Innovations in Silicon and Software

Defining the Next Generation of Enterprise Technology

Dave Zacks
Distinguished Engineer

Peter Jones
Distinguished Engineer

Hardware Innovations in ~~Silicon~~ and Software

Defining the Next Generation of Enterprise Technology

CISCO Live !

Dave Zacks
Distinguished Engineer

Peter Jones
Distinguished Engineer

#HighBitRate

Cisco Webex App

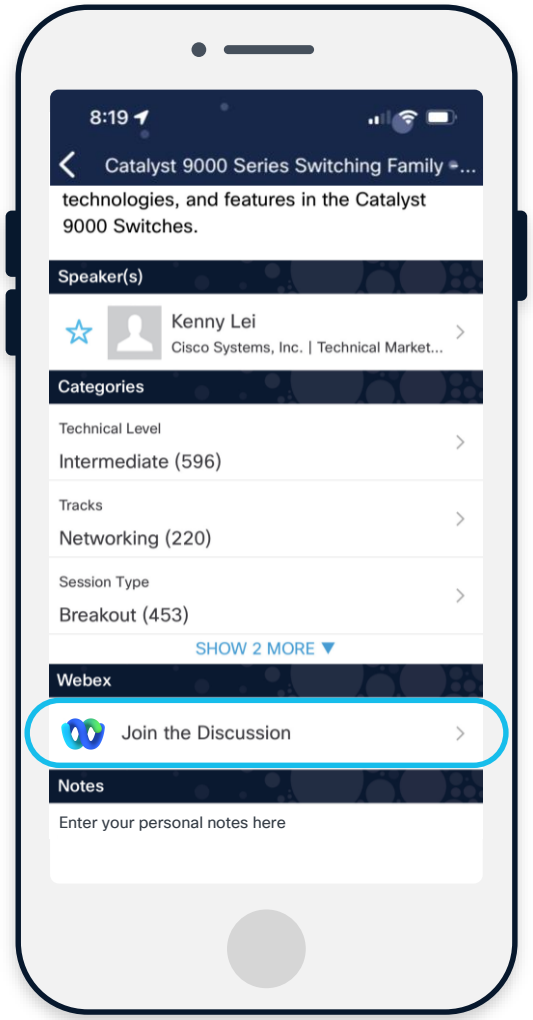
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 13, 2025.



By Way of Introduction ...

I am a **Distinguished Engineer** in the Cisco Security Innovations CTO team, and have been with Cisco for 25 years.

I work primarily with large, high-performance Enterprise network architectures, designs, and systems. I have over 30 years of experience with designing, implementing, and supporting solutions with many diverse network technologies.

I have a strong background in, and focus on, customer requirements, and integrating these into the products and solutions Cisco builds.
I have a special interest in **Flexible Hardware, Fabrics, Assurance** and **ML/AI**.

Dave Zacks
Distinguished Engineer

Email: dzacks@cisco.com
Bluesky: davezacks.bsky.social
LinkedIn: ln/dave-zacks-43677474/

By Way of Introduction ...

I am a **Distinguished Engineer** in the Cisco Networking Hardware team and have been with Cisco since 2005.

I work on system architecture and standards strategy across the portfolio. I was a key figure in the development of the UADP switching ASIC architecture and the Catalyst switches that use it.

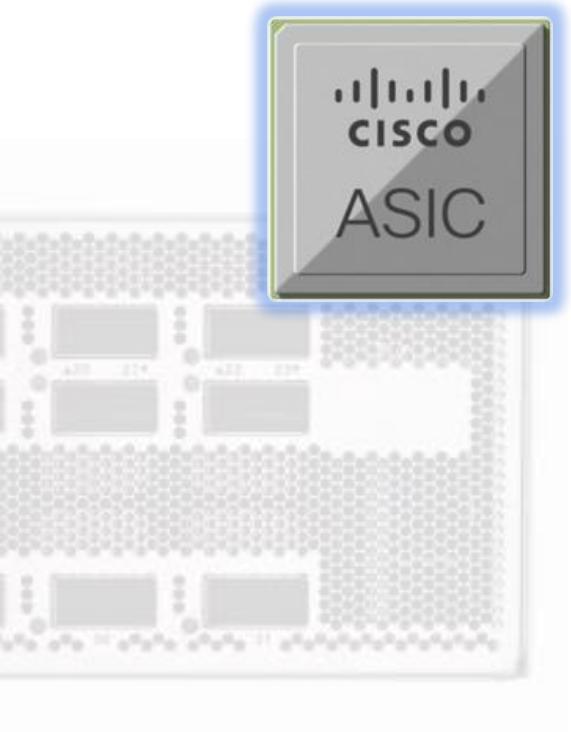
I work in defining and promoting new Ethernet standards in IEEE 802.3 and as Ethernet Alliance Chairman.

I am passionate about **Network Evolution, Adoptable Technology** and **Ethernet.**

Peter Jones
Distinguished Engineer

Email: petejone@cisco.com
Bluesky: petergjones.bsky.social
LinkedIn: in/petergjones/

Agenda



01 **Introduction**

02 **Creating Networking Silicon,**
Why and How

03 **Flexible Networking Silicon,**
Overview

04 **Campus Networking Silicon,**
Latest ASIC, Platform, and Infrastructure Developments

05 **Cisco Solutions,**
Latest Developments, ML/AI

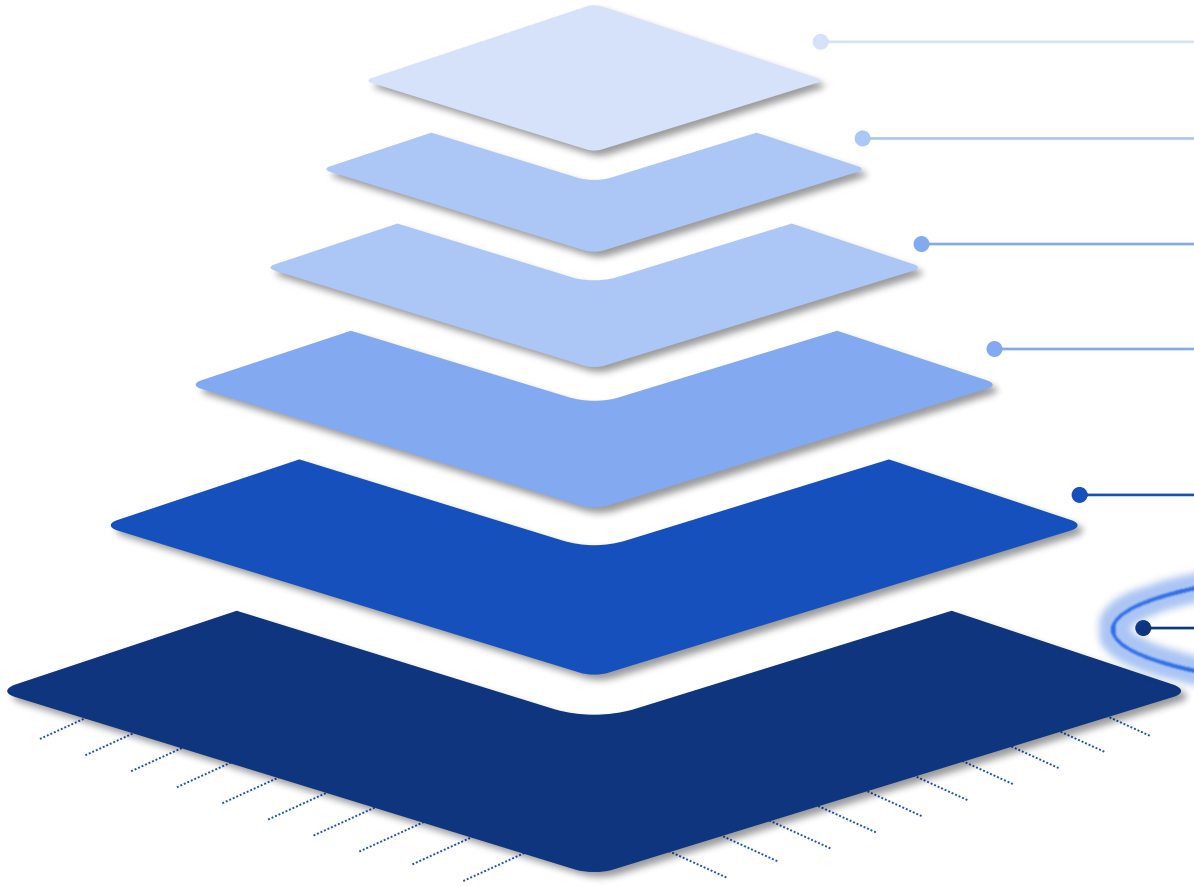
06 **Summary –**
Why Hardware Still Matters in a Software-Defined World

Introduction

Creating Networking Silicon, Why and How

The Technology Stack

Building on a Firm Foundation



Applications and Zero Trust

Highly Programable APIs

Cloud Management
On-Prem Management

Powered by model-driven
software Cisco IOS® XE

Physical and
virtual infrastructure

Cisco Application-Specific Integrated Circuit (ASIC)

Cisco Spaces

Cisco ISE

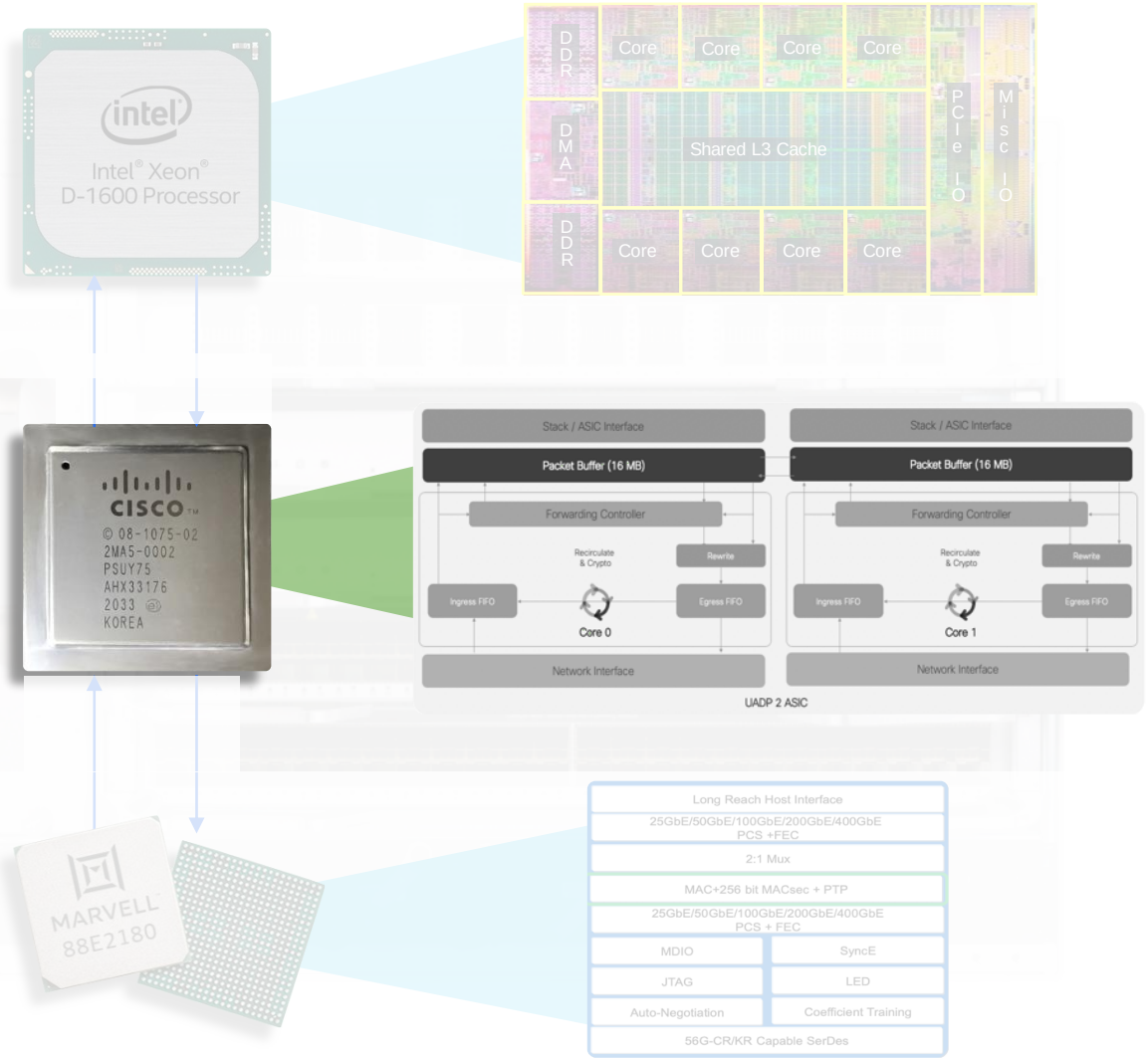
Cisco Catalyst Center

Meraki

Catalyst 9000 Family

Silicon One UADP

Software vs. Hardware



CPU/DRAM

Where the OS "software" runs. Includes control-plane, data-plane and system-management functions.

- **OS layer** – IOSXE (IOSd) and Features, etc.
- **System layer** – FMAN, CMAN, IOMD, FED, etc.

ASIC(s)

Where the "hardware" processing of traffic & services runs. Uses forwarding and state tables programmed by the software.

- **Forwarding** – L2, L3, ECMP, Encap, etc.
- **Services** – ACLs, QoS, Analytics, Encryption, etc.

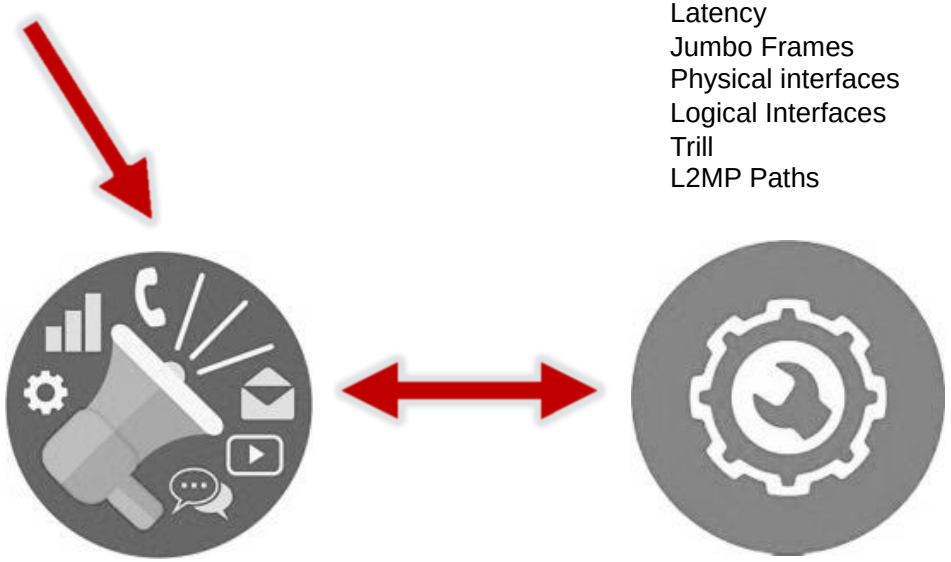
Stub/PHY(s)

Transforms electrical and optical signals, splits or combines signals, and other various "physical" layer functions, such as encryption and timestamping.

How is an ASIC built?

