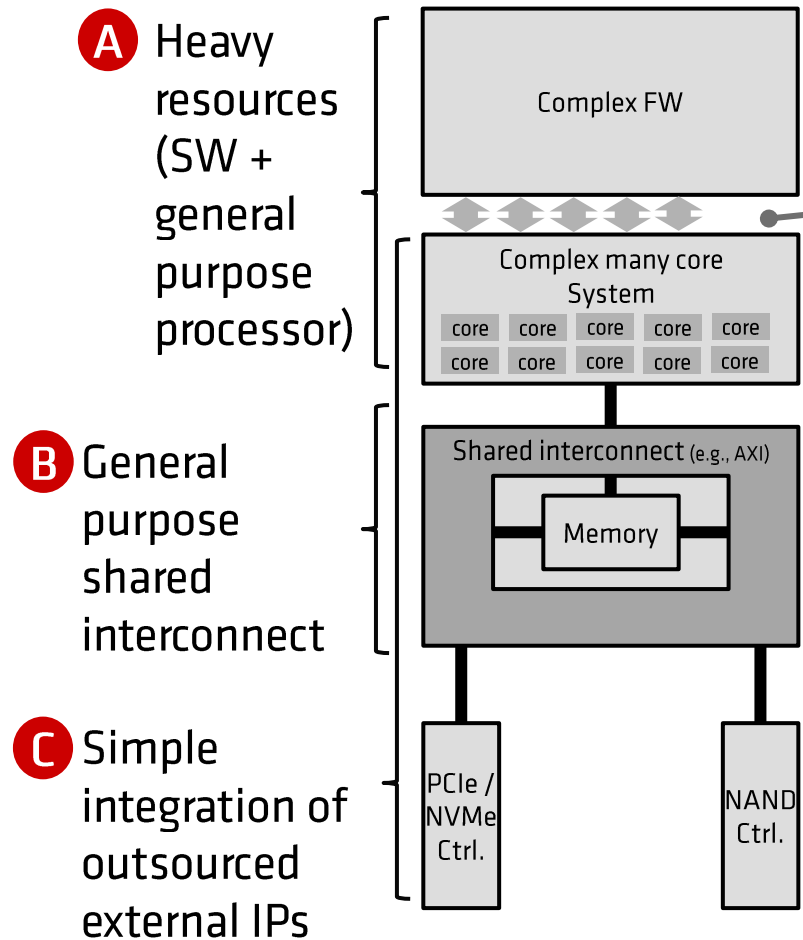


Challenges => Power consumption becomes critical barrier

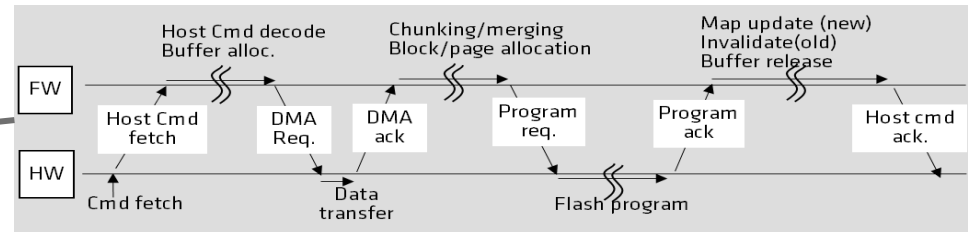
- The Interface speed is becoming incredibly faster than before
- Inherent peculiarities of Flash memory. It's getting worse
- Nobody cares about up to performance. What really matter is consistency in performance (QoS)
- **At last, power and thermal becomes the most crucial hurdle in enterprise SSD design**

	SATA SSD		Legacy high-end NVMe eSSD
Throughput (read)	6Gb/s = 600MB/s	4X	~3GB/s
IOPS (read)	80K	7X	~600K
Active power (read)	3W	7X	~20W
Idle power	1W	4X	~4W
IOPS/ Watt	27K / W	!!!	25K / W

In legacy architecture, entire operation is handled by software operated on top of heavy general purpose processors



1) Performance inefficient



2) Power and resource inefficient

Controller case study:

- ✓ 32 Cores @500MHz each w/ 192KB SRAM
- ✓ Logic and SRAM consume huge area power
- ✓ More than 15W per a controller

3) Unpredictable & Unbalanced

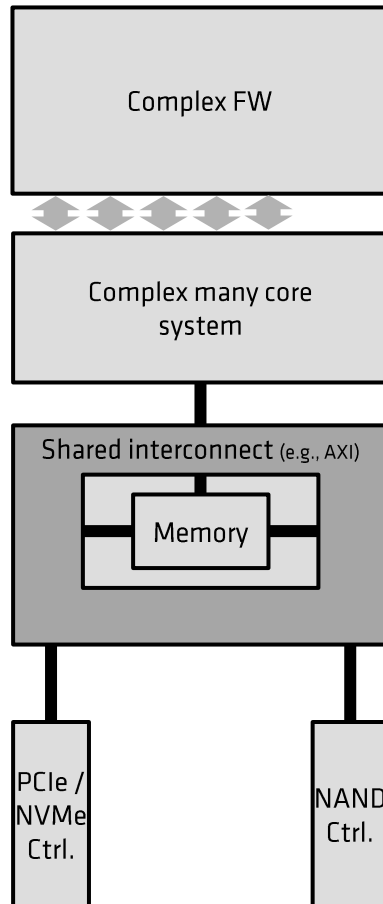
- ✓ Computation load unbalance among CPUs
- ✓ Bus congestion
- ✓ Interrupt latency and priority
- ✓ Context switch, Cache miss, and Control/data hazard
- ✓ Error-prone parallel programming

➔ **Fundamental limitation in Performance/watt as well as performance/\$**

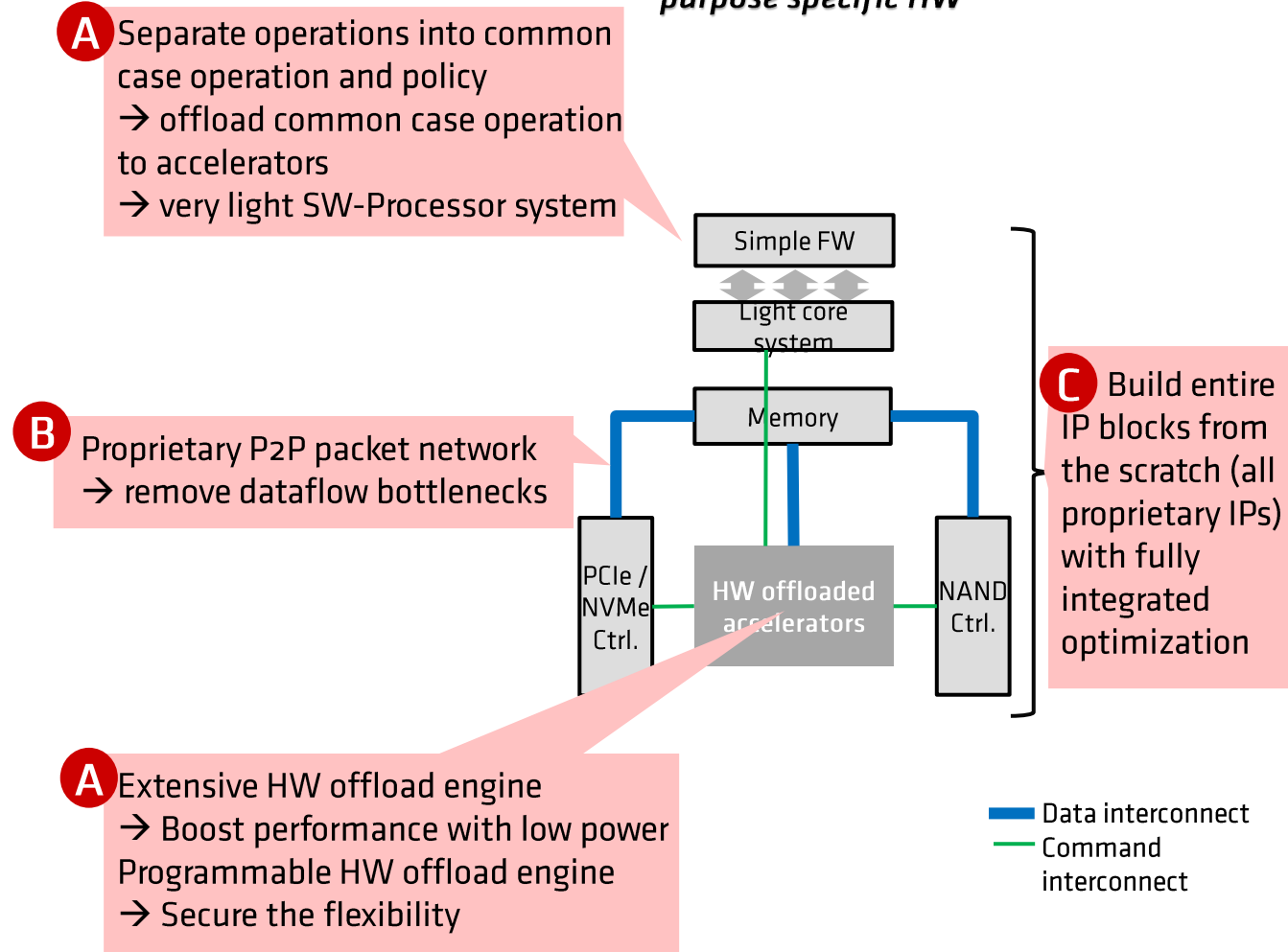
We introduced the innovation

Achieved application specific(SSD) optimization for the HW/SW full stack from silicon to the device scale

Entire operation is handled by Software operated on top of general purpose processors



Heavy SW-general HW operation is offloaded to purpose specific HW

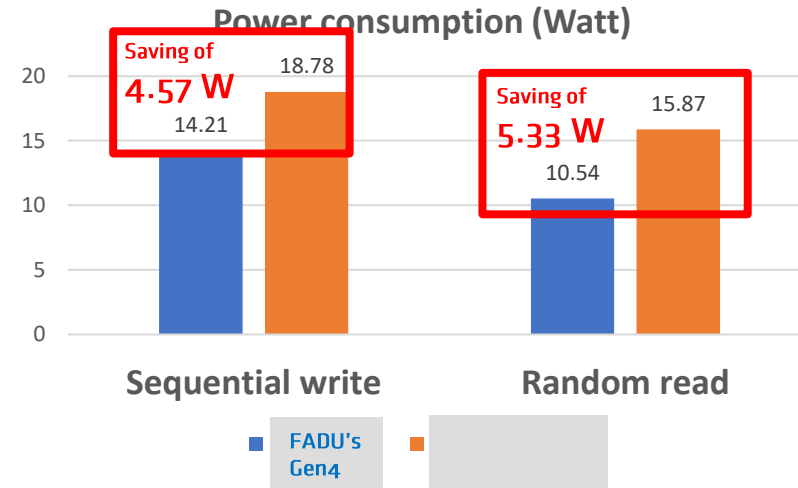
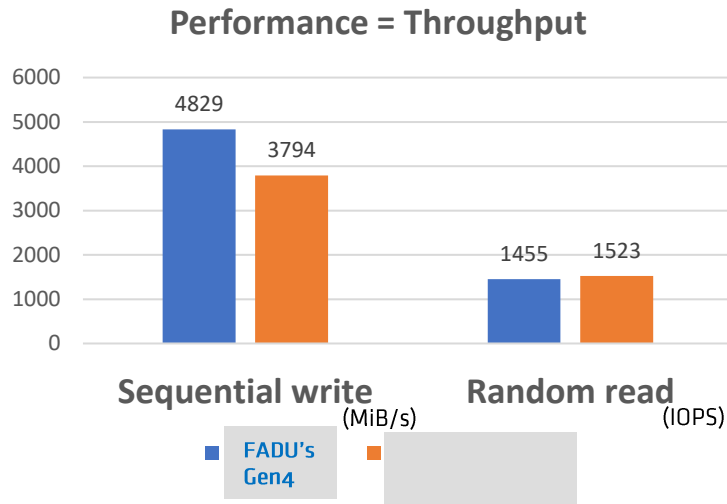


Our 2 delivered products clearly confirmed that our approach worked

	PCIe Gen3 (FADU 1 st Gen)			PCIe Gen4 (FADU 2 nd Gen)		
	FADU Bravo	Intel P4511	Samsung PM983	FADU Delta	Intel D7-5510	Samsung PM9A3
Sequential Read (GB/s)	3,500	2,000	3,000	7,300	6,500	6,500
Sequential Write (GB/s)	2,700	1,430	1,400	4,600	3,400	3,500
Random Read (K IOPS)	800	295	480	1,550	700	900
Random Write (K IOPS)	100	36	47	185	170	180
Power consumption (W)	6.0	8.25	7	13	18	13
Performance/power (K IOPS/W)	133	35.8	60	119	36	69
Controller size (mm ²)	26		37			

- FADU's the first ASIC proved superior competency against existing market leaders – Intel and Samsung → beat them in all key metrics
- FADU's advantage sustained even in 2nd ASIC

Key thesis: Industry needs “Low Power”=“Low TCO” Solution



Case study: OCP storage server rack

4 watts saving per drive
16 drivers per server
21U servers per rack (50%)

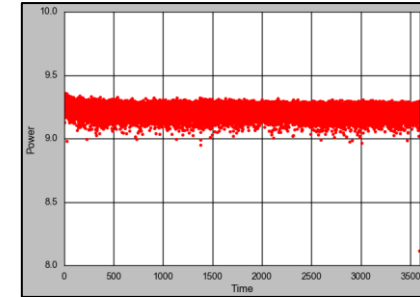
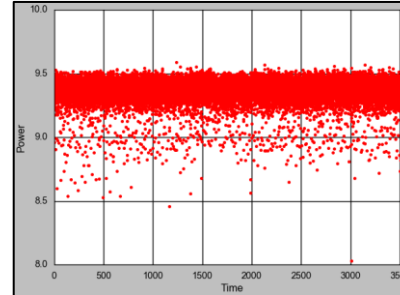
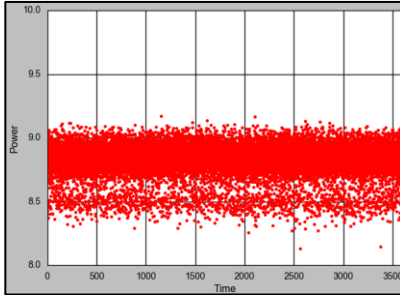
= **total 1344 watt (4x16x21) saving per rack**



1U server typically consumes
< 1000watt

Key thesis: Industry needs “*Low Thermal*” Solution

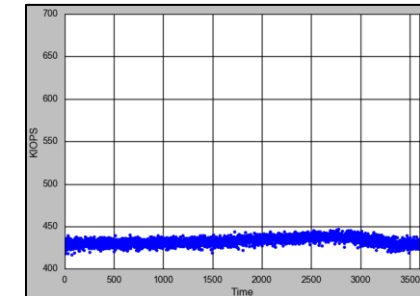
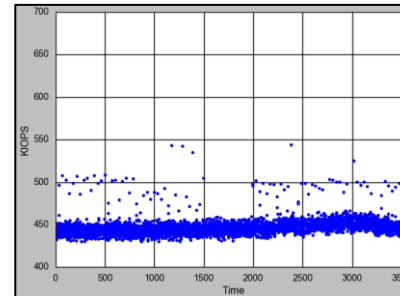
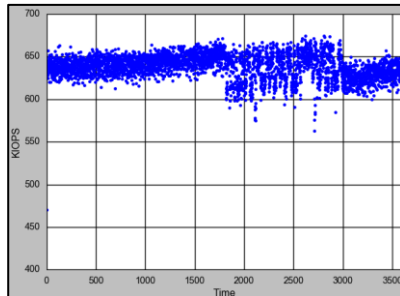
R9oW1o
Power
Trend
(W)



Hot spot
Temp(C)



R9oW1o
(IOPS)



FADU Delta

Competitor A

Competitor B

- Low thermal → High reliability → High availability
- Low thermal → Low cooling cost → Low TCO
- Low thermal → Low power → Low TCO

Agenda

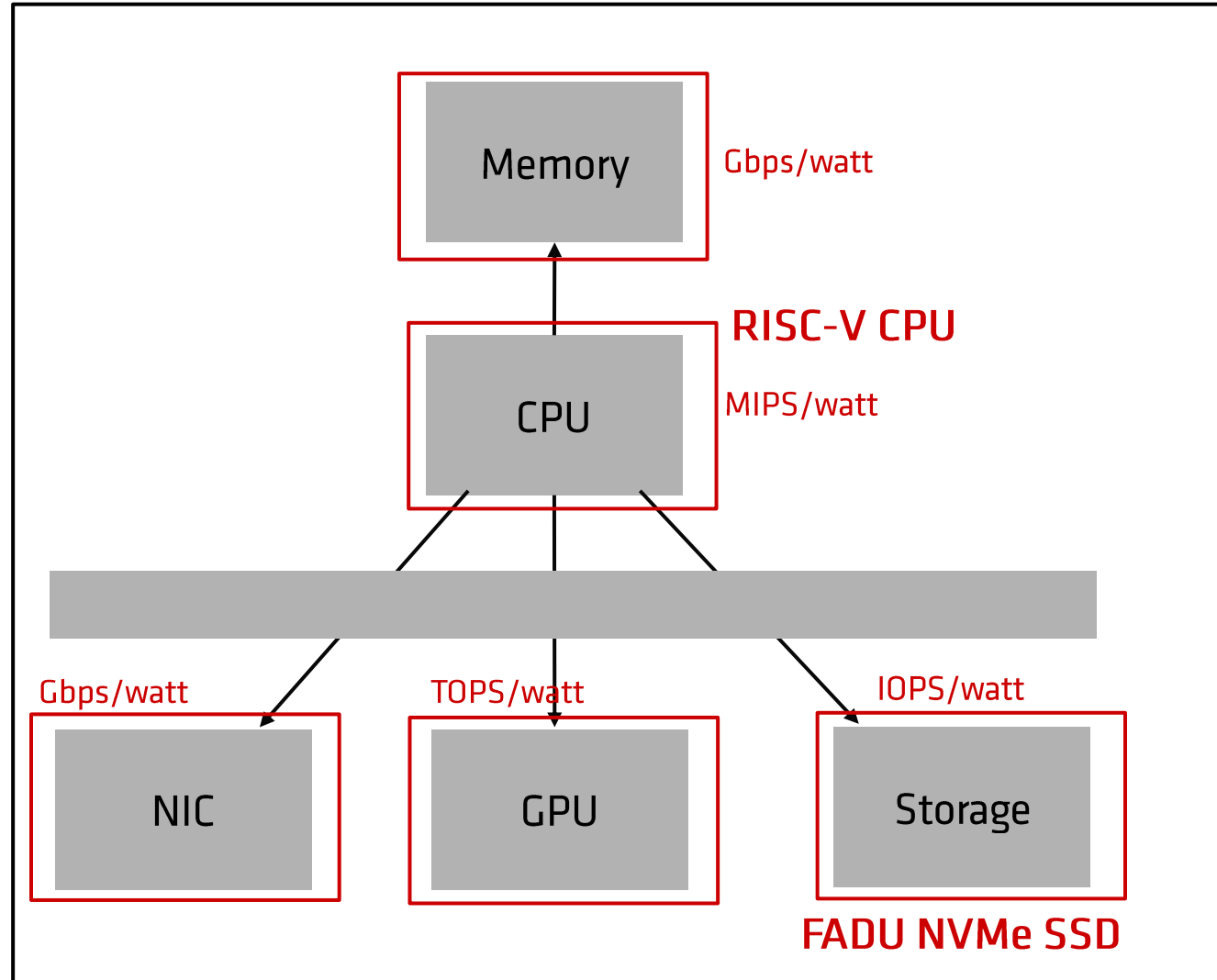
- FADU introduction and mega trend
- FADU NVMe SSD
- RISC-V CPU
- Next step

Optimization techniques for the computer system

Constraints & Requirements

- Cost increase should be minimized
- Power and thermal have become essential constraints
- Performance should be as high as possible

- Optimizing each component



FADU controller overview

- Most of the IPs (in red box) are developed from scratch several times to squeeze area and power consumption
- Die size is presumed around 30% smaller than that of competitors

PCIe

- 1 Gen3x4 or 2 Gen3x2
- 1 PF & 15 VF per port
- SRIS
- L1.2 and DPA
- MSI/MSI-X/legacy
- VPD
- More

Processor subsystem

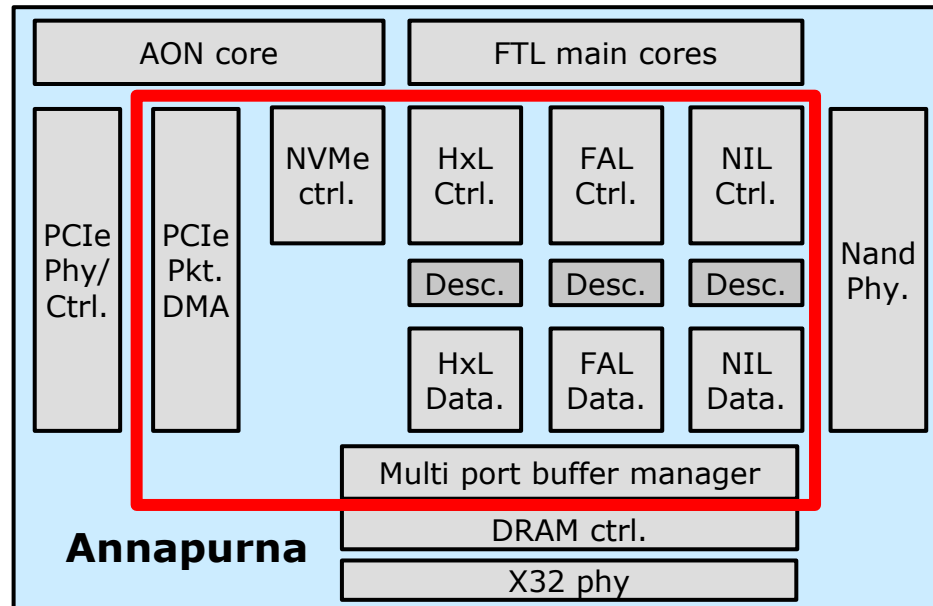
- Triple 64-bit RISC-V cores
- QSPI flash, I2C & SMBUS
- Temp sensor
- UART, GPIO, and various timer
- More

Reliability, Availability, and Serviceability

- ECC protection on SRAM/DRAM (SECDED)
- Host meta & T10 DIX/DIF
- PCIe AER & ECRC
- More

NVMe

- 128 Namespaces
- 32 streams
- SGL & PRP
- CMB/PMR
- Reservation
- Virtualization mgmt.
- 512KB Atomicity
- 2MB MDTs
- NVMe-MI
- More



NAND

- 8 ch.x8 CE (max. 32 LUNs/ch)
- ONFi 4.0/Toggle 3.0 up to 800 MT/s
- TLC/MLC/SLC
- 4KB/8KB/16KB page size
- More

DDR memory

- LPDDR4, up to 4GB @ 2.8 Gbps
- DDR4, up to 8GB @ 2.8 Gbps
- Single X32 w/ in-line ECC
- More