ASICs

From Definition to Deployment



Marketing

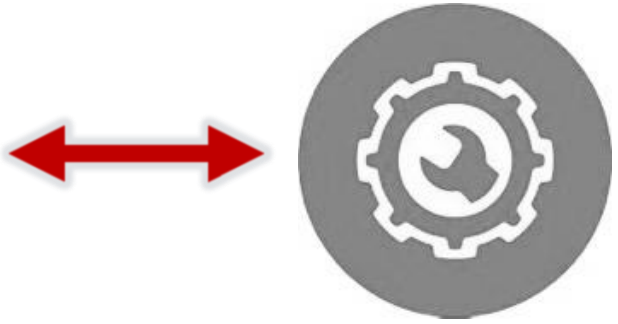
Engineering

FIB Table Size	ECMP Paths	LAG	WCCP	CoPP Policing
MAC Table Size	IGMP v1/2/3	RACL	CAPWAP	L2 CoPP
IPv6 unicast/multicast	PIM Snooping	PACL	EEE	L3 CoPP
IPv4 unicast/multicast	Netflow	QoS	802.1Qbg	FC/FCoE
Adjacency Table	Flow Masks	PBR	802.1Qbh	1588 PTP
MPLS P/PE/TE	EoMPLS	RBACL	802.1Qaz	802.1ae
Latency	VPLS	VACL	Adjustable MTU	Port Security
Jumbo Frames	GRE	VACL Capture	Byte Statistics	SGT
Physical interfaces	MPLSoGRE	GOLD	Packet Statistics	AES Encryption
Logical Interfaces	MVPN	ELAM	Broadcast Suppression	VOQ
Trill	SPAN	VRF Lite	Multicast Suppression	QPPB
L2MP Paths	ERSPAN	Policing	Unicast Suppression	PFC

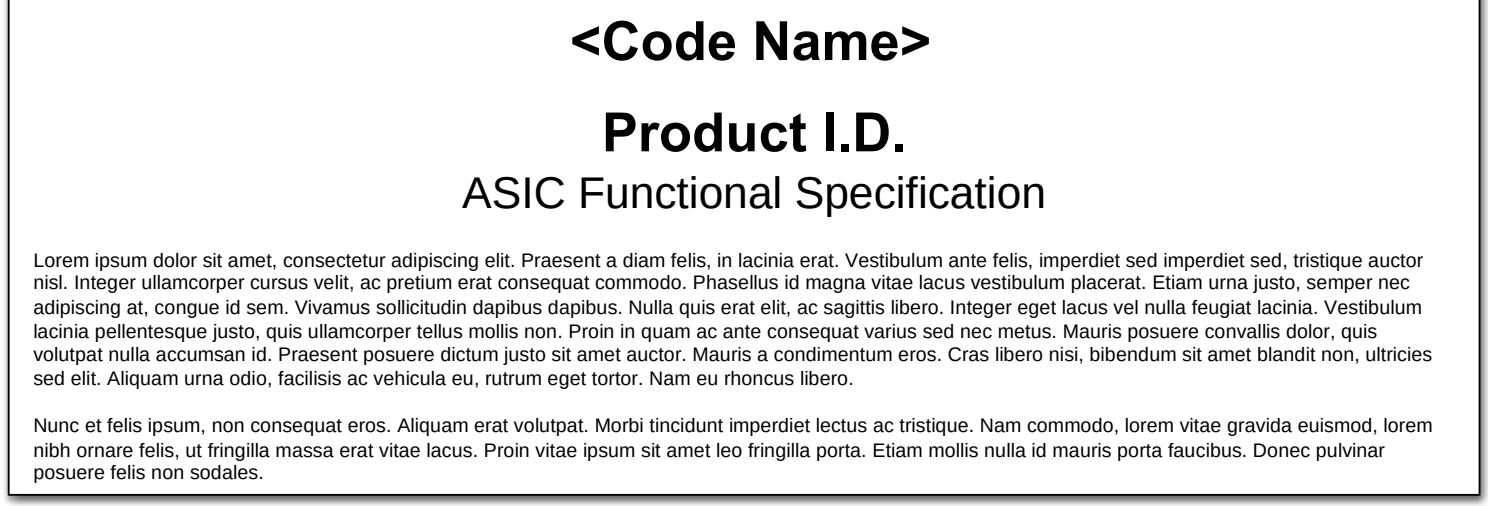
and more…

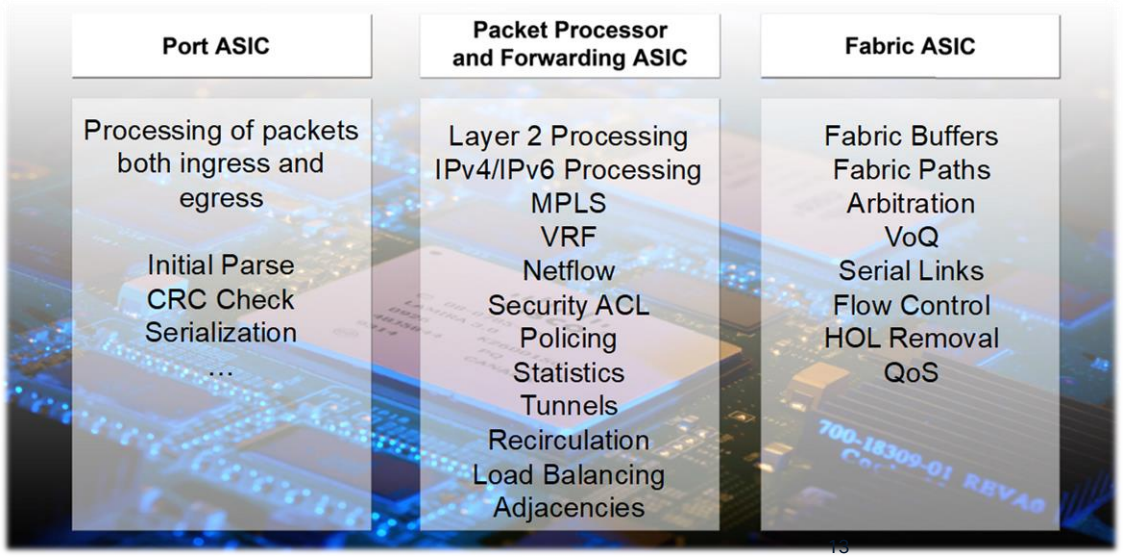
ASICs

From Definition to Deployment



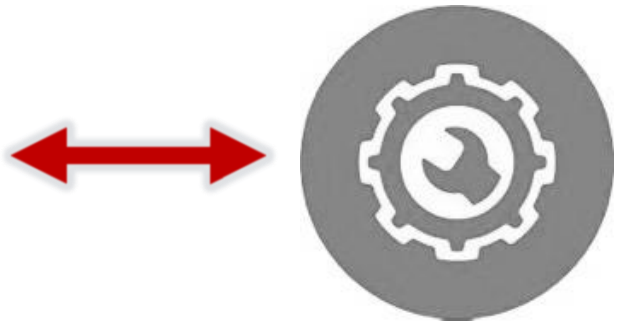
Marketing **Engineering**





ASICs

From Definition to Deployment



Marketing **Engineering**

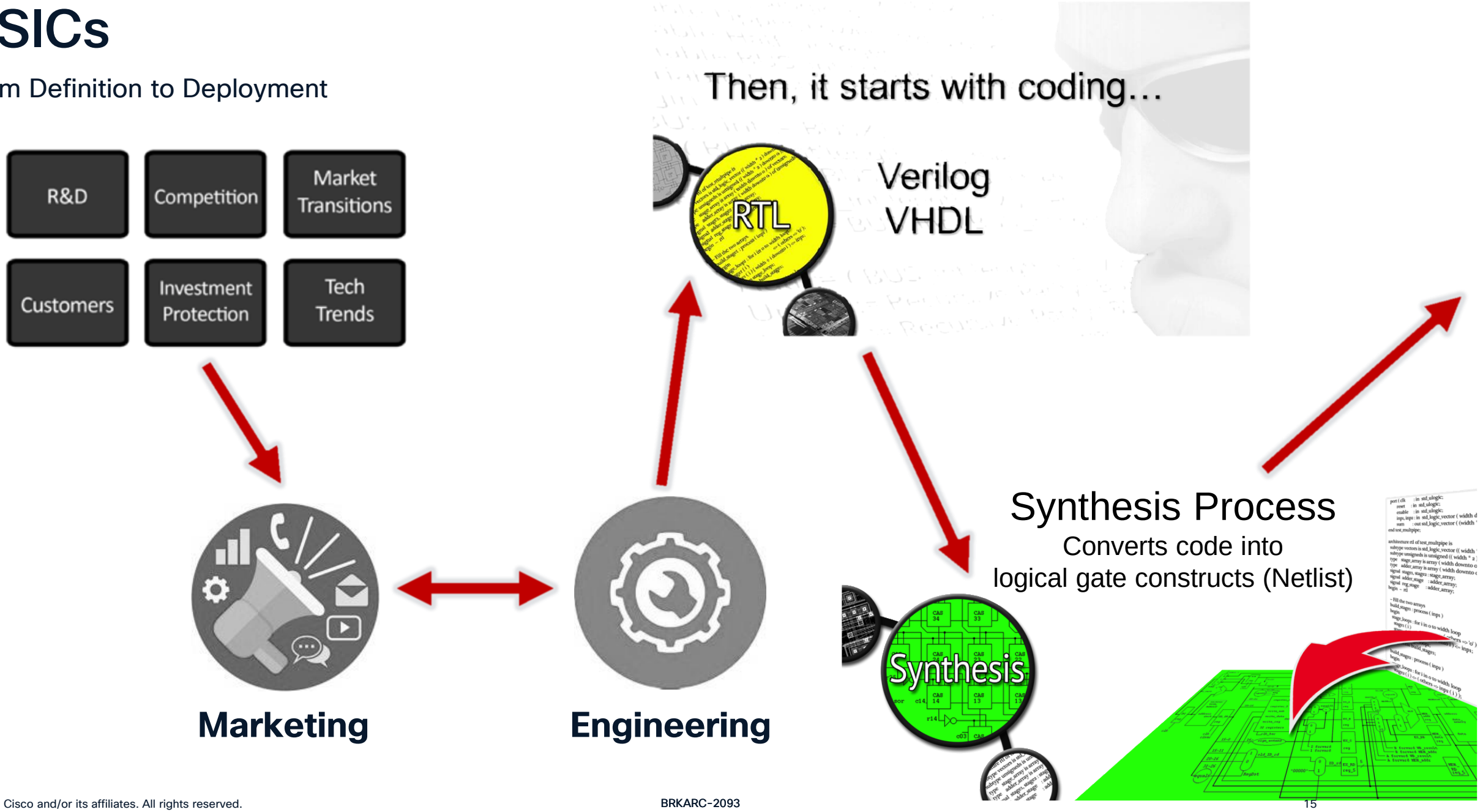
Then, it starts with coding…

Verilog
VHDL

```
library IEEE;
use IEEE.std_logic_1164.all;
use IEEE.numeric_std.all;    -- for the unsigned type

entity COUNTER is
  generic (
    WIDTH : in natural := 32);
  port (
    RST   : in std_logic;
    CLK   : in std_logic;
    LOAD  : in std_logic;
    DATA  : in std_logic_vector(WIDTH-1 downto 0);
    Q     : out std_logic_vector(WIDTH-1 downto 0));
end entity COUNTER;

architecture RTL of COUNTER is
  signal CNT : unsigned(WIDTH-1 downto 0);
begin
  process(RST, CLK) is
  begin
    if RST = '1' then
      CNT <= (others => '0');
    elsif rising_edge(CLK) then
      if LOAD = '1' then
        CNT <= unsigned(DATA); -- type is converted to
unsigned
      else
        CNT <= CNT + 1;
      end if;
    end if;
  end process;

  Q <= std_logic_vector(CNT); -- type is converted back to
std_logic_vector
end architecture RTL;
```

ASICs

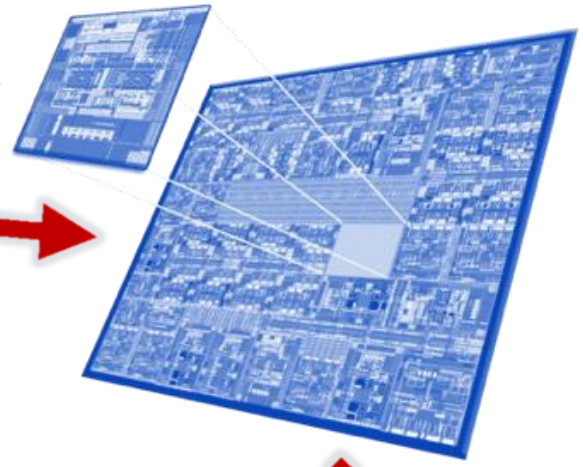
From Definition to Deployment



ASICs

From Definition to Deployment

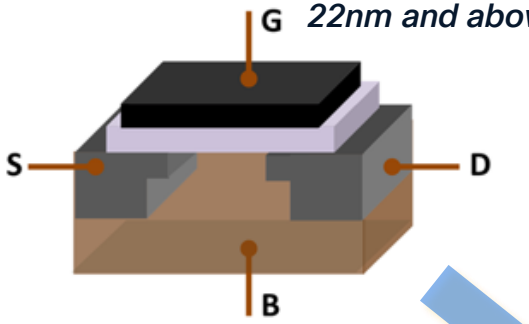
Arrange and interconnect constructs, connect power, minimize crosstalk, etc...



Floor Planning and Placement

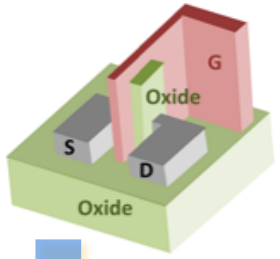
Discrete transistor

... mostly used @ 22nm and above



MOSFET

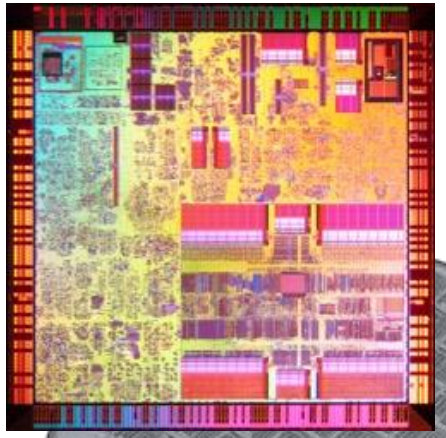
(metal oxide semiconductor field effect transistor)



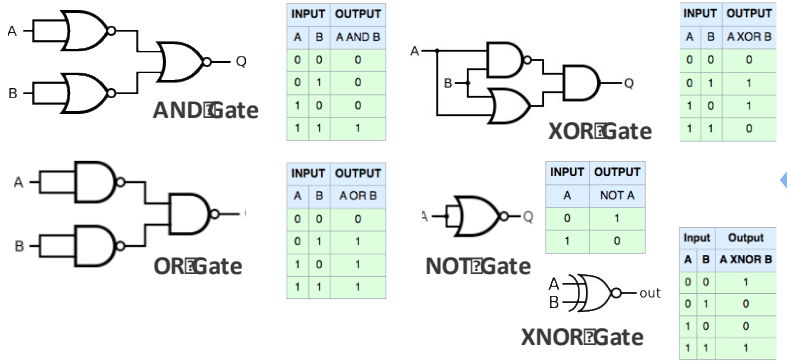
... mostly used @ 16nm and below

FinFET

(fin field effect transistor)



... which, when we put millions of them together on a silicon die, produce a chip!



INPUT		OUTPUT
A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

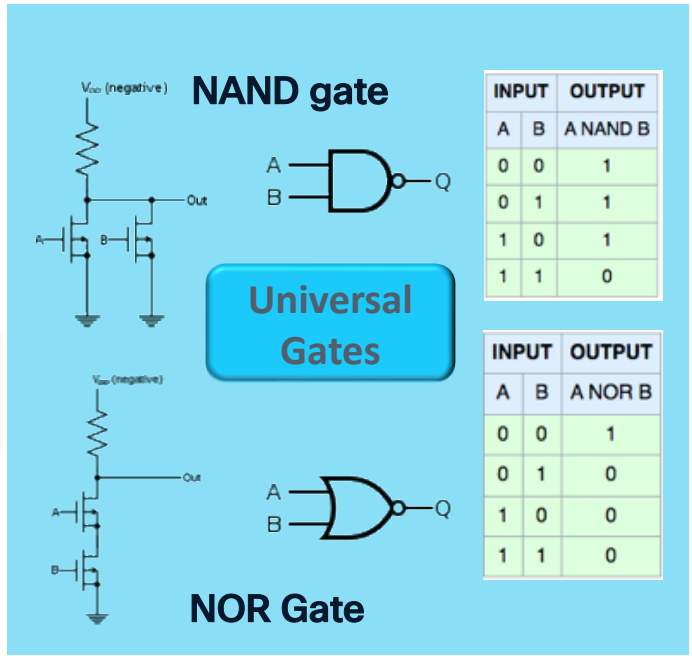
INPUT		OUTPUT
A	B	A OR B
0	0	0
0	1	1
1	0	1
1	1	1

INPUT	OUTPUT
A	NOT A
0	1
1	0

INPUT		OUTPUT
A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Input		Output
A	B	A XNOR B
0	0	1
0	1	0
1	0	0
1	1	1

... which can be used to build any of the other logic gates ...