Core SSD IPs

- Dedicated control & data packet network
- Programmable control plane
- Various hardware accelerators for optimizing common case operation
- 4K LDPC ECC
- In storage RAID
- Configurable data randomizer
- Parity check for FTL metadata
- AES128/256-XTS, TRNG
- More

The first meaningful adoption of RISC-V CPU in enterprise class ASIC

PCIe

- 1 Gen3x4 or 2 Gen3x2
- 1 PF & 15 VF per port
- SRIS
- L1.2 and DPA
- MSI/MSI-X/legacy
- VPD
- More

NVMe

- 128 Namespaces
- 32 streams
- SGL & PRP
- CMB/PMR
- Reservation
- Virtualization mgmt.
- 512KB Atomicity
- 2MB MDTs
- NVMe-MI
- More

Processor subsystem

- **Triple 64-bit RISC-V cores**
- QSPI flash, I2C & SMBUS
- Temp sensor
- UART, GPIO, and various timer
- More

Reliability, Availability, and Serviceability

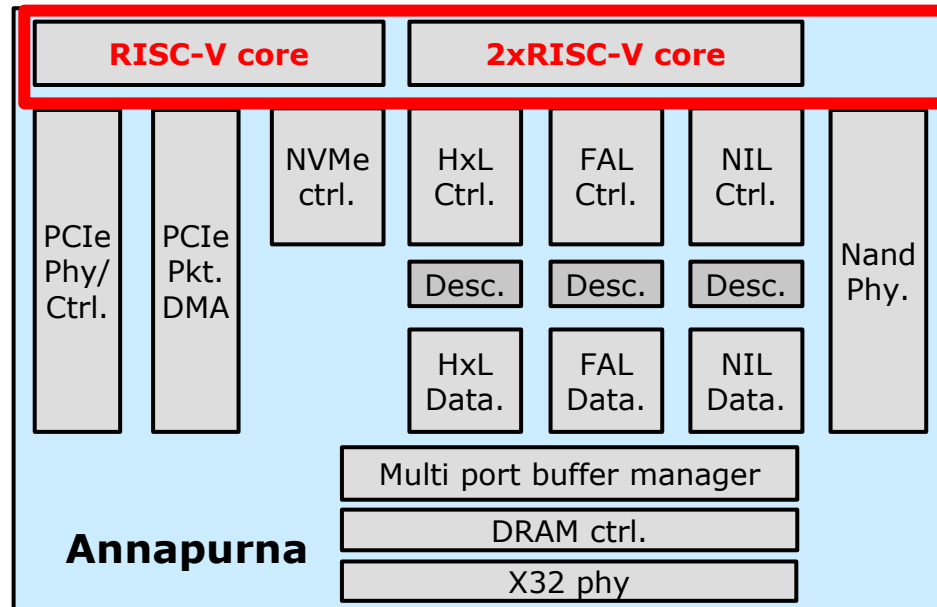
- ECC protection on SRAM/DRAM (SECDED)
- Host meta & T10 DIX/DIF
- PCIe AER & ECRC
- More

NAND

- 8 ch.x8 CE (max. 32 LUNs/ch)
- ONFi 4.0/Toggle 3.0 up to 800 MT/s
- TLC/MLC/SLC
- 4KB/8KB/16KB page size
- More

DDR memory

- LPDDR4, up to 4GB @ 2.8 Gbps
- DDR4, up to 8GB @ 2.8 Gbps
- Single X32 w/ in-line ECC
- More



Core SSD IPs

- Dedicated control & data packet network
- Programmable control plane
- Various hardware accelerators for optimizing common case operation
- 4K LDPC ECC
- In storage RAID
- Configurable data randomizer
- Parity check for FTL metadata
- AES128/256-XTS, TRNG
- More

Considered CPUs for SSD: Pros and Cons of conventional CPUs



arm



ARC
Synopsys



cādence



tensilica

Pros

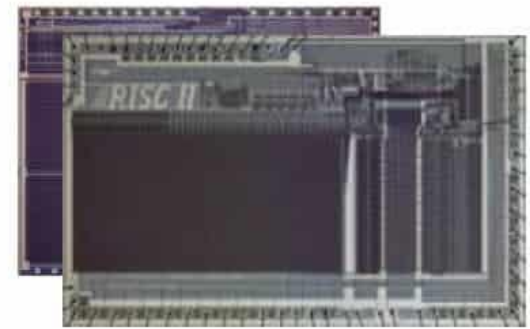
- Very mature and stable
- Superior SW development environment (IDE & SDK)

Cons

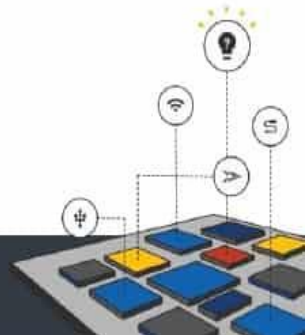
- Too expensive
- Too large
- Too slow and old
- Not desperate (to small company)
- Poor configurability and customizability
- Not support for native 64bit addressing

What is RISC-V

RISC-V is a high-quality, license-free, royalty-free ISA



- 5th Generation RISC design from UC Berkeley
- A high-quality, license-free, royalty-free RISC ISA specification
- Standard maintained by non-profit RISC-V Foundation
- Multiple proprietary and open-source core implementations
- Supported by growing software ecosystem
 - binutils/gcc/FreeBSD mainlined, Linux/glibc submitted to upstream
- Appropriate for all levels of computing system, from microcontrollers to supercomputers



Why did I choose RISC-V and SiFive?



Cons

■ Not proven and SW dev environment was suck

■ Debugging Engine

- Thread2 & thread3 have the same context (eg. \$PC) in GDB

```
(gdb) c
Continuing.
^C
Thread 3 received signal SIGINT, Interrupt.
0x0000000028016404 in uart_get_char_try () at uart.c:16
16      fifo = UART0_REG(UART_REG_RXFIFO);
(gdb) info thread
Id      Target Id      Frame
1      Thread 1 (hart 1) main_core0 () at core_1_main.c:252
2      Thread 2 (hart 2) 0x0000000028016404 in uart_get_char_try () at uart.c:16
* 3      Thread 3 (hart 3) 0x0000000028016404 in uart_get_char_try () at uart.c:16
```

- Downloading executable thru JTAG is too slow ($\approx 1\text{KB/s}$)
- Upper 32bit address space cannot be accessed thru JTAG
- Break points is applied to all cores
 - One core stops at break point, other cores also stop.

■ Debugging Engine (Cont'd)

- Openocd is occasionally disconnected with “unexpected wake up on hart 0” when SSD FW is running on GDB
- When SSD FW is compiled with O2 option, it halts with openocd disconnection.
- When SSD FW halts for unknown reasons, we press ctrl+c in GDB command shell to debug SSD FW, then openocd is also disconnected.

```
root@test-desktop: ~/glt/freedom-e-sdk
Error: Trying to use configured scan chain anyway...
Warn : Bypassing JTAG setup events due to errors
Info : accepting 'gdb' connection on tcp/3333
Error: unable to halt hart 0
Error: dmcontrol=0x00000001
Error: dmstatus =0x0030c82
./jtagrun: line 3: 14807 Aborted                    openocd -f openocd.cfg
root@test-desktop: ~/glt/freedom-e-sdk

root@test-desktop: ~/glt/freedom-e-sdk
[Switching to thread 1 (Thread 1)]
#0  _start () at entry.S:6
0      la t0, trapvec
[Switching to thread 2 (Thread 2)]
#0  0x0000000000000000 in ?? ()
---Type <return> to continue, or q <return> to quit---
[Switching to thread 3 (Thread 3)]
#0  0x0000000000000000 in ?? ()
(gdb) c
Continuing.

^CRemote connection closed
(gdb)
```

■ Toolchain infra

- Currently SiFive SDK is only running on ubuntu OS
 - CentOS 6.8 / Windows ...

■ Considerations for the future

- How to support debugging when PMU core & main core time?
 - In GDB, multi-process debugging (PMU core & main core)
 - Heterogenous CPU debugging (PMU – E31, Main – E51 core)
 - ✓ We are also seriously considering E51 as PMU core if that of debugging
- Currently, SiFive releases all the toolchains as source and compiles all toolchains.
 - It takes long time to build all toolchains.
 - It would be better to release all the toolchains as installer
 - It also helps to avoid the human error to install the toolchain

■ Convenience

- GUI support is strongly required

Why did I choose RISC-V and SiFive?



Cons

- Not proven and SW dev environment was suck ... but

Pros (necessary condition)

- Cheap and small
- Good configurability and customizability
- Support for native 64bit addressing

Pros (sufficient condition)

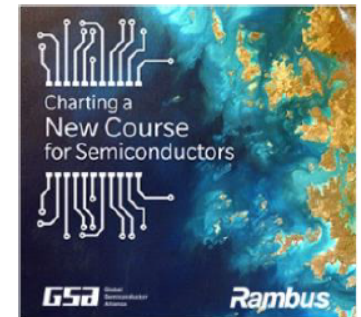
- Personal preference because RISC-V is a CPU with a lineage in CPU history
- Very desperate
- Very fast and new (no legacy)
- Same design principle even in a different domain
- Korean is founder and CTO
- The ISA in the textbook was changed from ARM to RISC-V at that time, and received a gift

Why did I choose RISC-V and SiFive?

RISC-V – “Linux” of Microprocessors

Free, open, extensible instruction set architecture (ISA) for **all** computing devices

- RISC-V foundation now >50 members
- Why RISC-V ?
 - Simple, free and open
 - Architected for security
 - Architected for customization—great for **AI & Machine Learning**
- SiFive is the center of gravity
 - SiFive Platforms are treated as RISC-V standards
 - Robust, open-source Software Ecosystem means less software investment required from within SiFive—lower cost, faster scale out
 - Everybody interested in RISC-V talks to SiFive
 - Passionate and huge fan base of early adopters



**Global Semi Alliance:
RISC-V is Linux of
Microprocessors**



Source: from company introduction material of SiFive

Why did I choose RISC-V and Si-Five?

Prof. Krste Asanovic'

Founders of Si-Five

Prof. David Patterson



Key members of RISC-V foundation

Why did I choose RISC-V and SiFive?

Who is SiFive?

Best-in-class team with technology depth and breadth

Founders & Execs



Yunsup Lee
CTO



Krste Asanovic
Chief Architect



Andrew Waterman
Chief Engineer



Jack Kang
VP Product/BD



Stefan Dyckerhoff
Interim CEO, VC



Sander Arts
CMO

Key Leaders & Team



Han Chen
Chip Implementation



Ali Habibi
Verification



Source: from company introduction material of SiFive

Why did I choose RISC-V and Si-Five?

Same observations even in a different domain



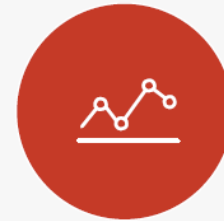
“Moore’s Law” is only for the rich

- Cost per transistor **no longer shrinking**
- **Exploding design cost, time complexity**



Computing is changing

- **IoT & cloud demanding better perf/Watt**
- **IoT, machine learning and security** require customization



Semi Biz Model Failing

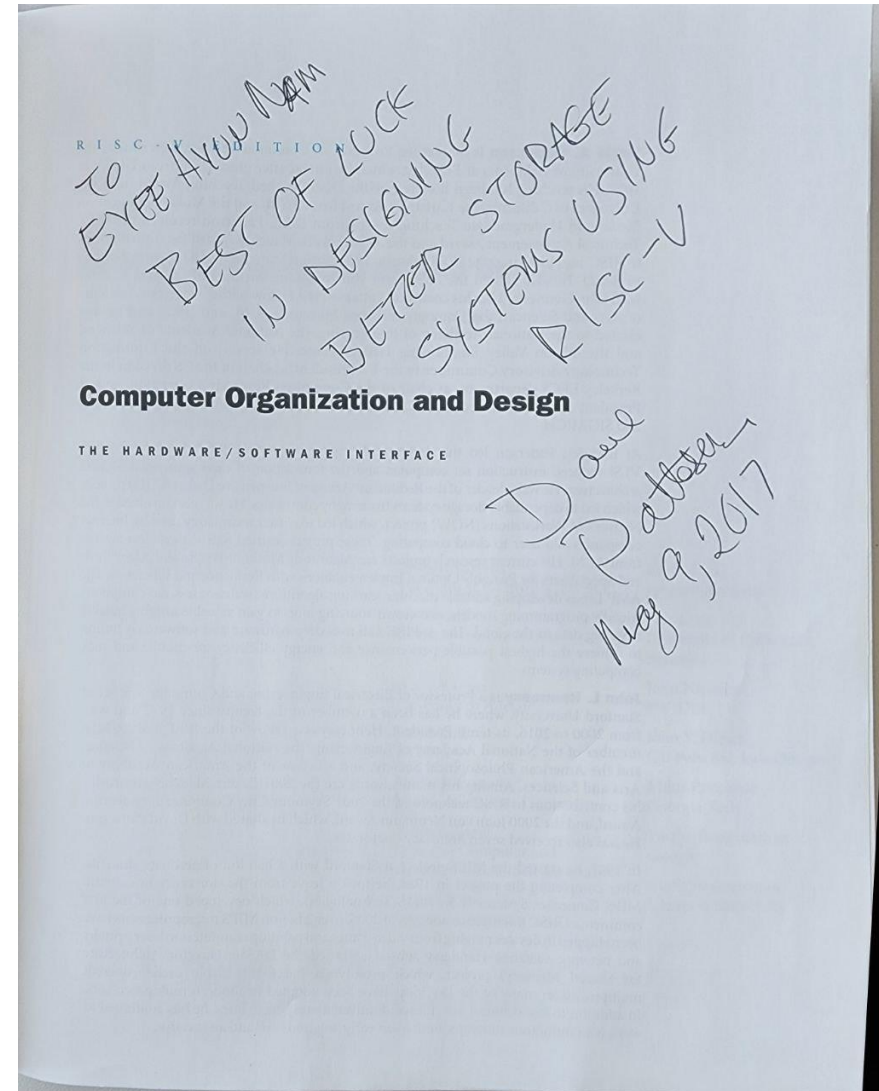
- ARM monopoly **crippling system innovation**
- Biz model **not aligned to demand from cloud & IoT**

Why did I choose RISC-V and Si-Five?

Why **SiFive Coreplex** IP?

- Designed by the inventors of RISC-V
- Leading-edge RISC-V support
 - Supports the latest RISC-V standards, such as Compressed Mode
- Compatibility with current and future RISC-V specifications
- Better Perf/Watt than ARM
- More Efficient Designs than ARM
- Better for Machine Learning and AI than ARM

Why did I choose RISC-V and Si-Five?



Success story

August 07, 2018

FADU Lo Leading Power V Core l

SAN MATEO, Calif. and s
leading provider of cor
company developing s
market, today announ
Controller and FADU Br
industry leading 64-bit

The FADU Annapurna S
Controller and provide
among its peers, while
FADU Annapurna, the F
supporting dual port a
lower power, and the n
consuming only 6-8W
competing solutions a
memory controller, and
embedded RISC-V Cor

"FADU is focused on bu
devices for our custom
Lee, CEO, FADU. "SiFive's
of competing solutions
optimizing our architec

Tweet



Congratulations @FaduTechnology or
announcement of Fadu's PCIe 5 x 4 /
OCP Cloud Spec 2.0 SSD Controller a
solutions, featuring the SiFive Essenti
Read more from @Tom's Hardware he
hubs.la/HO_6MS80
#sifive #riscv #essential #ocp

tom's**HARDWARE** US Edition

Home > News

FADU Technology Announces Speed
GBps PCIe 5.0 SSD, up to 3.5 Million

By Aaron Klotz · 2 days ago

New SSD offers nearly twice the bandwidth of Gen 4



(Image credit: Shutterstock)

Just as mainstream computers begin adopting PCIe Gen 5 support, FADU Technology has announced its first-ever PCIe Gen 5 SSD and SSD controller designs. According to GlobeNewswire, these SSDs will target 14.6 GB/s of performance, consume less power, and will be designed specifically for the cloud and data center market.

7:08 AM · Nov 13, 2021 · HubSpot

[단독] SK하이닉스, 메타(페이스북)와 SSD 공급 계약 체결

벤처기업 '파두'의 반도체 컨트롤러 차용한 SSD 공급

장문서 기자
입력 2022.01.25 11:25



SK하이닉스 이천캠퍼스 /SK하이닉스 제공

SK하이닉스가 미국 최대 소셜미디어(SNS) 기업 메타(구 페이스북)에 비휘발성 인터페이스 메모리(NVMe) 기반의 솔리드스테이트드라이브(SSD) 공급 계약을 체결한 것으로 알려졌다.

25일 금융투자업계에 따르면 SK하이닉스(93,500원 ▲ 4,400 4.94%)는 메타와 비휘발성 인터페이스 메모리(NVMe) 기반 저장장치(SSD) 계약을 체결했다. 정확한 금액 규모는 알려지지 않았지만, 수천억원에 달하는 것으로 추정된다.

SK하이닉스가 NVMe 방식의 SSD 제품으로 메타를 고객사로 확보한 것은 이번이 처음이다. 메타는 메타버스 사업을 준비하면서 클라우드용 SSD 수요가 늘어난 상태. SSD는 반도체를 이용해 데이터를 저장해 자기디스크를 이용하는 하드디스크(HDD)보다 데이터를 읽거나 쓰는 속도가 더 빠르다는 장점이 있다. 이번 계약에서 데이터용 SSD 제품 규격 등 주요 사안에 합의했다.

이번 계약과 관련해 반도체 설계를 담당하는 팍리스 스타트업 파두(FADU)가 핵심 역할을 했다. SK하이닉스가 파두가 제작한 핵심 반도체인 컨트롤러를 차용해 메타와 계약했기 때문이다. 파두는 SSD에 들어가는 핵심 반도체인 컨트롤러를 개발하는 업체다. 국내에서는 삼성전자 등이 SSD 컨트롤러를 독자적으로 설계하는 능력을 갖췄다.

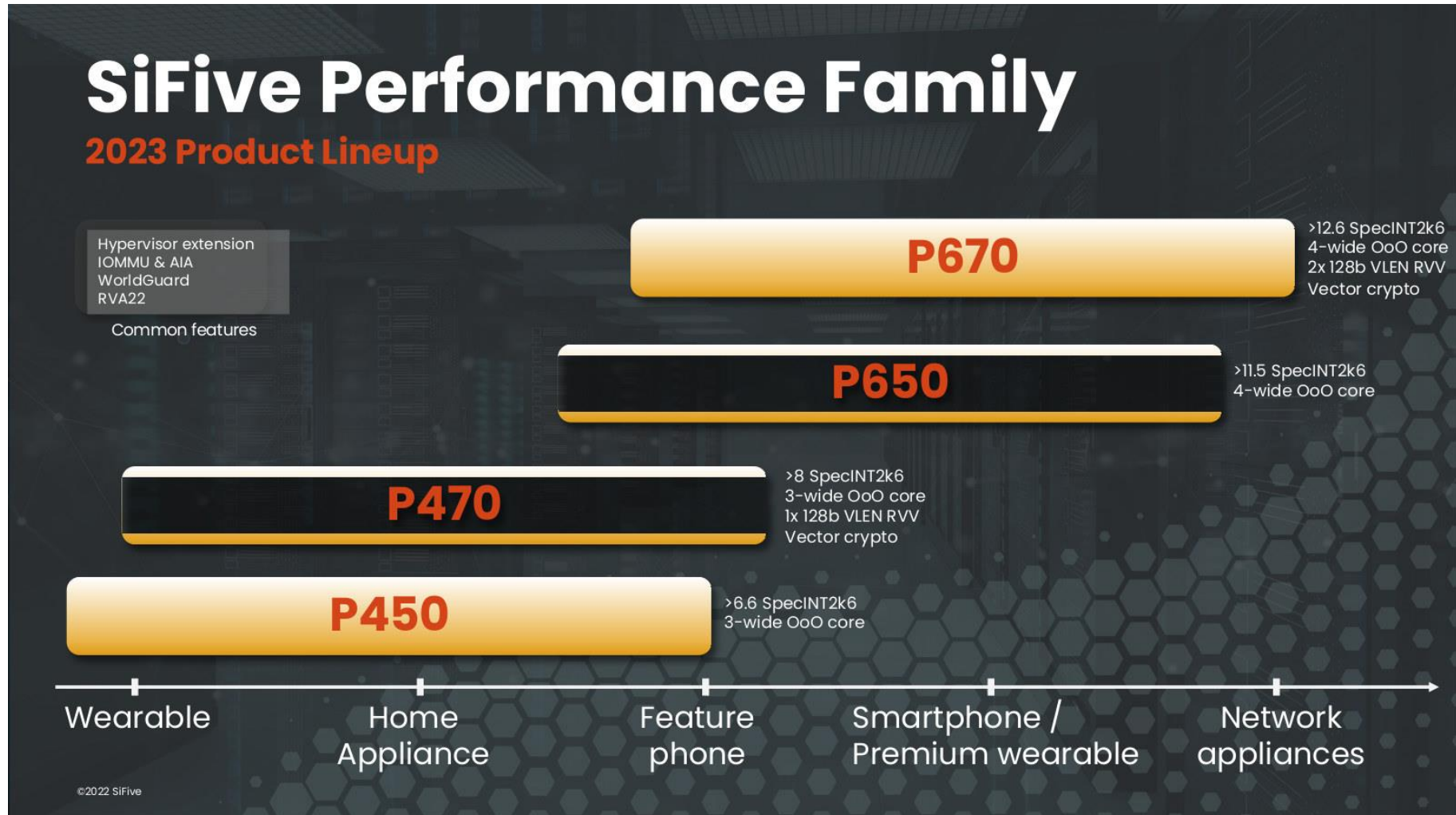
Samsung
Newsroom

삼성전자의
다양한 소식을
빠르게 만나볼 수
있습니다.