INPUT		OUTPUT
A	B	A NAND B
0	0	1
0	1	1
1	0	1
1	1	0

INPUT		OUTPUT
A	B	A NOR B
0	0	1
0	1	0
1	0	0
1	1	0

ASICs

Benefits of the Ever-Shrinking Transistor

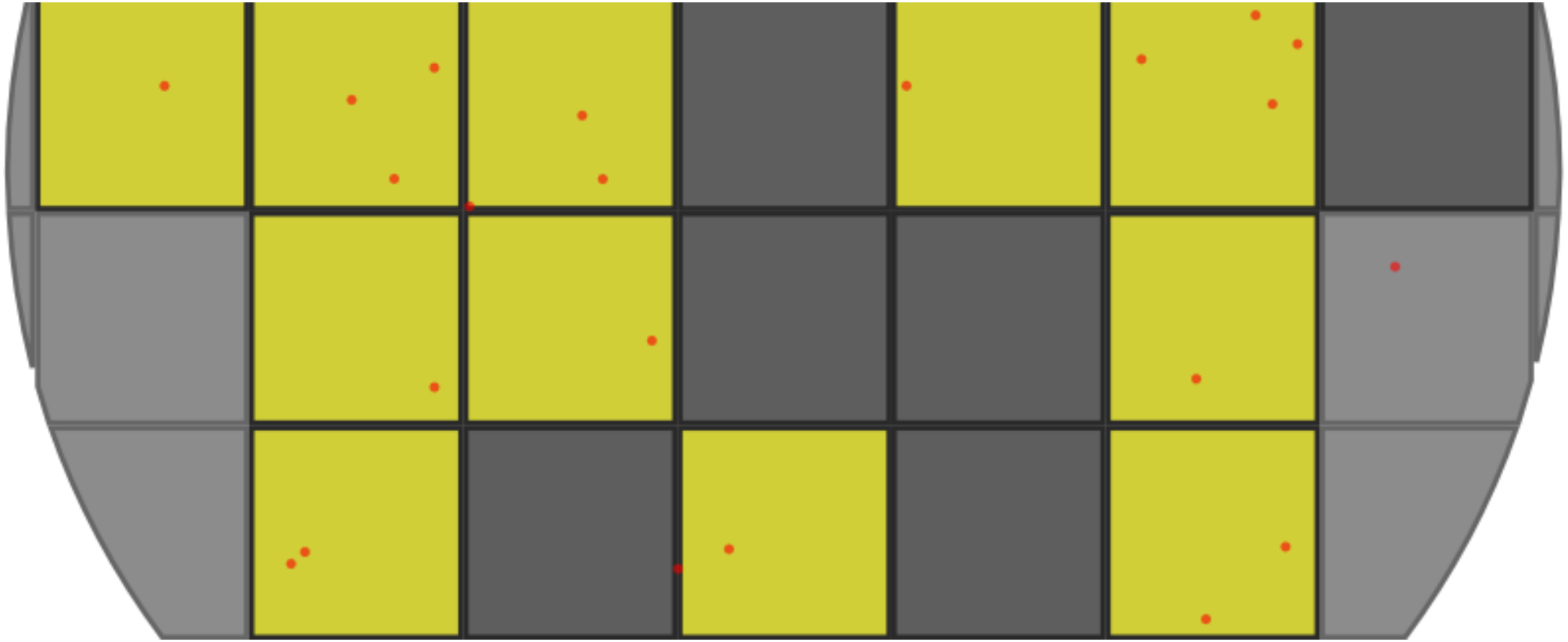
LOWER POWER CONSUMPTION

HIGHER PERFORMANCE

LOWER COST

ASICs

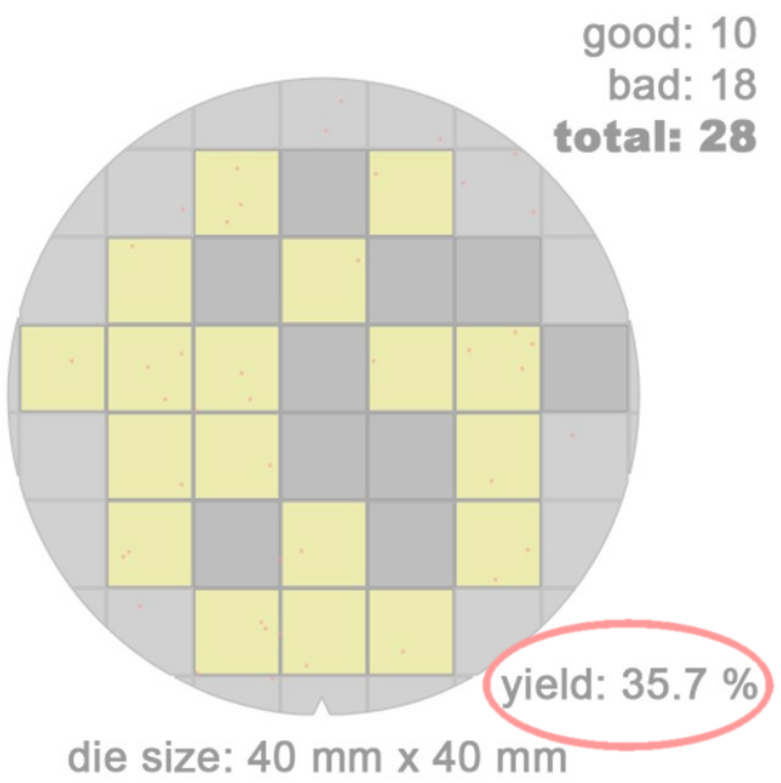
From Definition to Deployment

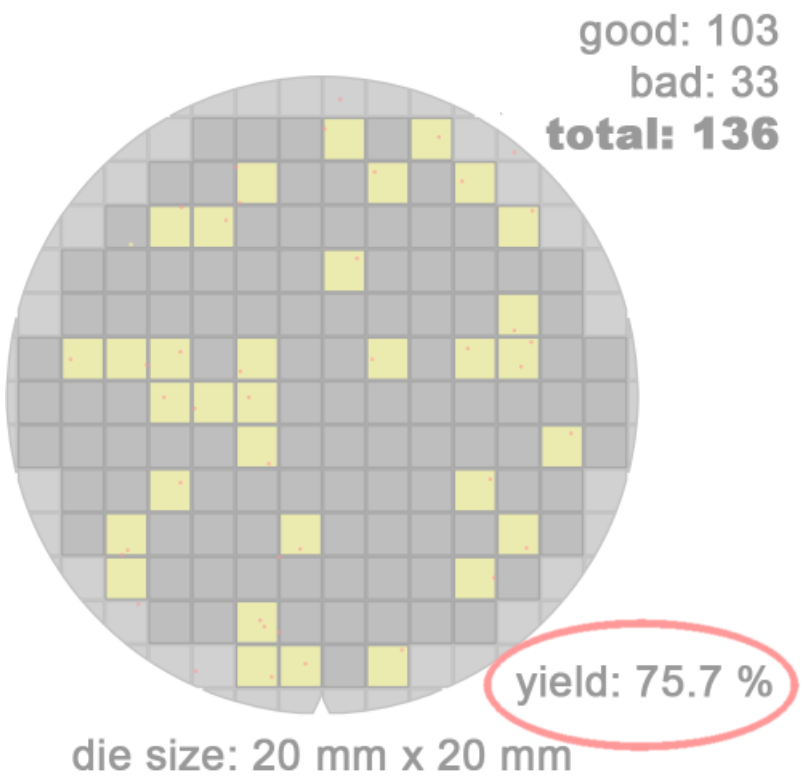


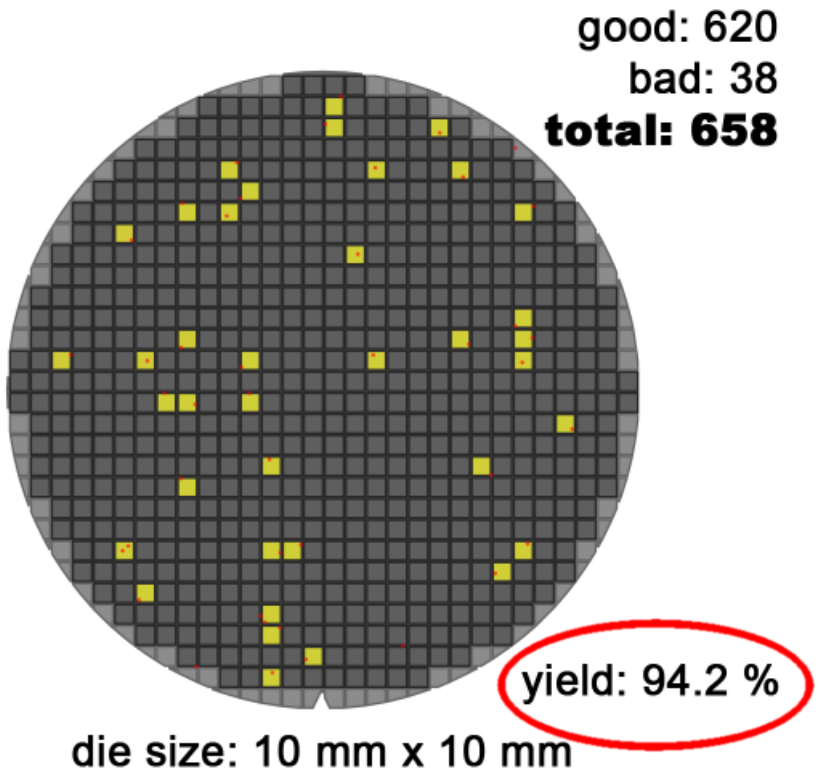
Every silicon die contains **imperfections** (impurities, micro-scratches, etc)
... these affect the **manufacturing yield** and thus the **cost** ...

ASICs

From Definition to Deployment

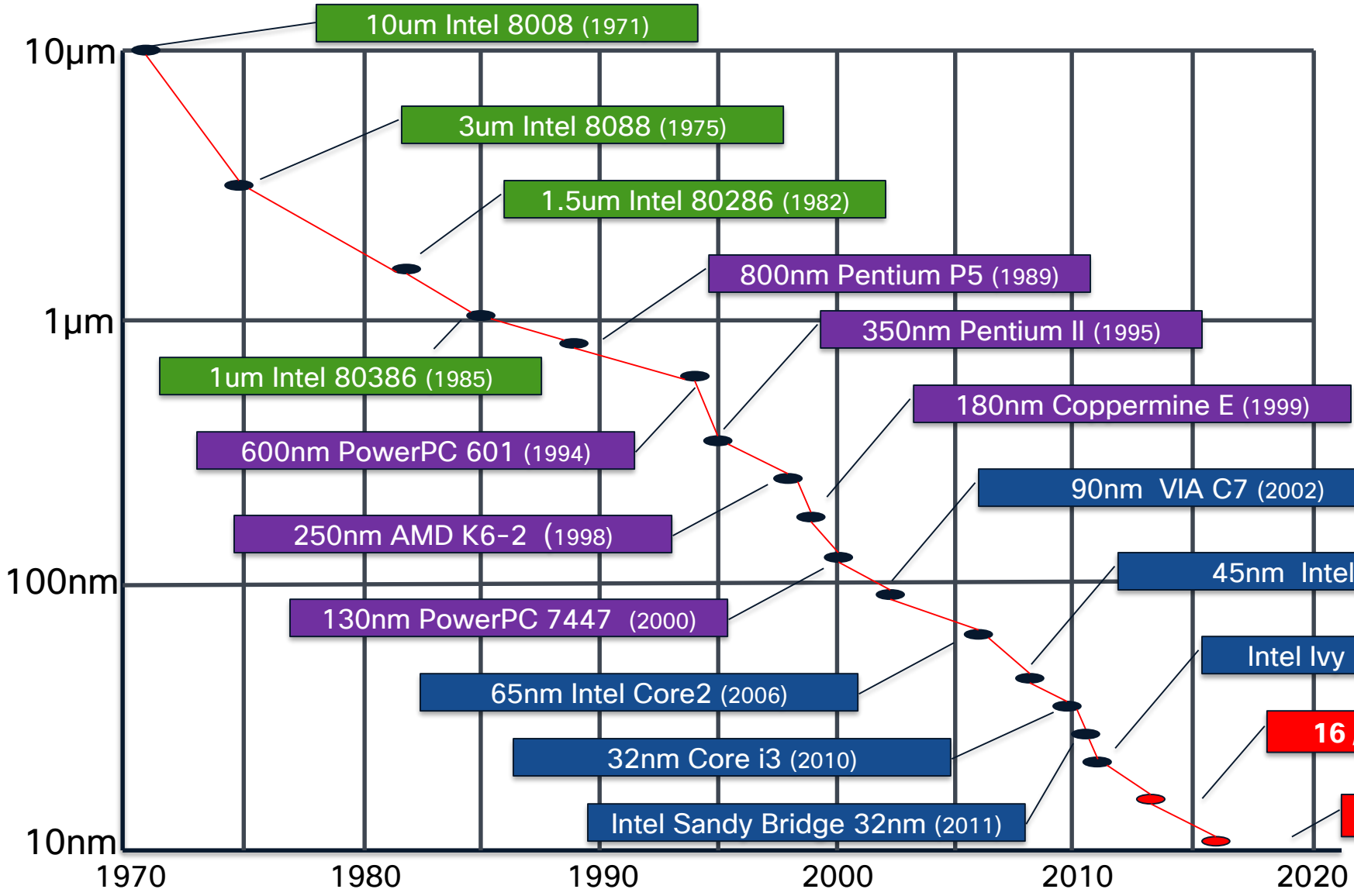




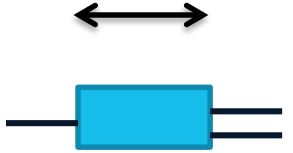


Source: http://commons.wikimedia.org/wiki/File:Wafer_die%27s_yield_model_%2810-20-40mm%29_-_Version_2_-_EN.png

Silicon Density Progression – Over Time



Transistor Width



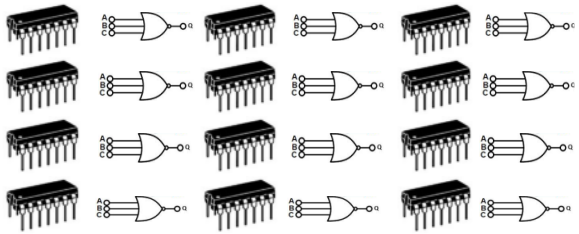
µm/nm is often a reference to width of the transistor

And even smaller today ... (more later)

Fun Fact!

Apollo Guidance Computer

... which was built from nothing but **that** ...



4100 ICs, each of which contained a **single 3-input NOR gate**

We put a human here ...

... using this ...

In other words ...
we put a man on the moon with less than 10,000 transistors ...
It takes 19.2 billion transistors to route your packets!

With the appropriate security, segmentation, QoS, encryption, fragmentation, etc, etc ...

More Fun Facts!

Source: https://forrestheller.com/Apollo-11-Computer-vs-USB-C-chargers.html#:~:text=The%20Anker%20PowerPort%20Atom%20PD%202%20USB-C%20Wall%20Charger,MHz%20Apollo%2011%20Guidance%20Computer).

Apollo Guidance Computer

has less performance than an Anker USB-C charger ...