많이 본 뉴스

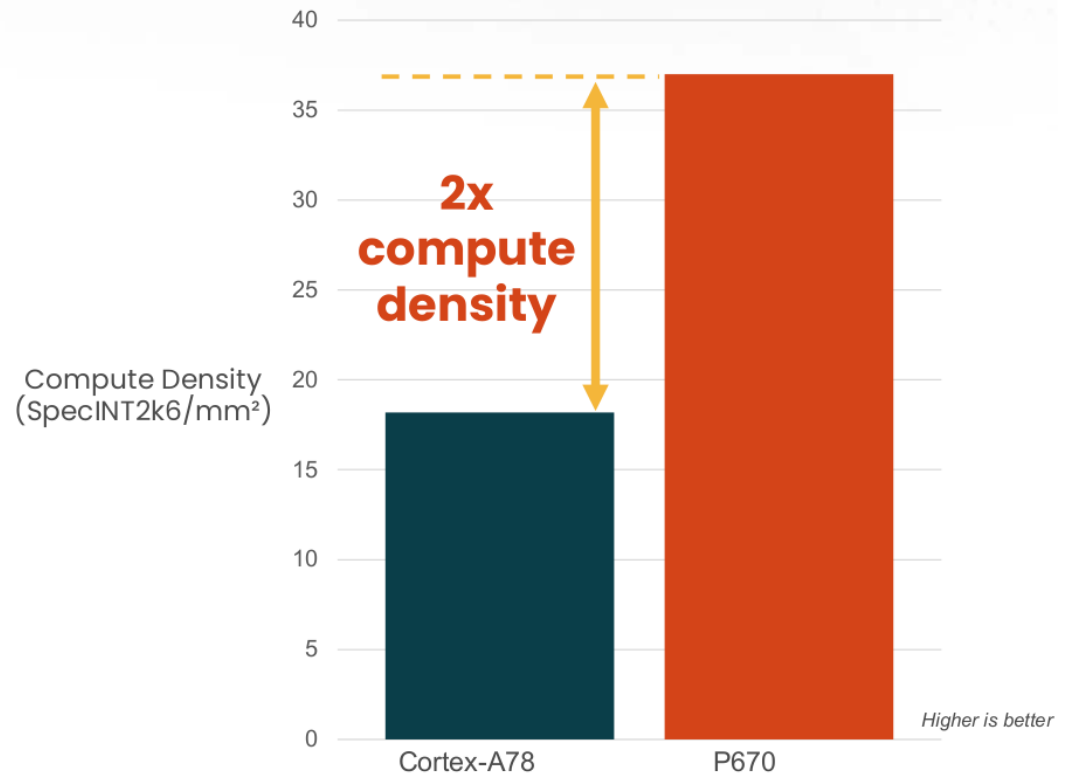
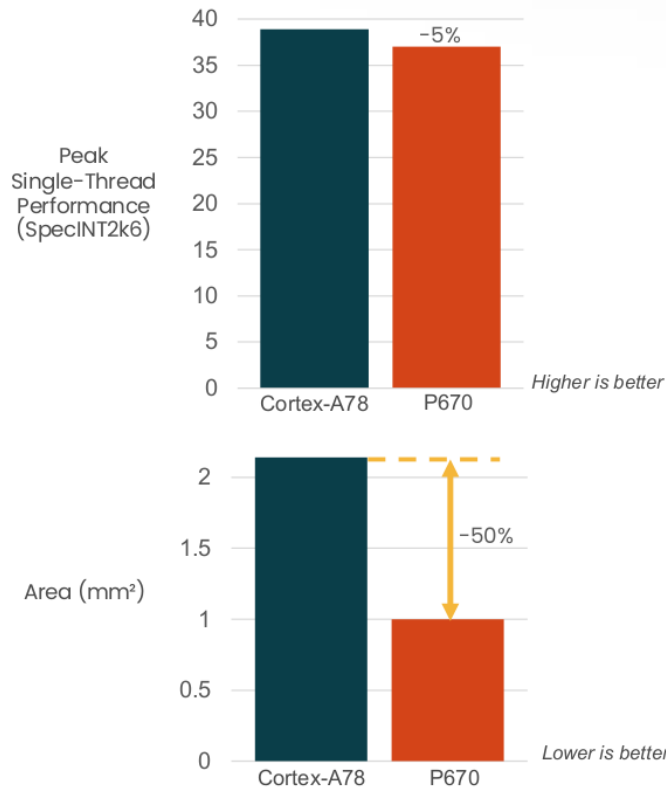
- 저축은행, 적금 금리도 시중은행에 추월 당해... "더 올리고 싶지만"
- 한전 대규모 전차에 채권시장 교란... 끝없는 탈원전 나비효과
- 로또 청약은 열광...수도권 아파트 미계약 물량 작년 대비 3배
- "순, 발이라도 그림니다"...시에 밀려 설 곳 잃은 화가들
- 인천 송도동 송도예유포레 푸르지오아파트 84㎡ 7억2000만원에 거래
- '지스타 2022' 모바일 넘어 K콘술 신작 대거 공개되다

Prospect on the RISC-V technology



Prospect on the RISC-V technology

P670 Performance & Efficiency

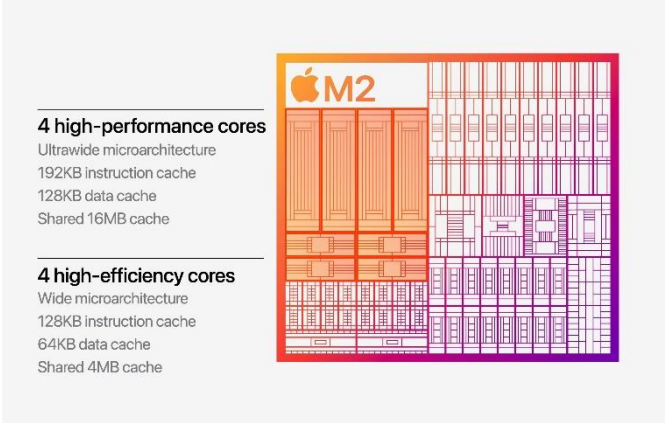
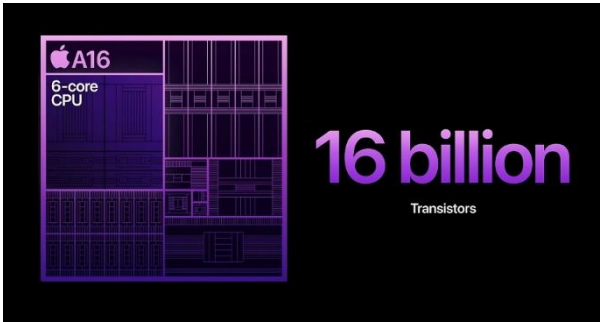
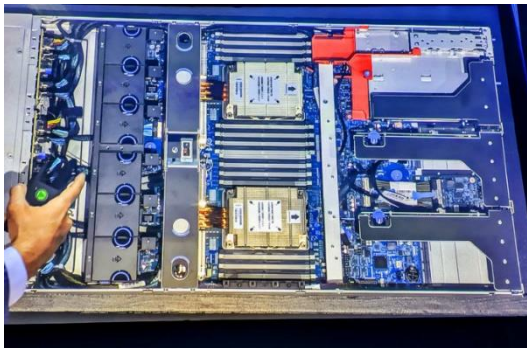


The Cortex-A78 was first used in the Samsung Exynos 1080 and 2100 SoC, introduced in November and December 2020 respectively.^{[7][8]} The custom Kryo 680 Gold core used in the Snapdragon 888 SoC is based on the Cortex-A78 microarchitecture.^{[9][10]} The Cortex-A78 is also used in the MediaTek Dimensity 8000 series.

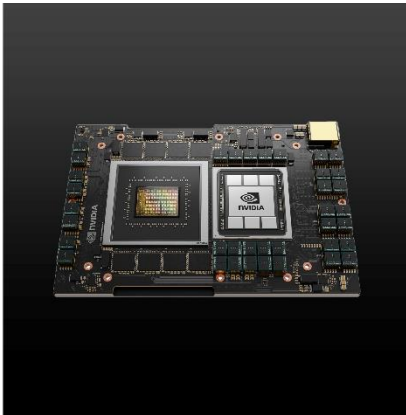
Samsung Galaxy S21/Samsung Galaxy S21+/Samsung Galaxy S21 UltraSamsung Galaxy S21 FE/Samsung Galaxy Z Fold 3/Samsung Galaxy Z Flip 3

Source: <https://www.cnx-software.com/2022/11/02/sifive-p670-and-p470-risc-v-processors-add-risc-v-vector-extensions/>

Prospect on the RISC-V technology



- NVIDIA GRACE**
Datacenter Ready
- NVIDIA's First Server CPU
 - 72 Arm v9.0 cores
 - SVE2 support
 - Virtualization Extensions: Nested Virtualization, 5 EL2 support
 - RAS v1.1
 - GIC v4.1
 - SMMU v3.1
 - Built on TSMC 4N process node



expect to the rise of RISC-V based mobile phone, laptop, and server subsequently within a couple of years



Blog

RISC-V Celebrates Upstreaming of Android Open Source Project RISC-V Port | RISC-V International

By RISC-V Community News | October 23, 2022 | No Comments

Success in the mobile and consumer device market depends on a vibrant software ecosystem. Over the past decade the Android Open Source Project (AOSP) has become one of the top operating systems for phones, tablets and a variety of other mobile devices. Today we are happy to share that the upstream enablement of RISC-V has started within AOSP. In order to celebrate this truly global community effort, let us take a closer look at the work to date and plans to expand and accelerate continued development of AOSP on RISC-V.

In 2020, engineers and software developers from the Chinese Academy of Sciences' PLCT Lab began to port Android 10 to the RISC-V architecture in an effort to open this important ecosystem to the RISC-V community. Since the early days of the effort, the Alibaba Cloud division has been a close collaborator and leader in this pioneering work and has kept the development current with newer Android versions.

"We are glad to see more support from Google for building AOSP targeting RISC-V! Alibaba Cloud has been committed to supporting the RISC-V community through a series of innovations, such as progressing the porting of basic Android functions onto RISC-V, which proves the feasibility of using RISC-V based devices in scenarios ranging from multimedia to signal processing, device interconnection, and artificial intelligence. We look forward to engaging with the Android team to contribute to the thriving RISC-V community down the road," said Dr. David Chen, Director of Ecosystem from Alibaba Cloud and Vice Chair of the Applications & Tools Horizontal Committee at RISC-V International.

"RISC-V has grown in popularity through the sheer demand for flexibility and choice across the full spectrum of computing, from the smallest embedded devices to the largest scale out cloud implementations," said Calista Redmond, CEO of RISC-V International. "This demand has made RISC-V inevitable as the most prolific open standard ISA of our time, accelerating innovation and adoption with the strongest ecosystem of global stakeholders."

Agenda

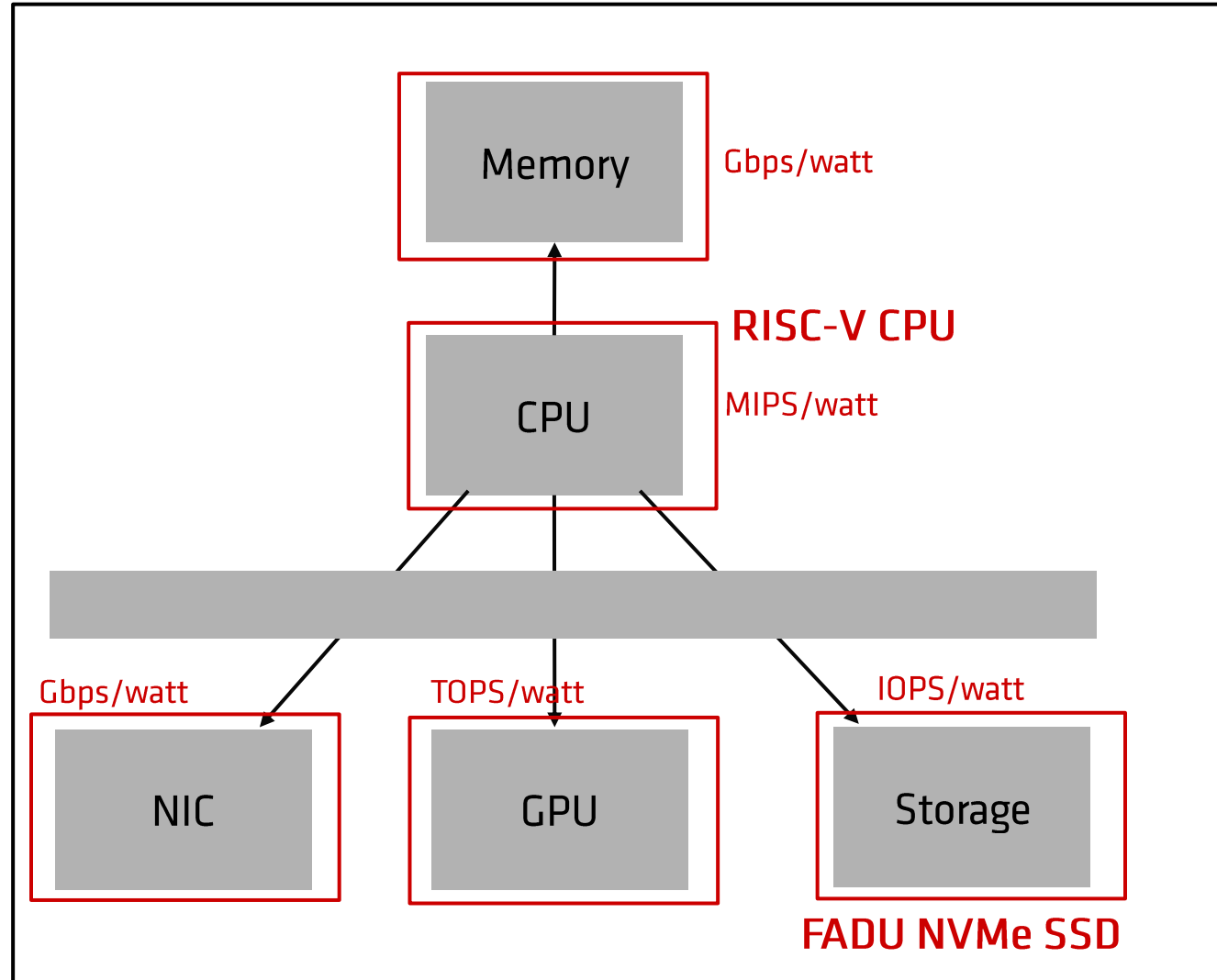
- FADU introduction and mega trend
- FADU NVMe SSD
- RISC-V CPU
- Next step