Flies to the Moon and back

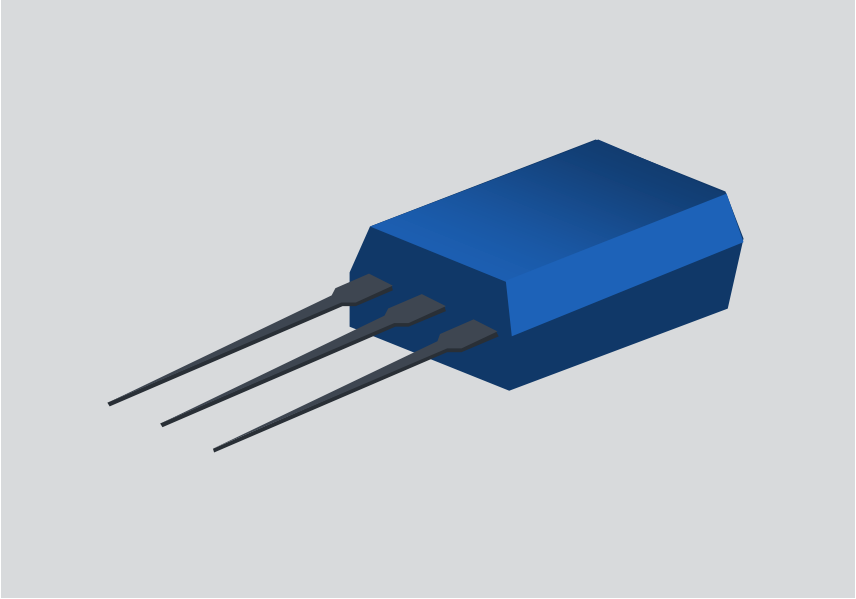
Based on: Discrete components
Clock speed: 1.024Mhz
RAM: 2048 15-bit words
Storage: 36,864 15-bit words

Charges 2 devices

Based on: Cypress CYPD4225
Clock speed: 48Mhz
RAM: 8KB
Storage: 128KB Flash

563x Faster!

We are talking about transistors sized in Nanometers



A Perspective

A **human hair** is

~100,000

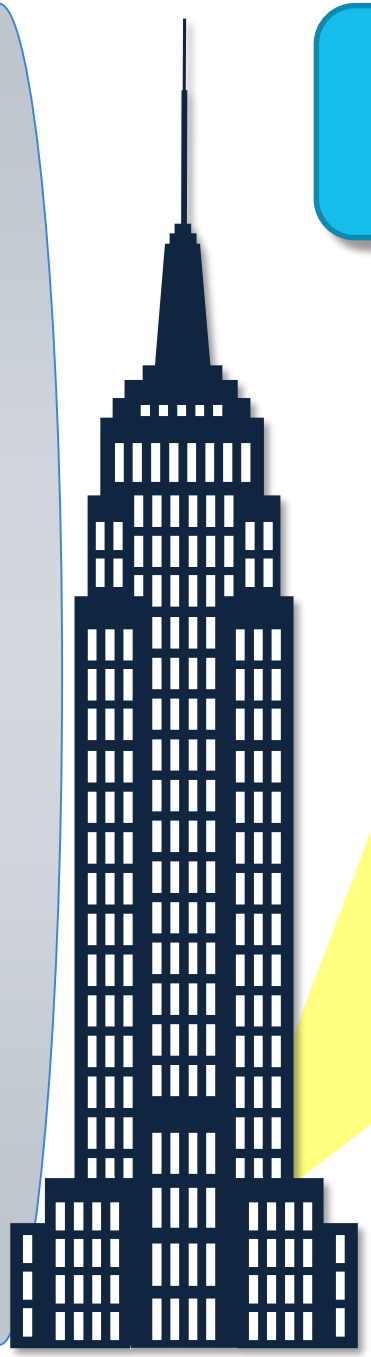
nanometers

in width

How SMALL is SMALL?

Single human hair

~ 100,000 nm



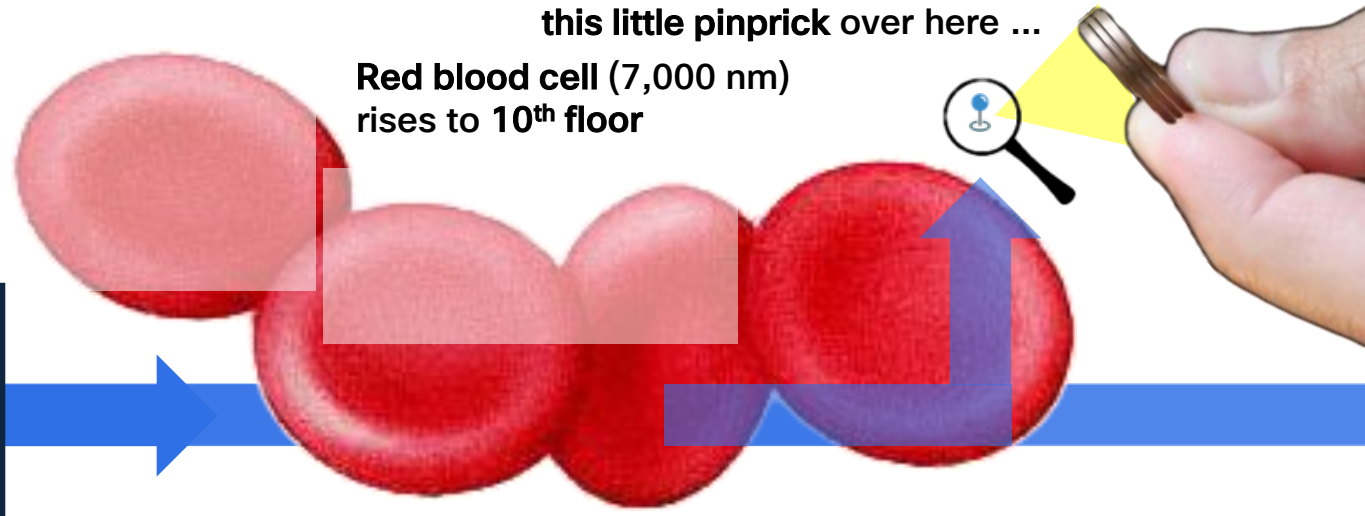
Empire State Building = 1454 feet to tip = **443 meters**

ONE NANOMETER – less than 1/4th of an inch!

... about the same thickness as three pennies on this scale ...

... and then we come to **this little pinprick** over here ...

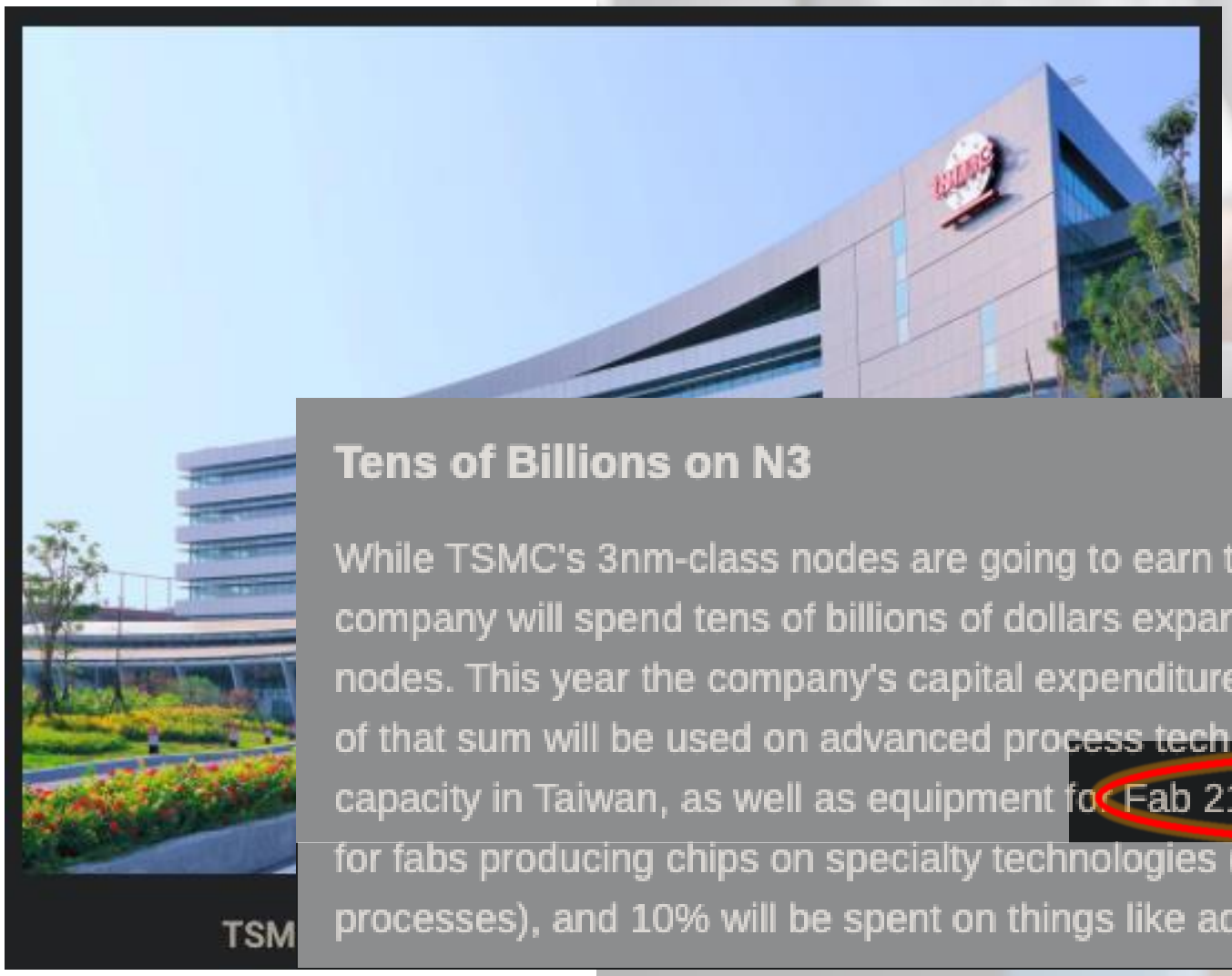
Red blood cell (7,000 nm) rises to **10th floor**



*... and we **build transistors** measured in **nanometers** ...*

The Current State of the Art

Ever Smaller ...



Advertised PPA Improvements of New Process Technologies
Data announced during conference calls, events, press briefings and press releases

	TSMC	
	N3 vs N5	N3E vs N5
Power	-25-30%	-34%
Performance	+10-15%	+18%
Logic Area	0.58x	0.625x
Reduction* %	-42%	-37.5%
Logic Density*	1.7x	1.6x
SRAM Cell Size	0.0199µm² (-5% vs N5)	0.021µm² (same as N5)
Volume Manufacturing	Late 2022	H2 2023

Tens of Billions on N3

While TSMC's 3nm-class nodes are going to earn the company a little more than $4 billion in 2023, the company will spend tens of billions of dollars expanding its fab capacity to produce chips on various N3 nodes. This year the company's capital expenditures are guided to be between $32 billion - $36 billion. 70% of that sum will be used on advanced process technologies (N7 and below), which includes N3-capable capacity in Taiwan, as well as equipment for Fab 21 in Arizona (N4, N5 nodes). Meanwhile 20% will be used for fabs producing chips on specialty technologies (which essentially means a variety of 28nm-class processes), and 10% will be spent on things like advanced packaging and mask production.

The Next Step

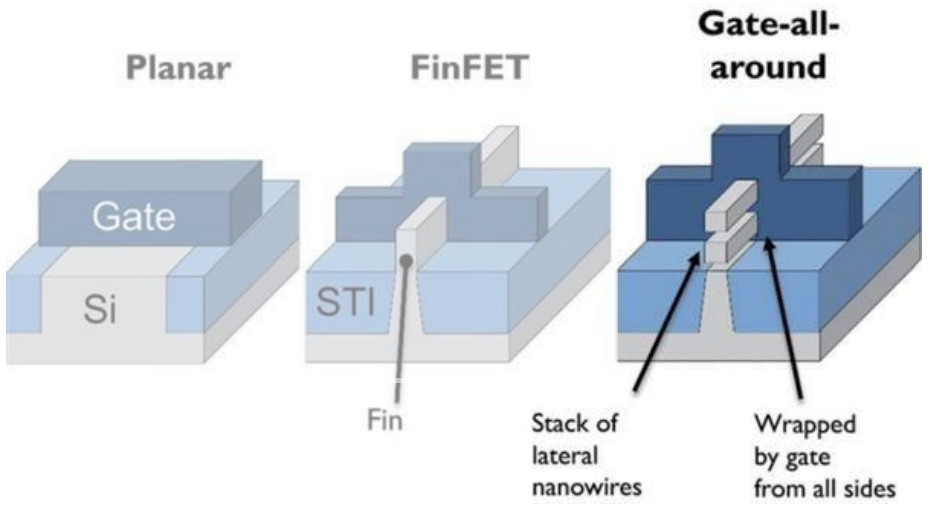
Ever Smaller …

TSMC Reveals 2nm Node: 30% More Performance by 2025

By Anton Shilov published June 16, 2022

TSMC details nanosheet-based GAAFET N2 process node.

Taiwan Semiconductor Manufacturing Co. today offici
class) manufacturing technology, its first node that wil
effect transistors (GAAFETs), at its 2022 TSMC Technol
fabrication process will offer a full-now performance a
when it comes to transistor density, it will barely impr
online.



	N2 vs N3E	N3E vs N5	N3 vs N5
Speed Improvement @ Same Power	10% ~ 15%	+18%	+10% ~ 15%
Power Reduction @ Same Speed	-23% ~ -30%	-34%	-25% ~ -30%
Chip Density	>1.1X	1.3X	?
HVM Start	H2 2025	Q2/Q3 2023	H2 2022

Sources:

https://www.tomshardware.com/news/tsmc-reveals-2nm-fabrication-process

https://www.ednasia.com/all-you-need-to-know-about-gaa-chip-manufacturing-process/

The Next Step

Ever Smaller ...

TSMC 2nm Update: Two Fabs in Construction, One Awaiting Government Approval

by Anton Shilov on January 19, 2024 11:15 AM EST

Anticipated to enter high-volume manufacturing in the **second half of 2025** (Baoshan site), and 2026 (Kaosiung site), using **Gate-All-Around (GAA) nanosheets**.

Second-generation 2nm process, using backside power delivery, **anticipated for 2026**.

Source: https://www.anandtech.com/show/21241/tsmc-2nm-update-two-fabs-in-construction-one-awaiting-government-approval

The Next Step

Ever Smaller …

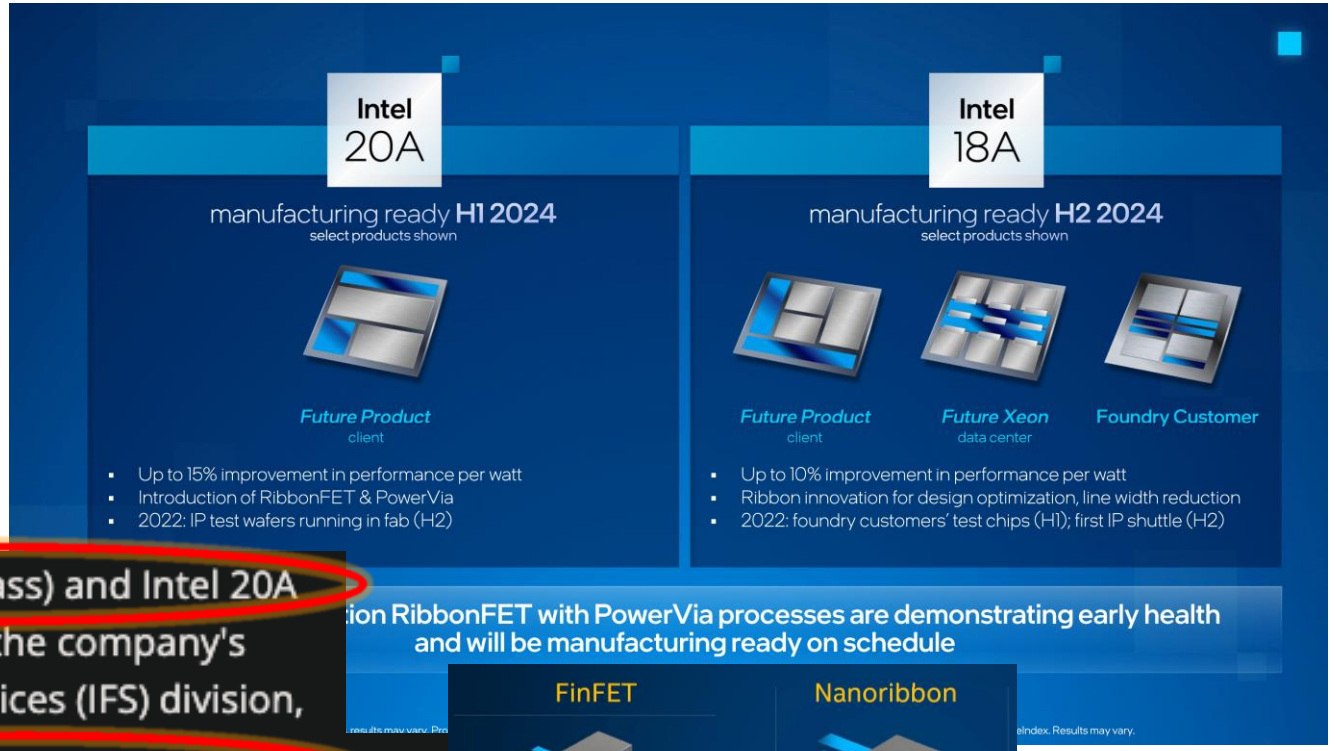
Intel Tapes Out Chips on 1.8nm and 2nm Production Nodes (Updated)

By Anton Shilov published 30 days ago

Intel moves a step closer to fabrication technology leadership.

Intel has completed chip tape-outs of its Intel 18A (1.8nm-class) and Intel 20A (2nm-class) fabrication processes that will be used to make the company's products, as well as chips for clients of its Intel Foundry Services (IFS) division,

Intel's 20A fabrication technology will rely on gate-all-around RibbonFET transistors and will use backside power delivery. Shrinking metal pitches, introducing all-new transistor structures and adding backside power delivery at the same time is a risky move, but it is expected that 20A will allow Intel to leapfrog the company's competitors — TSMC and Samsung Foundry. Intel plans to start using this node in the first half of 2024.

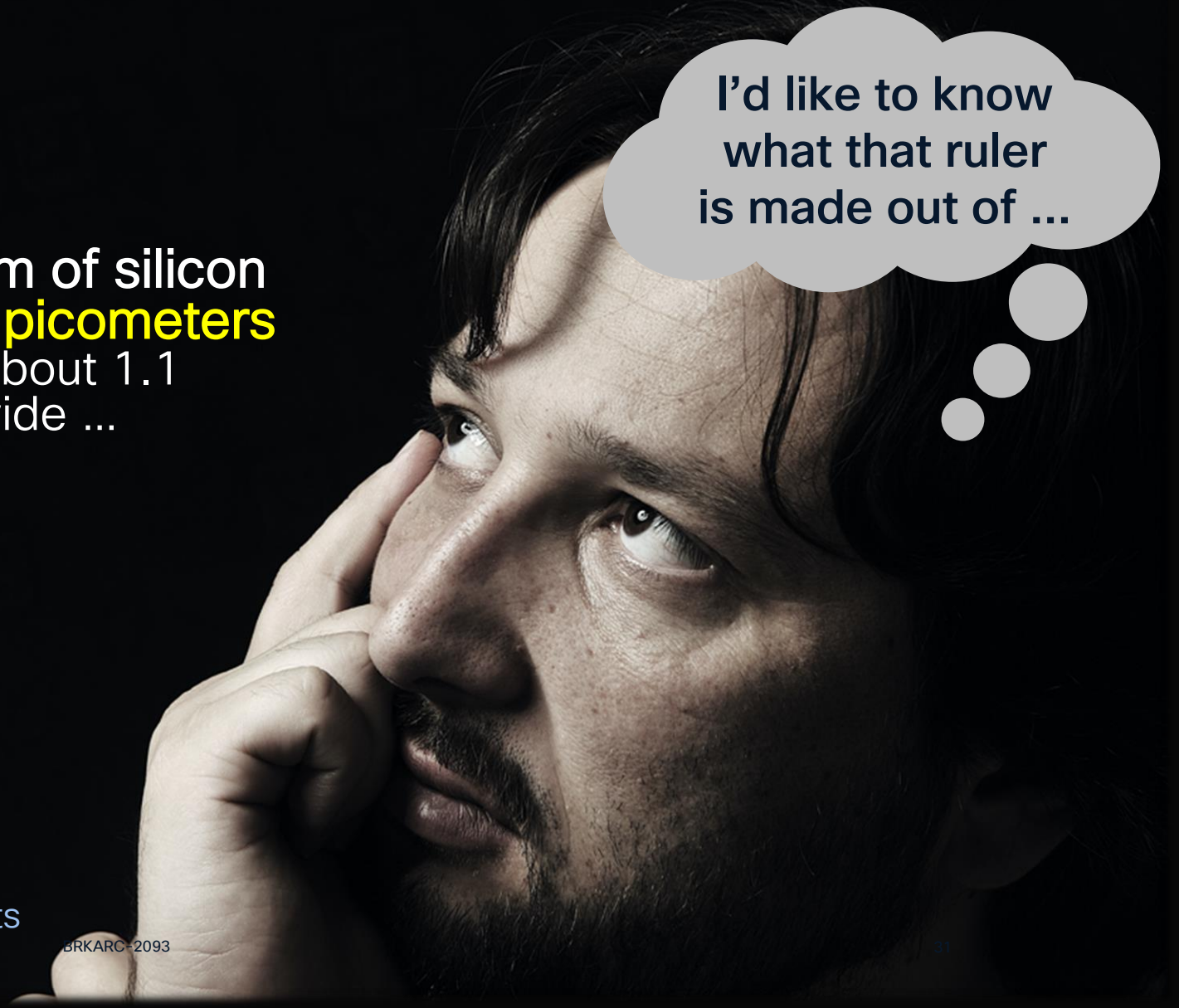


The company itself expects its 1.8nm-class manufacturing technology to be the industry's most advanced node when it enters high-volume manufacturing (HVM) in the second half of 2024.

Sources:
https://www.tomshardware.com/news/intel-completes-development-of-18a-20a-nodes
https://www.anandtech.com/show/16823/intel-accelerated-offensive-process-roadmap-updates-to-10nm-7nm-4nm-3nm-20a-18a-packaging-foundry-emib-foveros/3

How Low Can You Go?

A **single atom of silicon** is about **111 picometers** (0.11nm, or about 1.1 angstroms) wide ...



so it will be hard to get smaller than that! ☺

...but of course quantum interference effects will dominate long before that time ...

How do Humans Build Something This Small?

The size of a bus

Weight: 180 tons

Cost: $150M USD

It takes something **BIG** to make something **small** ☺

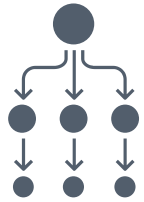
ASICs

From Definition to Deployment

How much exactly?

Numbers vary, but SemiAnalysis puts it at **$50M-$75M** for TSMC 7nm and mask costs have increased from **~$10M at 7nm** to **~$40M for 3nm**.

Source: https://www.semianalysis.com/p/the-dark-side-of-the-semiconductor



Marketing requirements | Architecture | RTL design | Synthesis | Floor planning | Fabrication

2 to 5 years

Building a new ASIC takes a lot of time and money

Simpler Deployment Options

Better **Insight and Optimization**

Increased **Security**

Most Appropriate **Scalability**

Why Does Cisco Develop Our Own Silicon?

Flexibility and Investment Protection *via Programmability*

Functional – Fast – but **FIXED** in nature

May not handle new protocols and encapsulations

May limit adoption of new technologies and solutions

WHAT'S IN A CHIP?

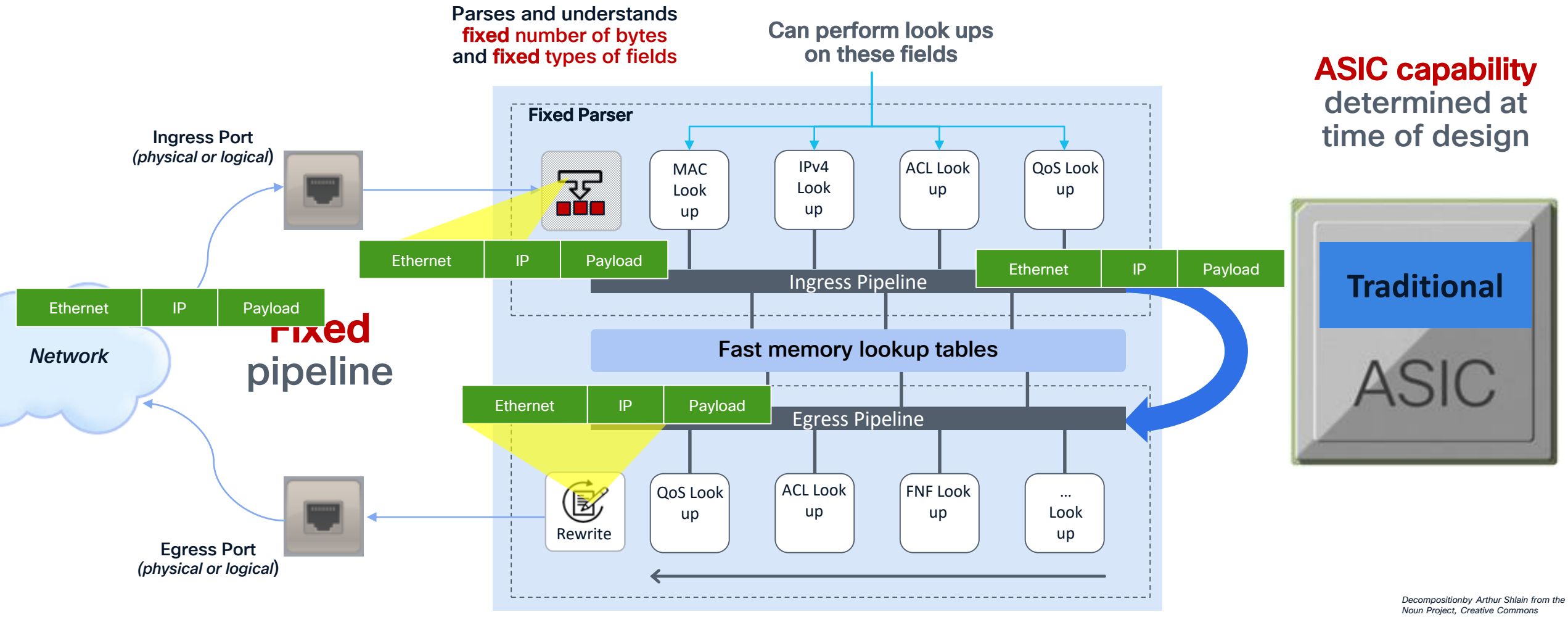
Functional – Fast – but **FIXED** in nature

May not handle new protocols and encapsulations

May limit adoption of new technologies and solutions

WHAT'S IN A CHIP?

The Past – Fixed Pipelines



Decompositionby Arthur Shlain from the Noun Project, Creative Commons

Icon courtesy of Takao Umehara from the Noun Project, Creative Commons

Network Evolution …

Automation

Security

Analytics

New **CAPABILITIES** …

Network-Wide
Assurance …

End-to-End
Policy …

New **PROTOCOLS** …

LISP

VXLAN

TrustSec

Not supported in Hardware

Options: **Punt to CPU (slow) or Drop**

Enter

PROGRAMMABLE Silicon

Much **faster** than a CPU

Much more **capable** than an FPGA

ASIC Engineer

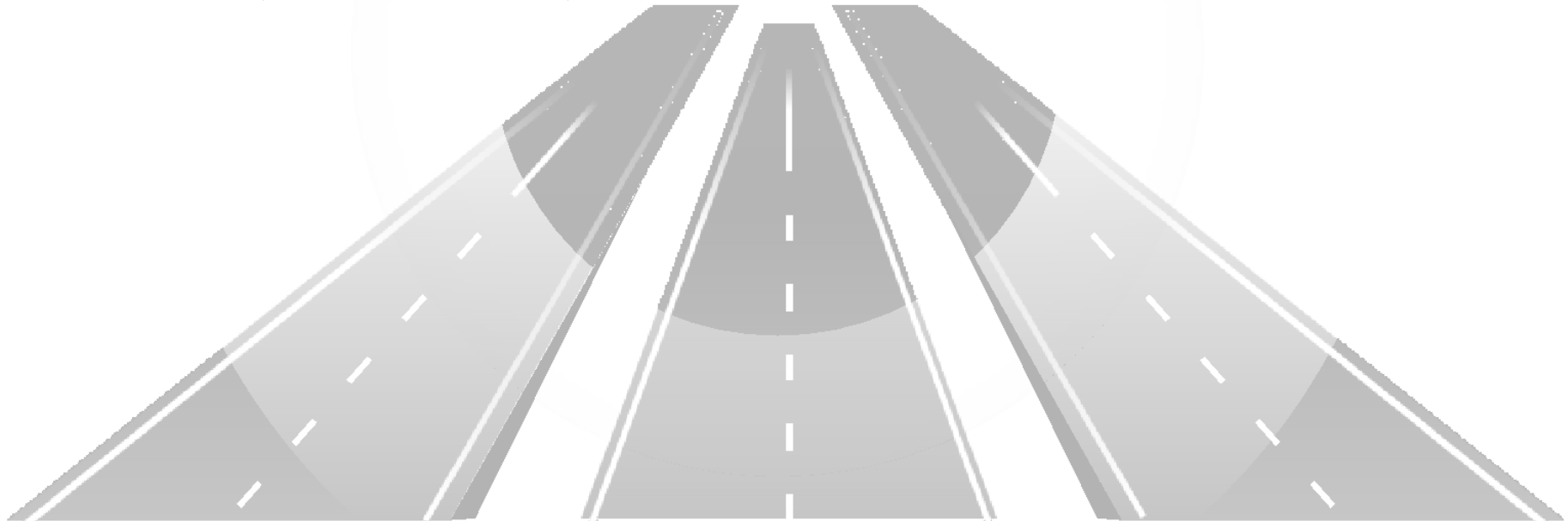
Microcode Update

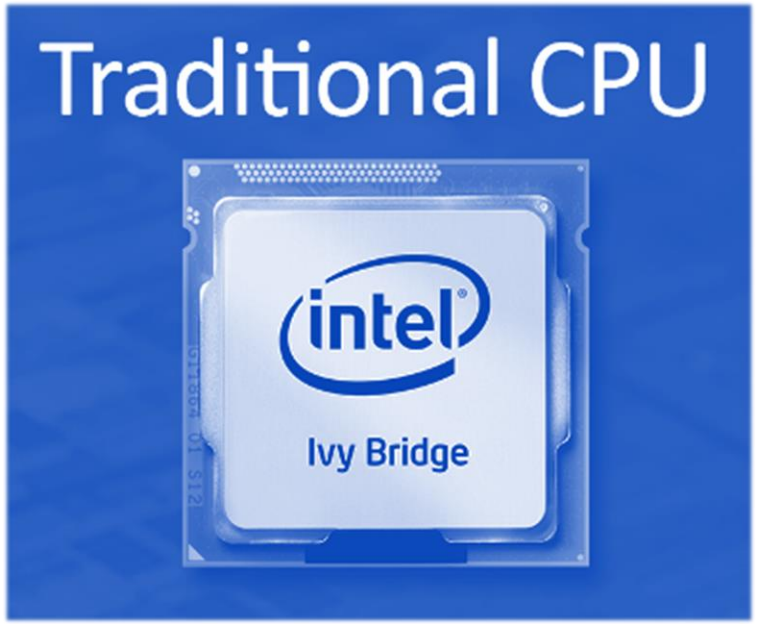
New ASIC Functionality

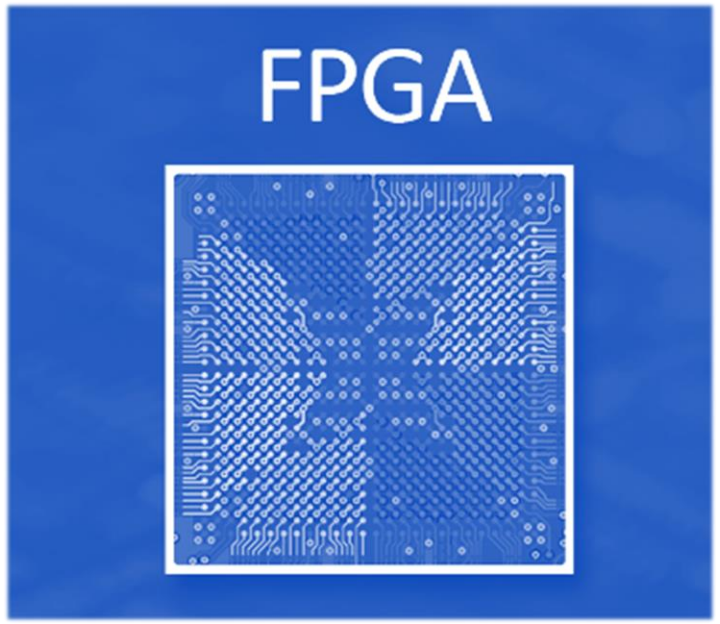
Hold on!