Optimization techniques for the computer system

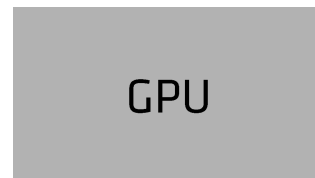
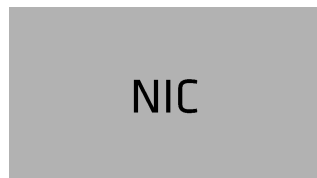
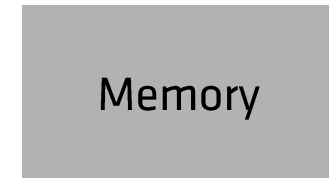
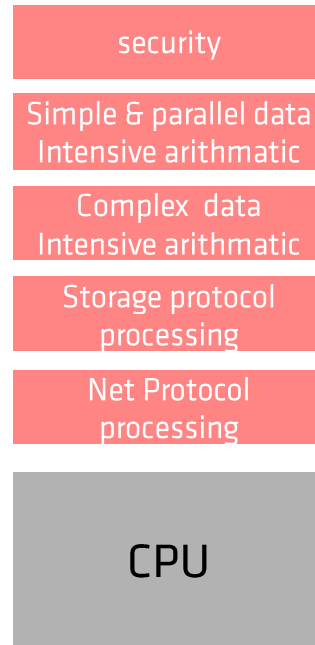
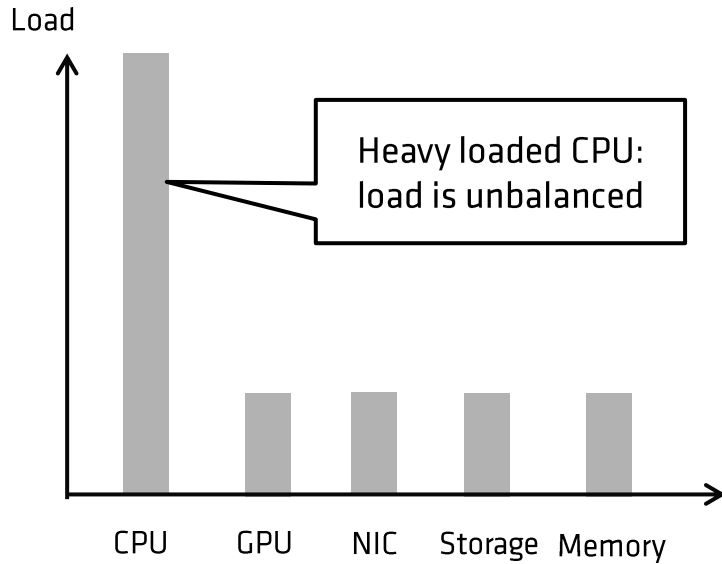
Constraints & Requirements

- Cost increase should be minimized
- Power and thermal have become essential constraints
- Performance should be as high as possible