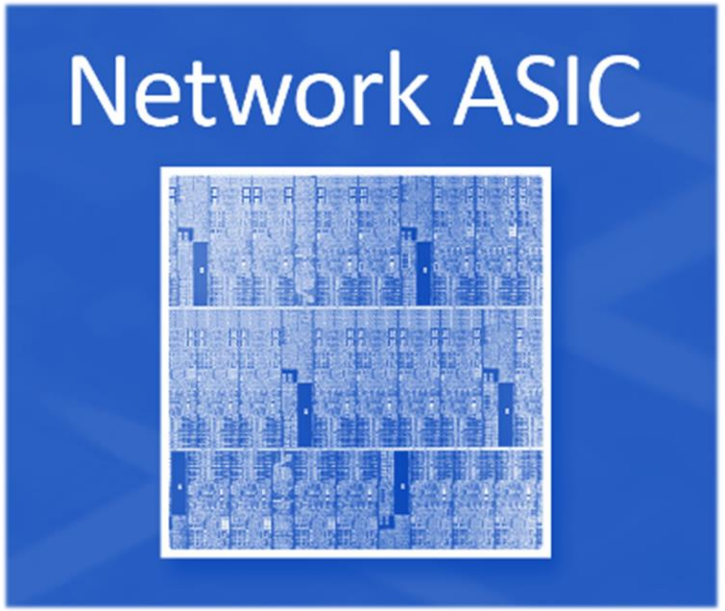
Why can't we **just forward traffic** the **traditional ways** we always have in networking devices –

CPU, **FPGA**, and **Network ASIC?**









All of these options have their

Own Tradeoffs

(Central Processing Unit)

Very flexible

Offers only **moderate performance** for networking applications

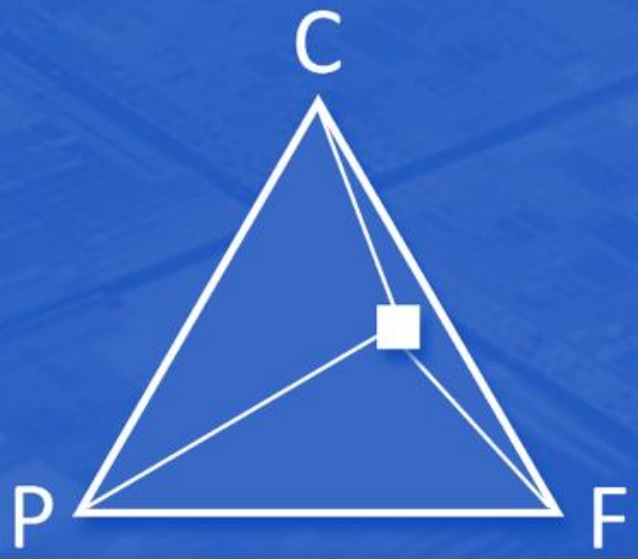
Ranges in cost from **inexpensive** to **quite expensive**

(Central Processing Unit)

Traditional CPU

Control-plane use in devices

Data-plane use for **relatively low performance needs** (WAN edge, etc)



Cost

Performance

Flexibility

Very flexible

Offers only **moderate performance** for networking applications

Ranges in cost from **inexpensive** to **quite expensive**

(Field Programmable Gate Array)

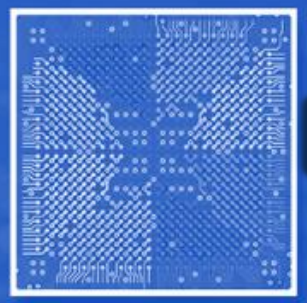
Fairly flexible

Offers **better performance**
than CPUs for networking applications

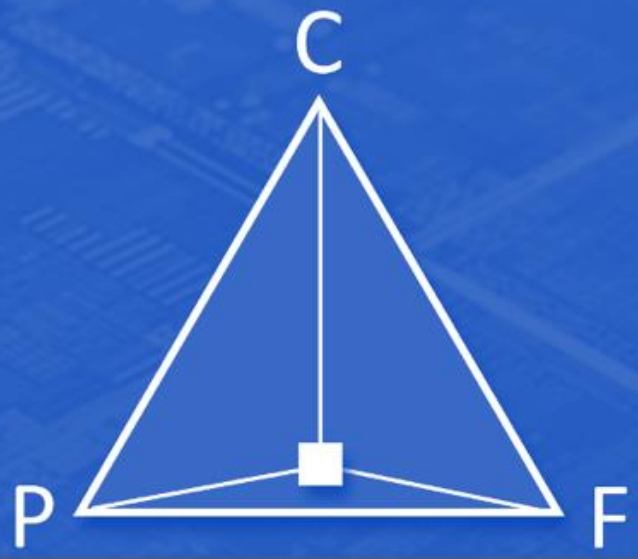
But are **quite expensive**
in terms of cost per bits moved
(not typically networking-optimized)

(Field Programmable Gate Array)

FPGA



Typically used as “gap fillers” in network devices paired with or replacing a traditional ASIC, and providing a **few key data plane capabilities** that a traditional fixed ASIC might lack



Cost

Performance

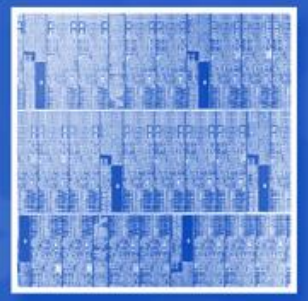
Flexibility

Fairly flexible

Offers **better performance**
than CPUs for networking applications

But are **quite expensive**
in terms of cost per bits moved
(not typically networking-optimized)

Traditional Fixed Network ASIC



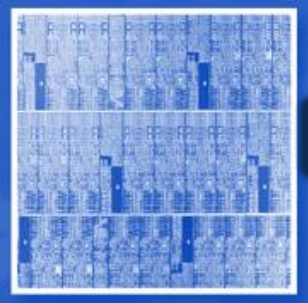
(Application Specific Integrated Circuit)

Offers excellent **performance**
for networking applications

And are **quite cost-effective**
for the performance
they provide

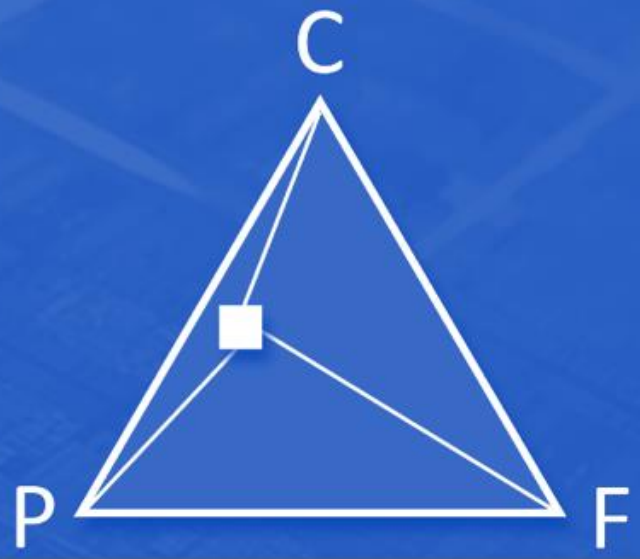
but are **not very flexible**

Traditional Fixed Network ASIC



(Application Specific Integrated Circuit)

Used as data-plane in many networking devices providing performance into terabits per second, but not readily adaptable to handle new protocols and encapsulations as they emerge



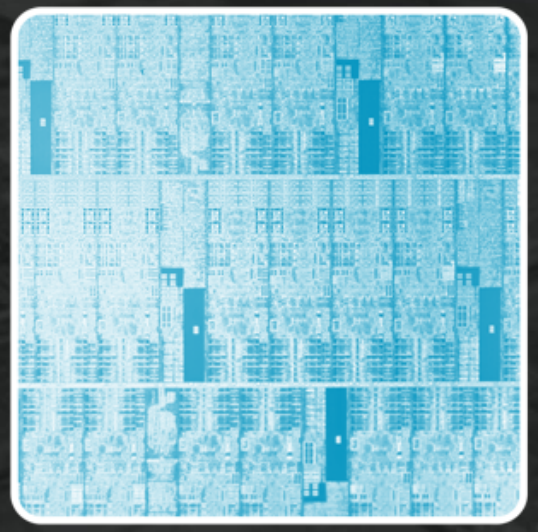
Cost	✓
Performance	✓
Flexibility	✗

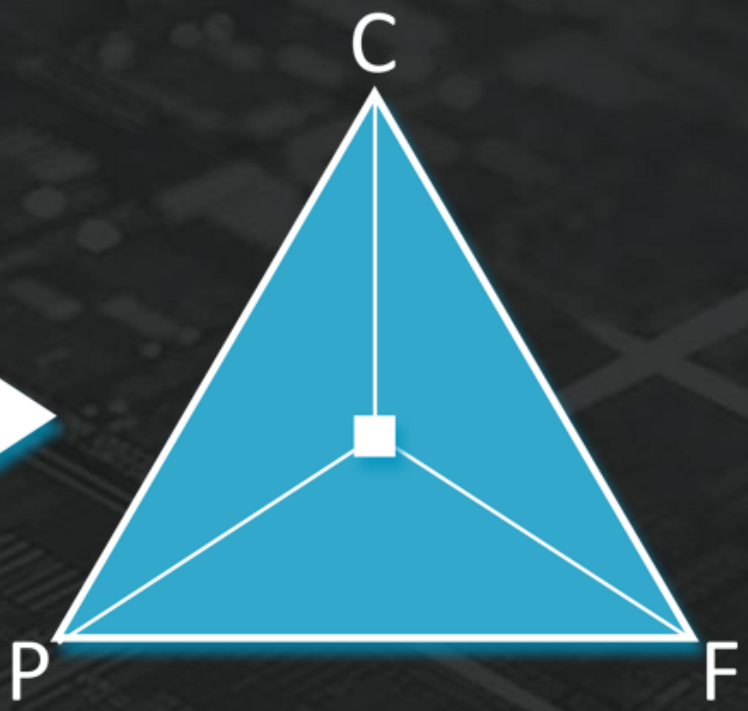
Offers excellent **performance** for networking applications

And are **quite cost-effective** for the performance they provide

but are **not very flexible**

Programmable ASIC





COST ✓

PERFORMANCE ✓

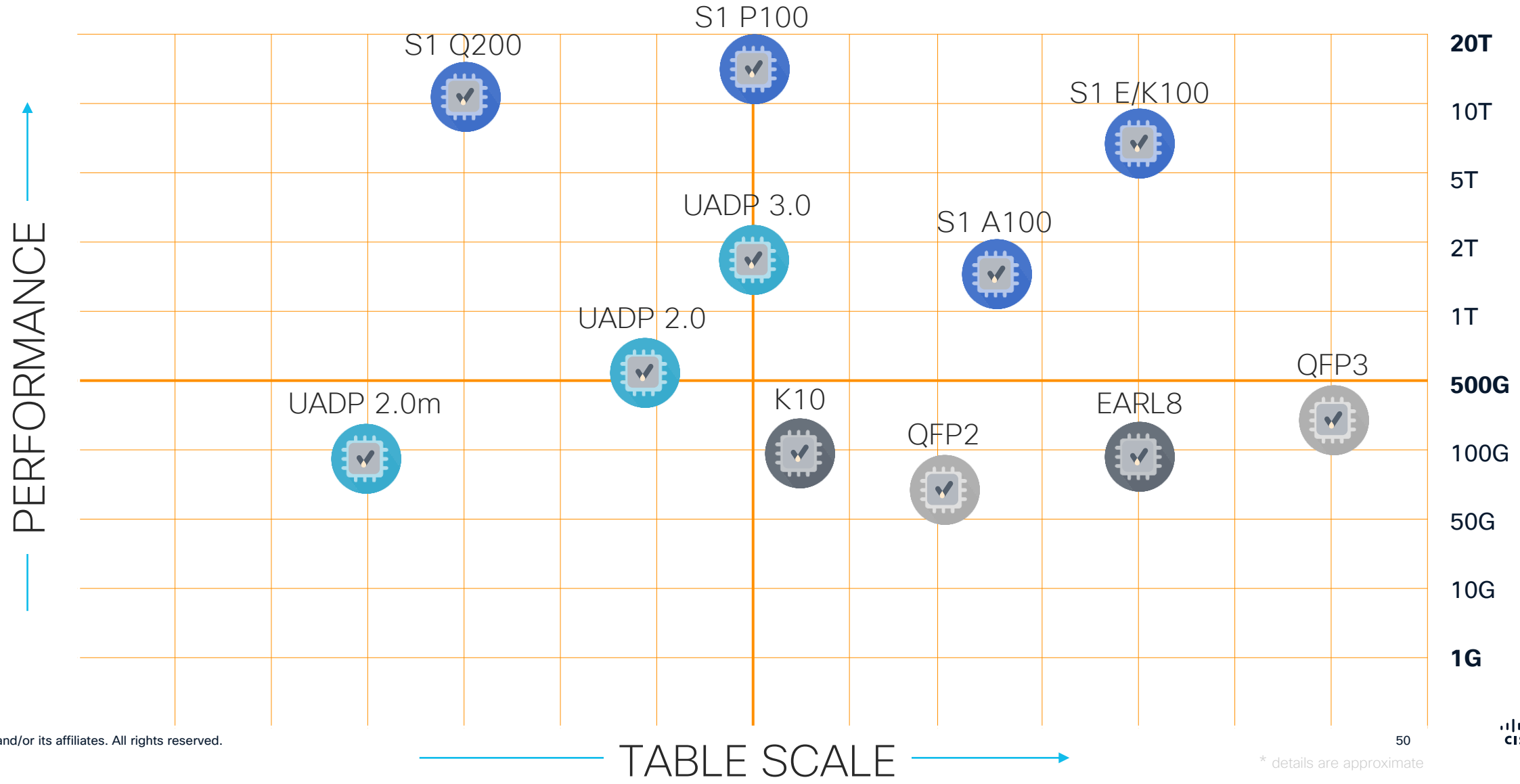
FLEXIBILITY ✓

Balance All Three Factors

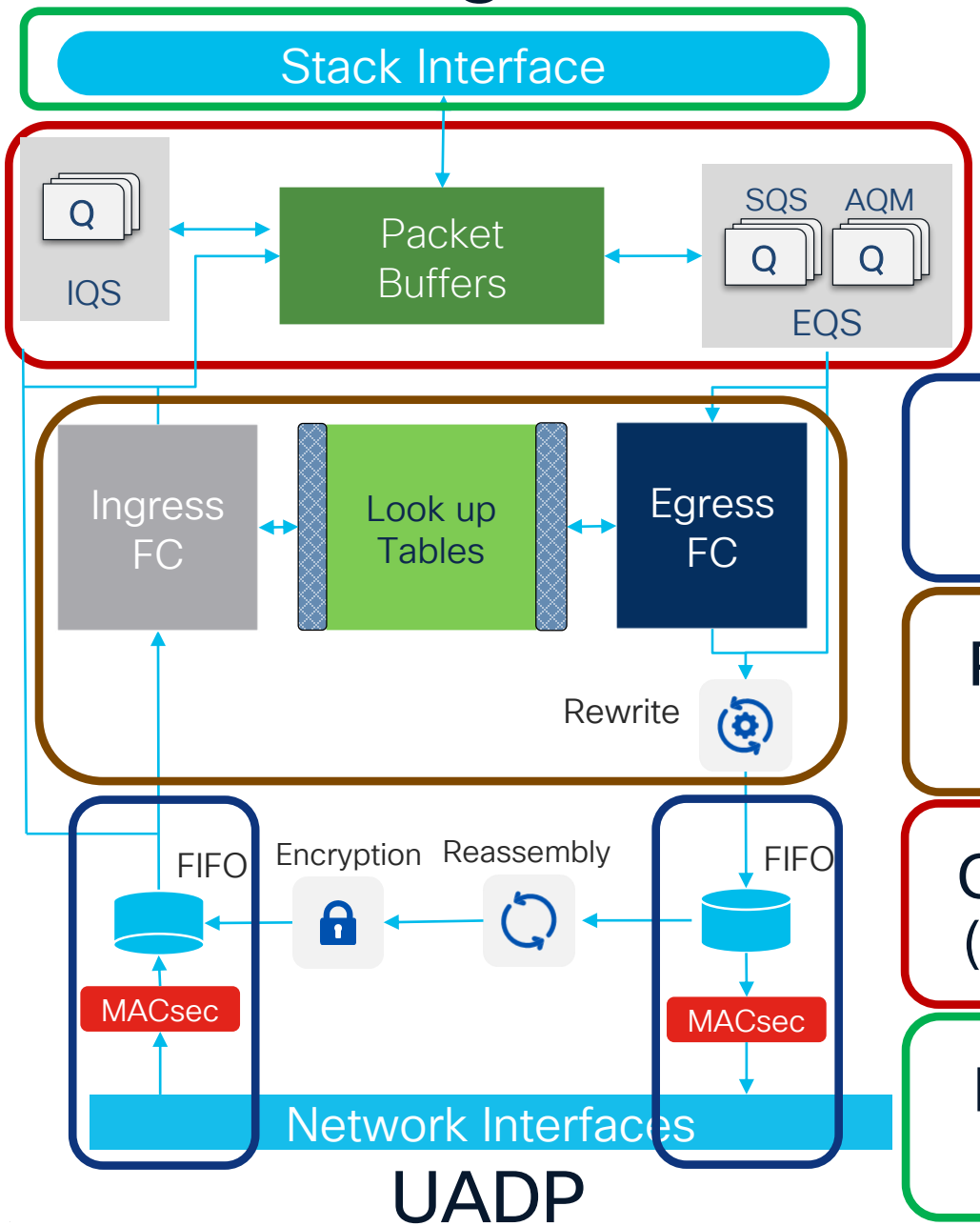
Flexible Networking Silicon – Overview

Cisco Networking ASICs

Performance vs. Features



Networking ASICs: Main blocks

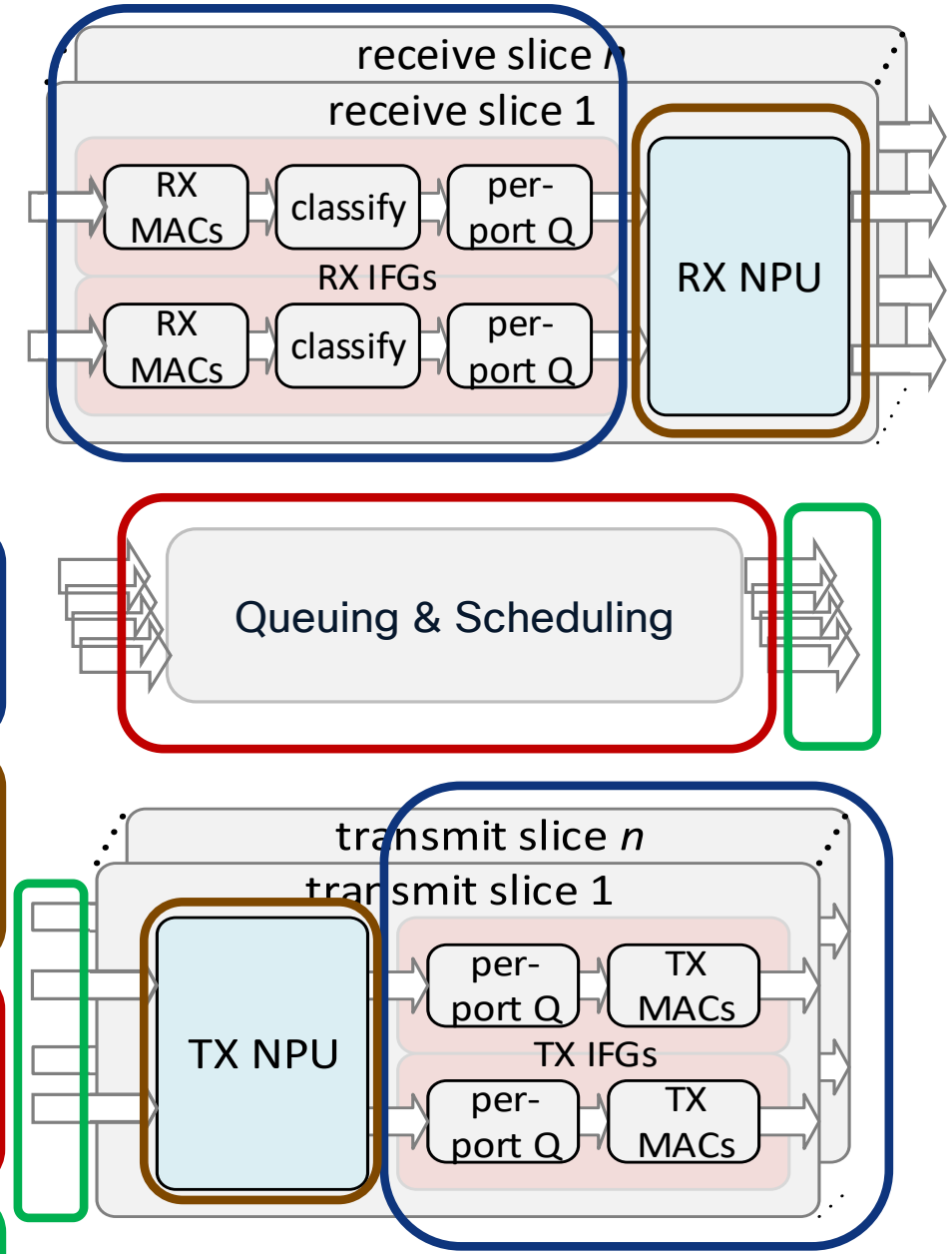


Port Interfaces
(aka NIF, IFG/EFG)

Packet Processor
(aka FC, NPU)

Queuing/Buffering
(aka TM, PBC, I/EQS)

Packet Exchange
(aka Stack, Fabric)



Silicon One

cisco

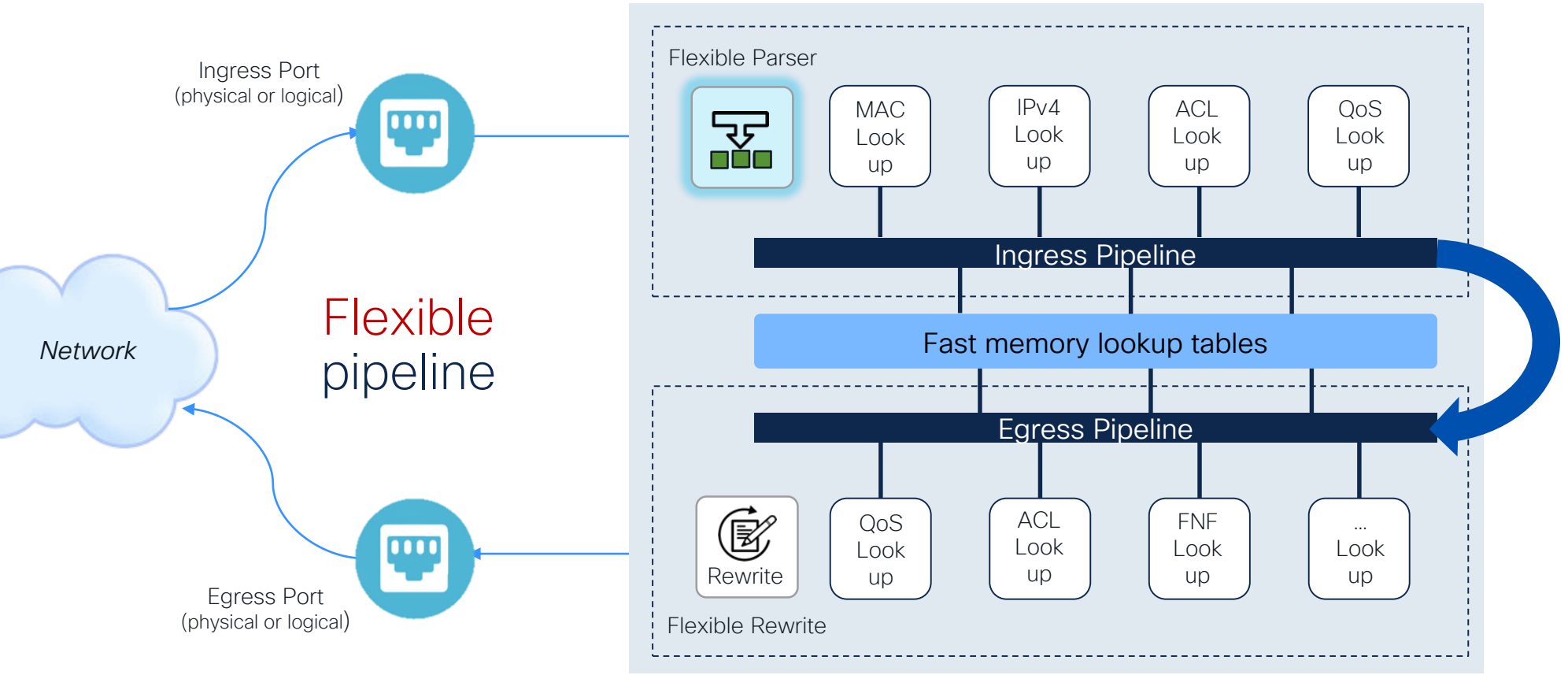
Programmable Network ASICs

Flexible Parsing

Flexible Parsing

Look deep into the packet header, with **programmable field parsing**

Parses and understands multiple programmable headers with flexible field definitions