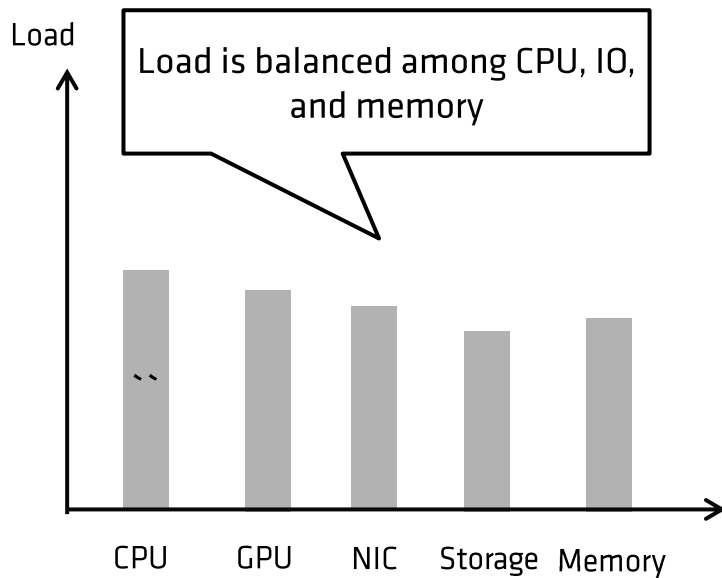
- Optimizing each component



Movitation: Application specific acceleration and offloading



Movitation: Application specific acceleration and offloading



No need to adopt unnecessarily powerful CPU for specific application or service

CPU

Simple & parallel data
Intensive arithmetic

security

PIM

Storage protocol
processing

Net Protocol
processing

security

DPU

Simple & parallel data
Intensive arithmetic

security

GPGPU/NPU

Complex data
Intensive arithmetic

security

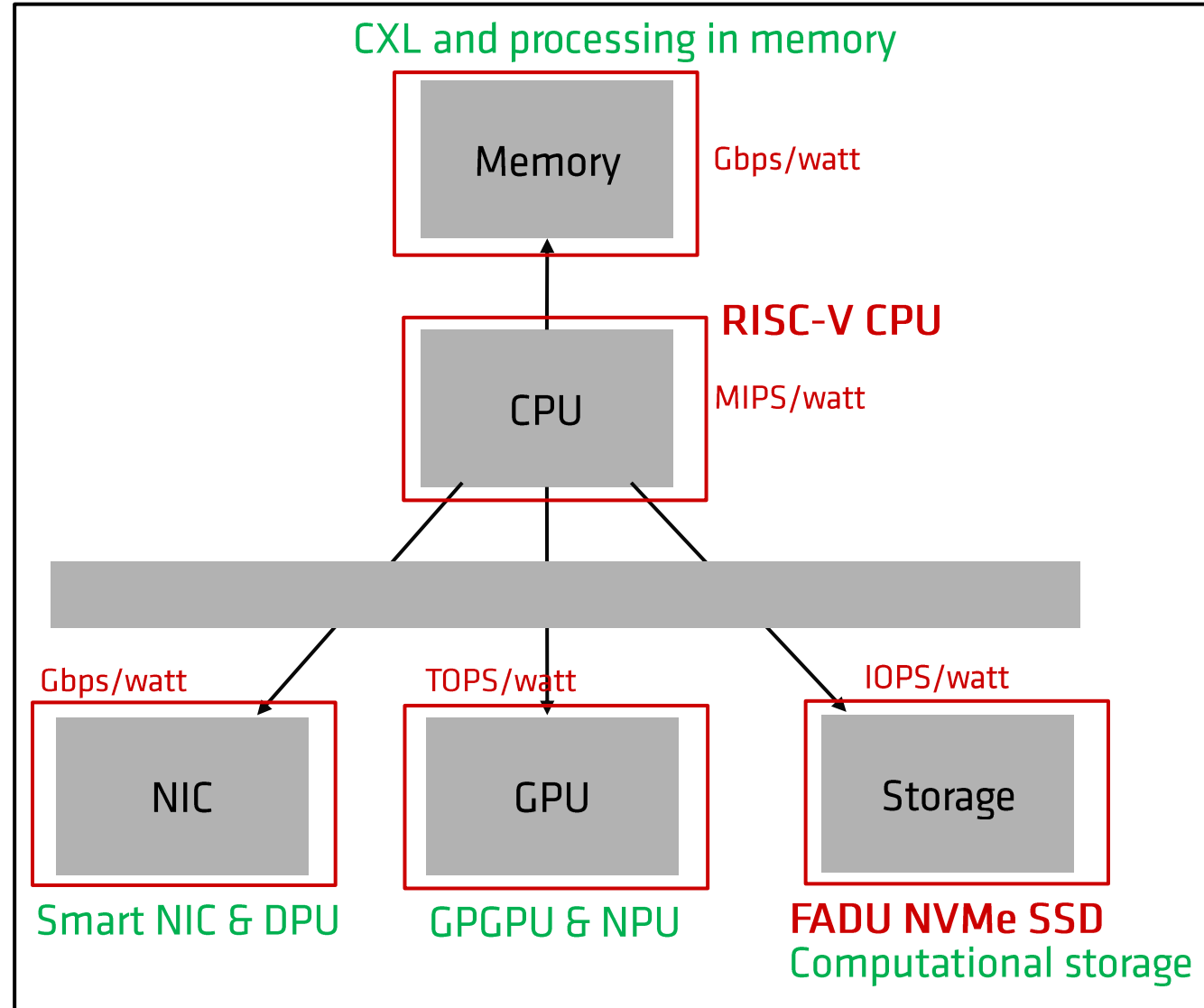
Computational
storage

Optimization techniques for the computer system

Constraints & Requirements

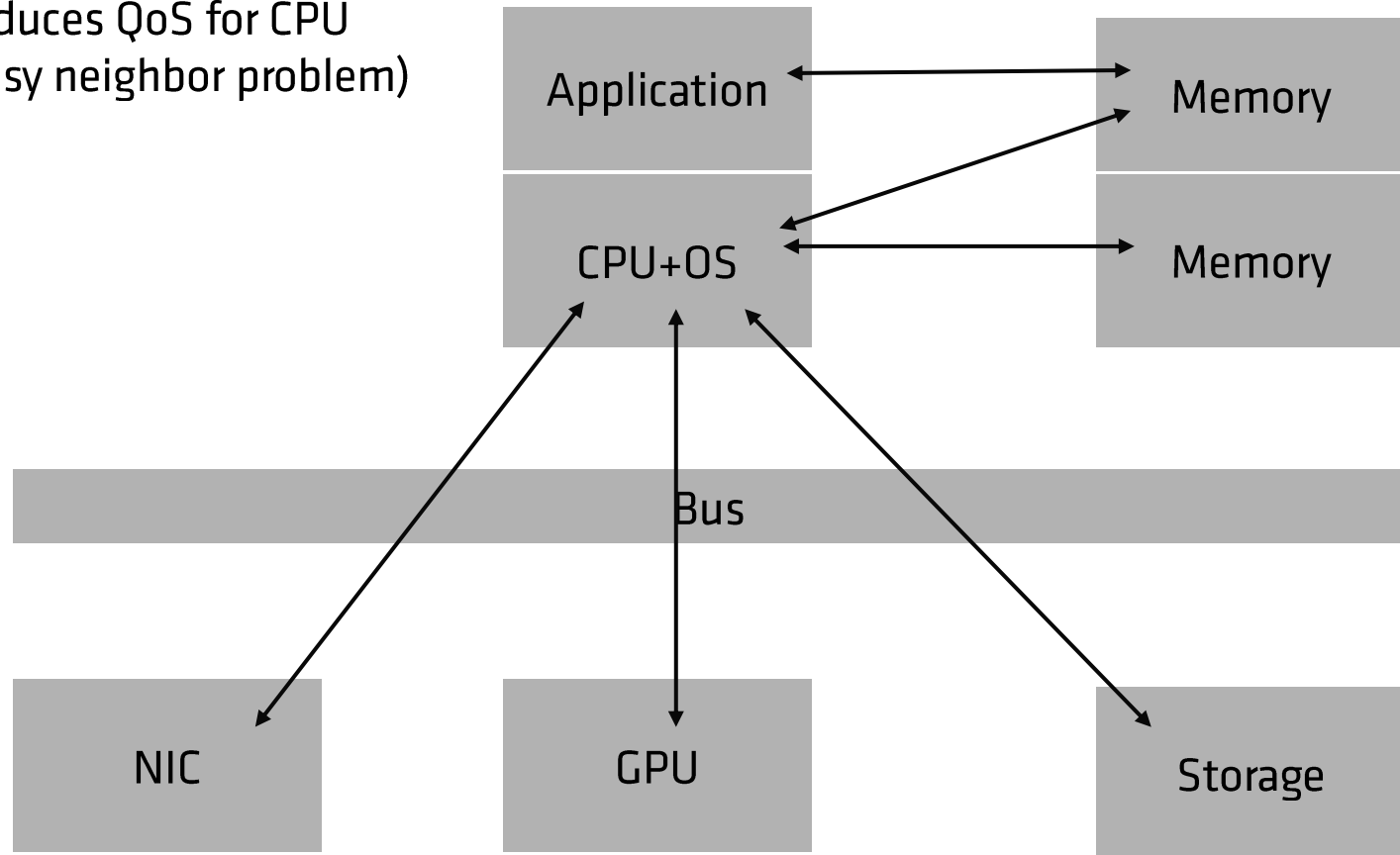
- Cost increase should be minimized
- Power and thermal have become essential constraints
- Performance should be as high as possible

- Optimizing each component
- Application specific acceleration and offloading

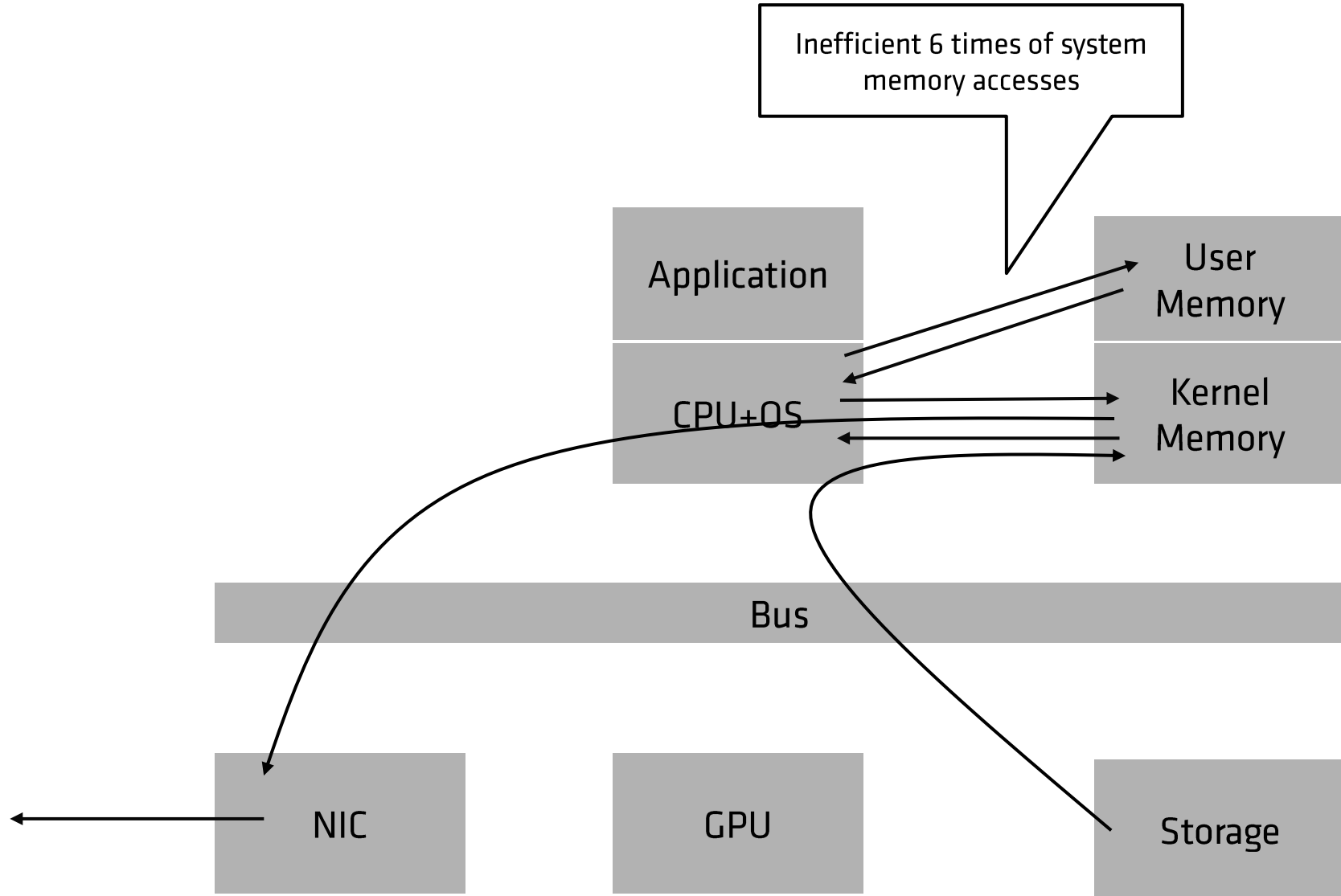


Problems of CPU coordinated data movement

- PCIE EPs are getting faster and faster (e.g., NVMe SSD, RDMA NIC, GPGPU)
- Bounce buffering all I/O through system memory is a waste of system resources and reduces QoS for CPU memory (the noisy neighbor problem)

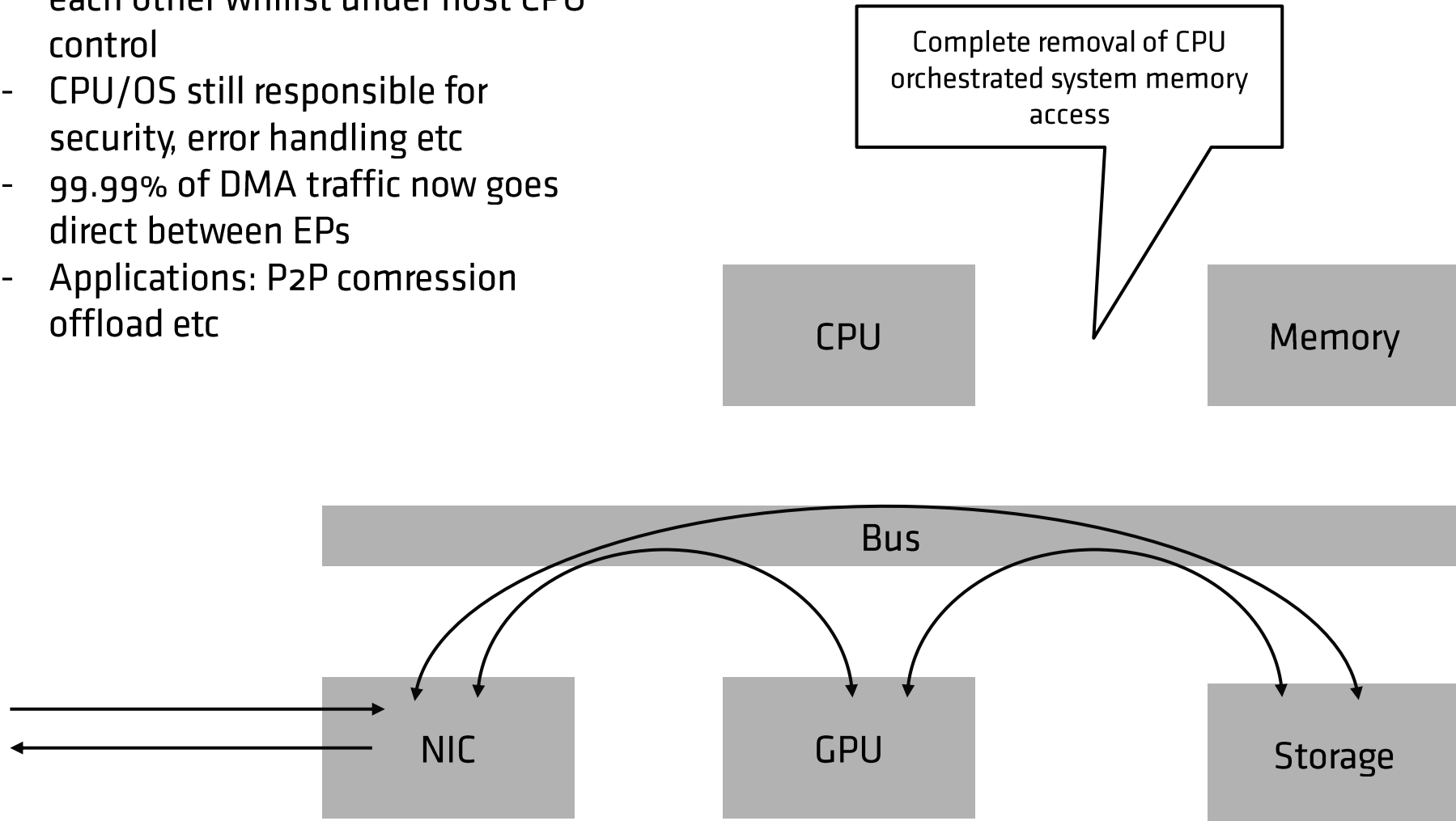


Problems of CPU coordinated data movement: Example



P2P communication

- P2P allows PCIe Eps to DMA to each other whilst under host CPU control
- CPU/OS still responsible for security, error handling etc
- 99.99% of DMA traffic now goes direct between EPs
- Applications: P2P compression offload etc

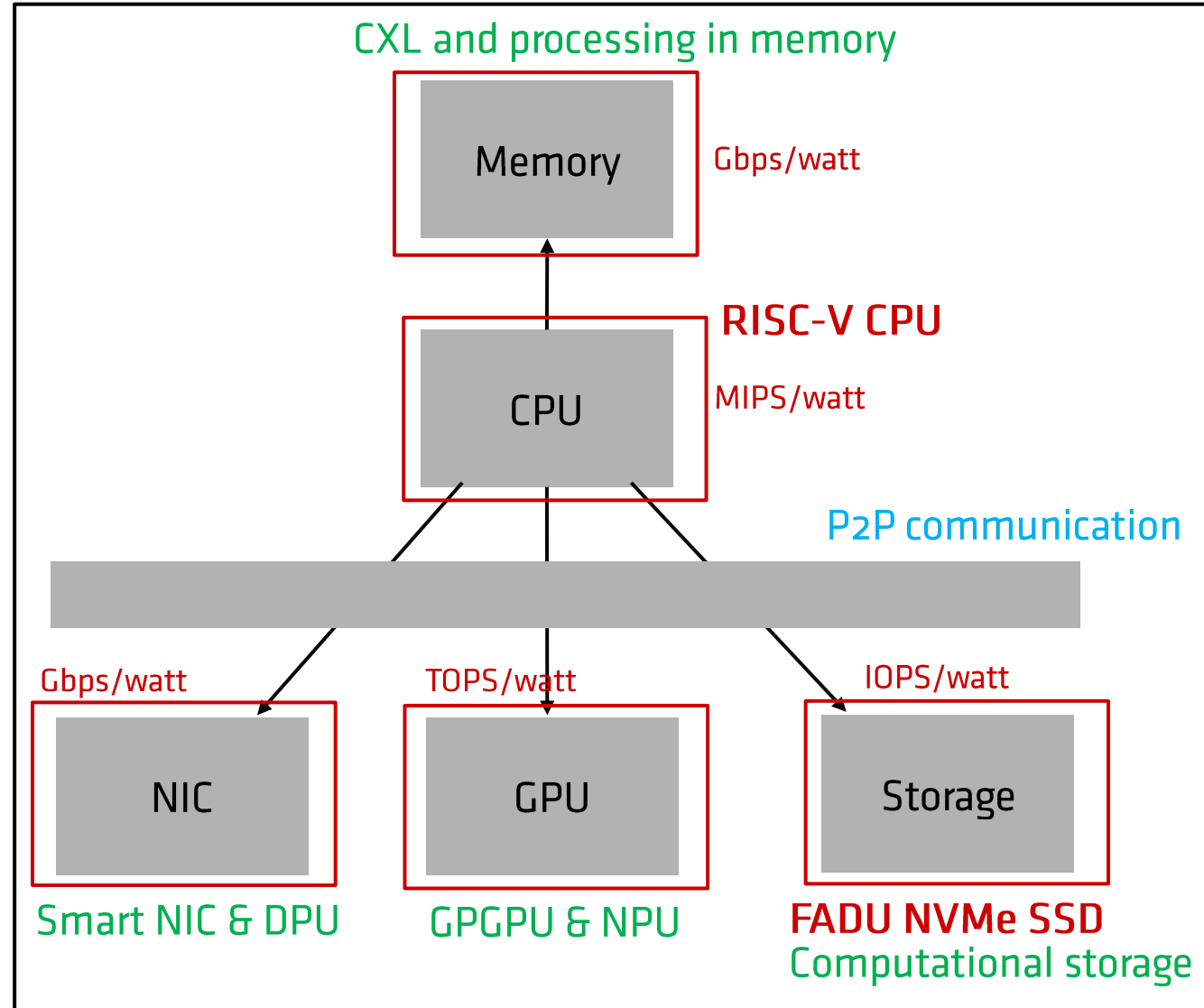


Optimization techniques for the computer system

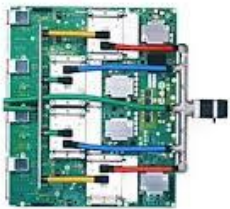
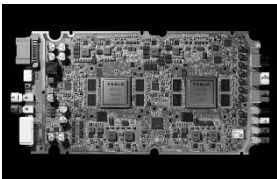
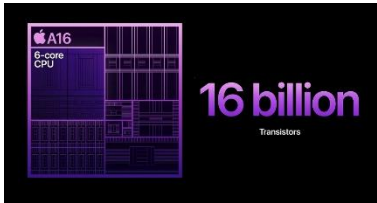
Constraints & Requirements

- Cost increase should be minimized
- Power and thermal have become essential constraints
- Performance should be as high as possible

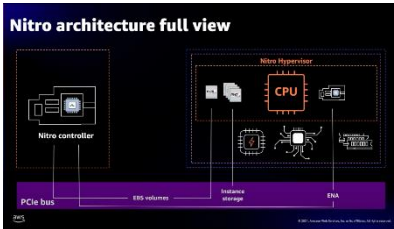
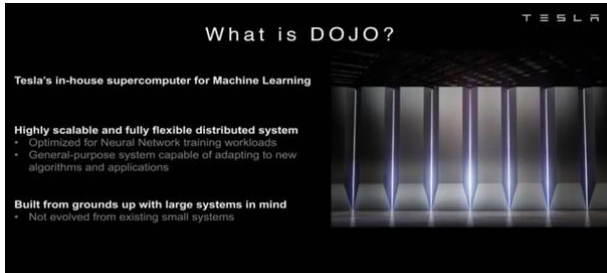
- Optimizing each component
- Application specific acceleration and offloading
- Efficient data movement



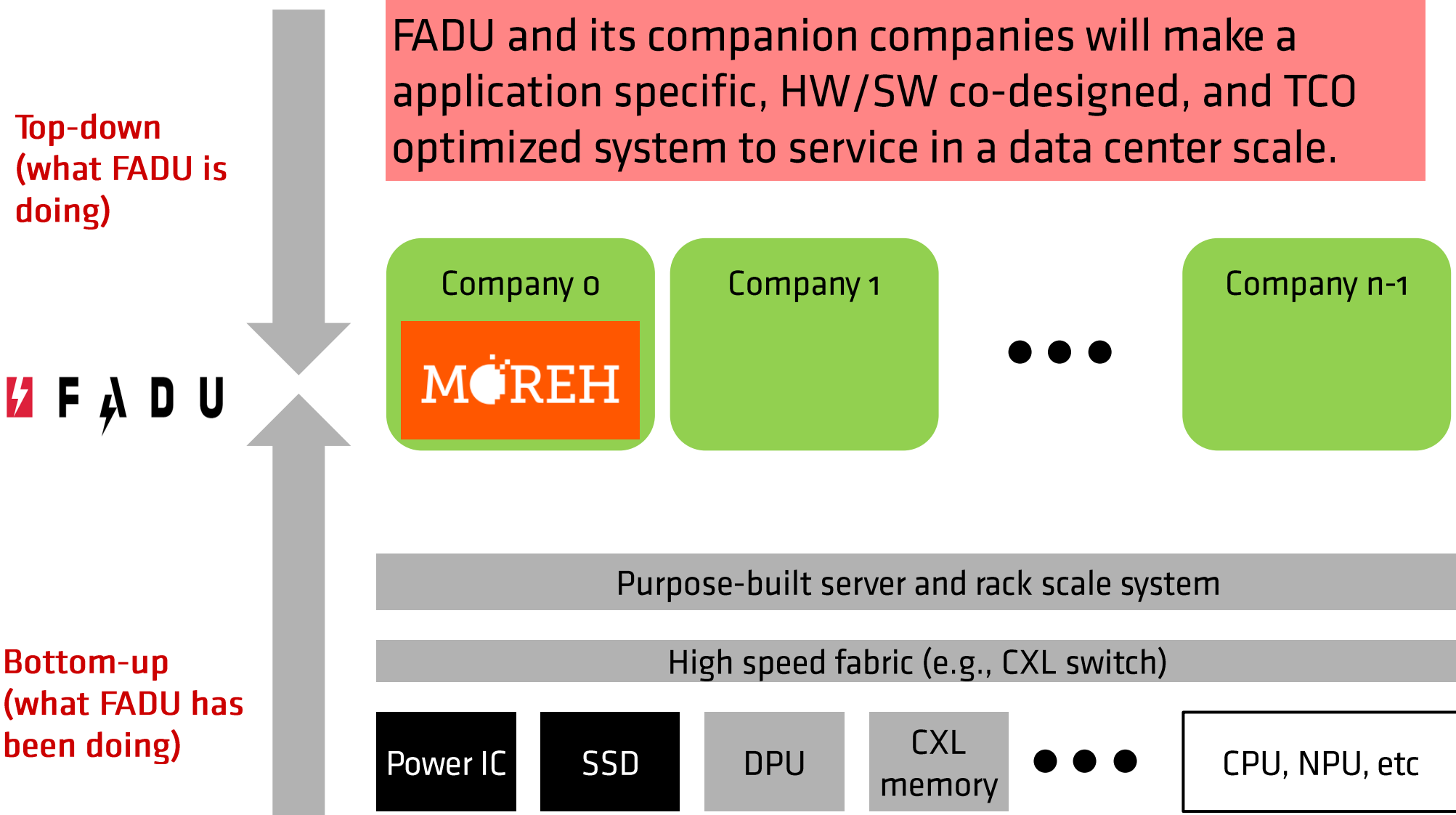
Global tech-companies have in common: **top-down** vertical optimization



TPU^{v4}



FADU's vision



Representative companion company

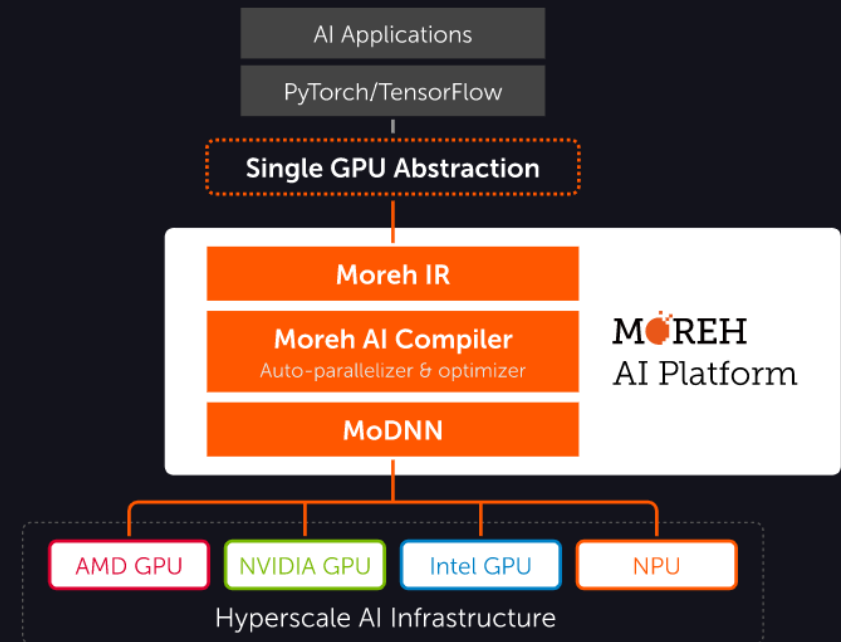
Technology ■

Moreh AI Platform

World's most compatible and flexible AI infrastructure software

Moreh AI Platform is a software stack for tomorrow's AI models and infrastructures. Moreh AI Platform enables users to accelerate AI applications seamlessly on a variety of processors and on a diverse system scales, from single chips to large clusters, with the help of our compiler technology. Customers can focus on important AI problems regardless of complicated hardware architecture and parallelization techniques. By eliminating proprietary components from our software stack, AI systems can be built in a cost-effective way without being tied to a specific hardware vendor. We provide full compatibility with standard deep learning frameworks including PyTorch and TensorFlow. The rapidly expanding AI software ecosystem can be powered by our platform without any code modification.

[CONTACT](#)



Ready to go with strong alliance



Design service



Infrastructure software



Memory provider

Secured customer



Accelerate
Data processing unit



Central processing unit



Server hardware



Memory provider



Concluding remark

- Engineering is money. Whether it's a paper or a product, it must sell (read) well.
- With the end of Moore's Law, the time has come when hardware becomes money.
- You need to find a blue ocean at the point in time when you go out into society.
- At a young age, supported by mental strength and physical strength, one must dig into one field to the end. You can expand it later
- Good companies often come from good research results. (Google, Sifive, FADU)
- There are many people who are good at using tools that someone has made, or who have a good understanding of someone else's book. I hope you to become a person who can make even the smallest things on his own.
- Graduate school will be one of the best place to exercise engineering discipline

 **F**  **A** **D** **U**