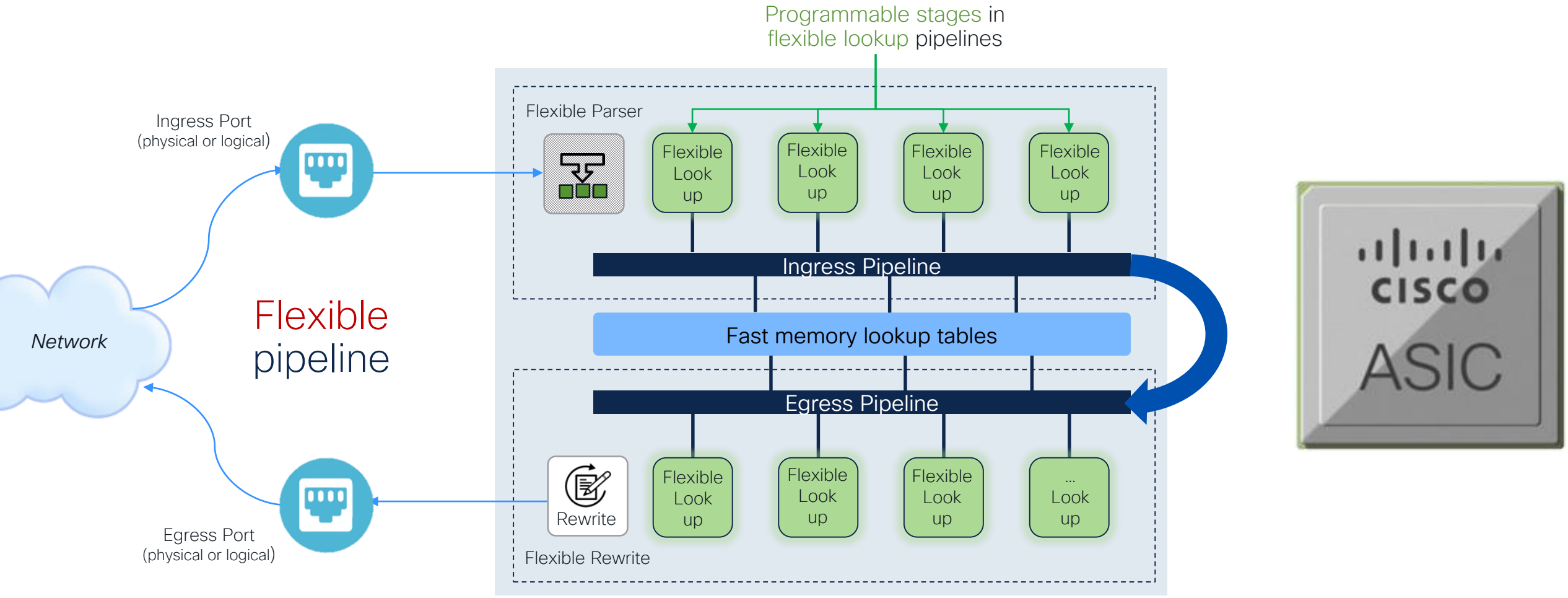


Programmable Network ASICs

Flexible Lookups

Flexible Lookups

Multi-stage packet handling, with **flexible packet lookups** at every step

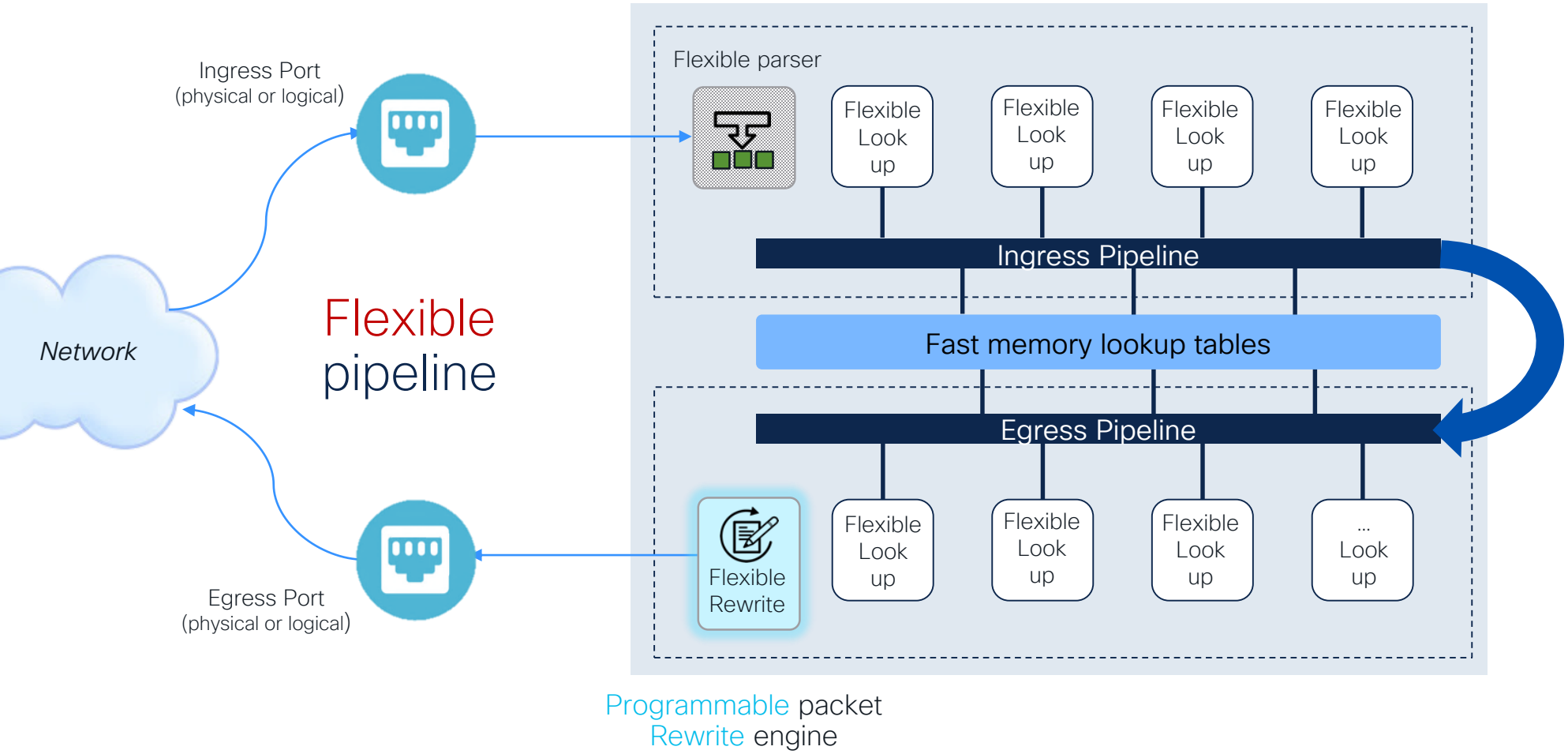


Programmable Network ASICs

Flexible Rewrites

Flexible Rewrites

Flexible packet handling and forwarding, with a **programmable packet rewrite**



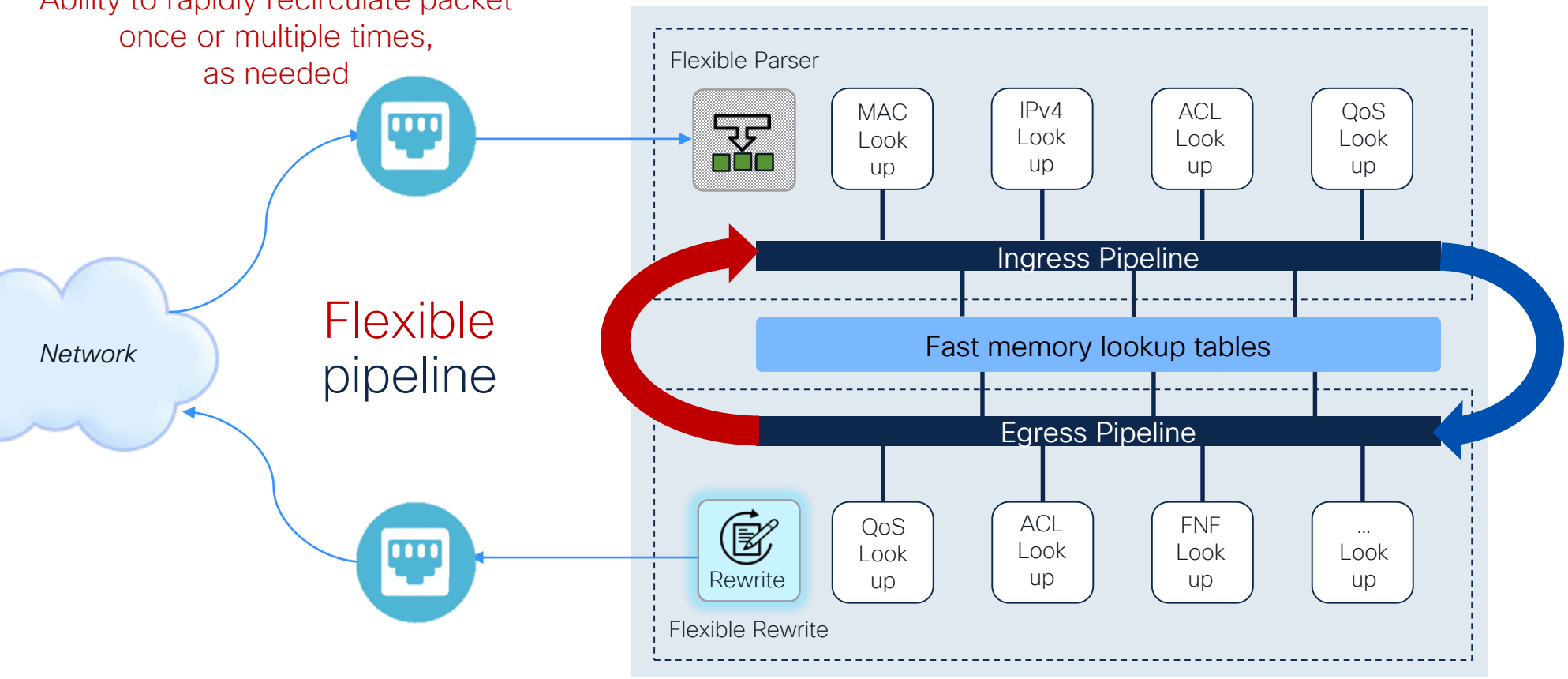
Programmable Network ASICs

Optimized Recirculation

Optimized Recirculation

Highly **optimized recirc path** for packet header addition / removal / forwarding

Ability to rapidly recirculate packet once or multiple times, as needed



Campus Networking Silicon

Cisco Forwarding Silicon

Cisco UADP

UADP 2.0mini 120G | UADP 2.0/XL 240G | UADP 2.0sec 480G | UADP 3.0 1.6T | UADP 3.0Sec 1.6T

Cisco Silicon One™

A100 1.5T | Q202L 3.2T | Q201 L 6.4T | E100 6.4T | K100 6.4T | Q211L 8T | Q200L 12.8T | P100 19.2T | G100 25.6T | G200 51.2T

Flexible Pipelines
Investment Protection

Adaptable Tables
Universal Deployment

Scalable Resources
Enhanced Scale and Buffering

Flexible and Programmable ASICs – Adapt to New Technologies

Catalyst 9000 Family

Common Building Blocks

Programmable x86 multicore CPU

Application hosting
Secure containers

Open Cisco IOS XE Polaris

Model-driven APIs
Modular patching

Flexible, programmable ASICs UADP and Silicon One

Programmable pipeline
Flexible tables

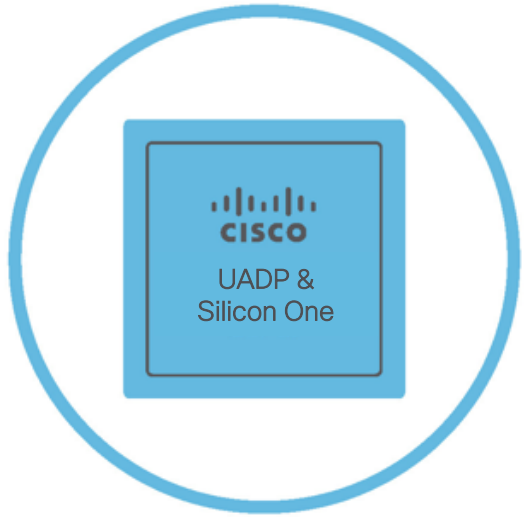
Cisco 9000 Series – Common Building Blocks

Multi-Core x86 CPU

Application Hosting
Secure Containers

Cisco IOS XE

Model-Driven APIs
Modular Patching



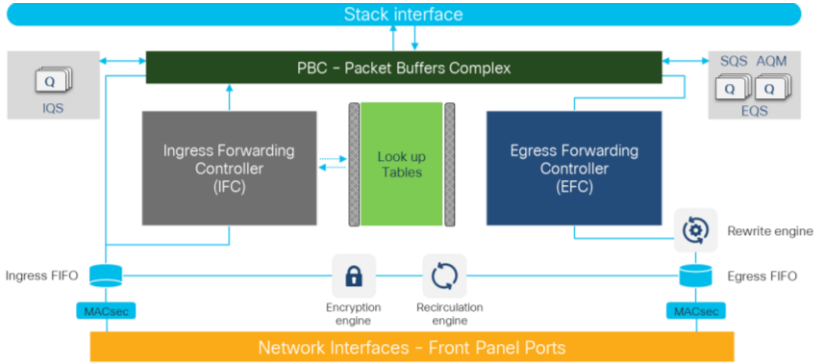
Cisco UADP & Silicon One

Programmable Pipeline
Flexible Tables

Unified Access Data Plane
UADP (aka Doppler)

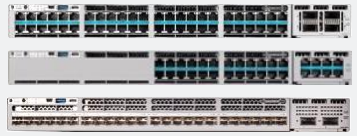
Catalyst 9000 with UADP

The industry's first Programmable Campus ASIC Family



Catalyst 9200CX
UADP 2.0 mini

Catalyst 9300X/LM
UADP 2.0sec

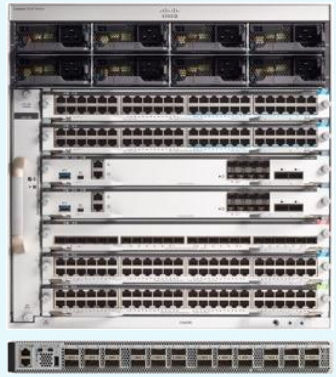


Catalyst 9400X-SUP2
UADP 3.0sec

Catalyst 9200/L
UADP 2.0 mini

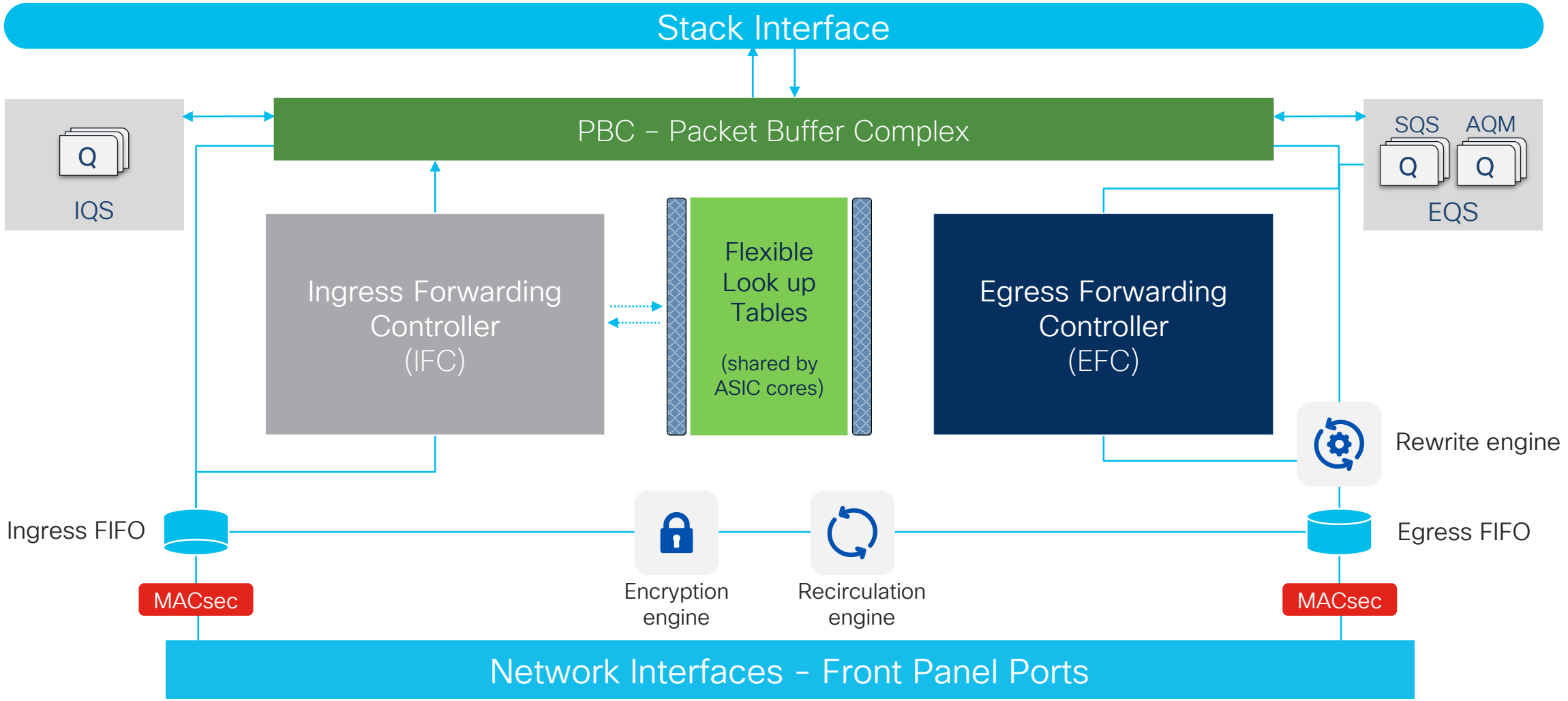
Catalyst 9300/L
UADP 2.0

Catalyst 9400-SUP1 & 9500
UADP 2.0XL



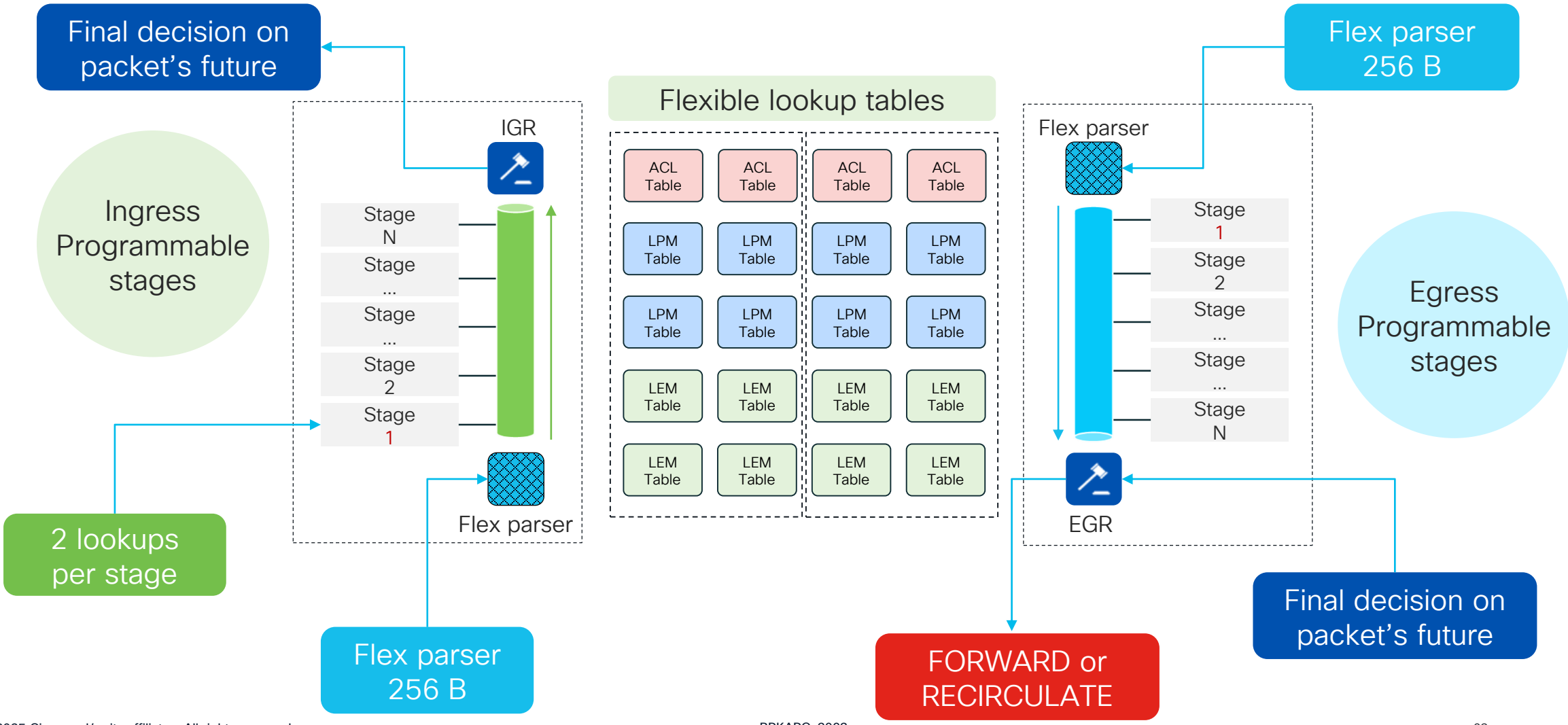
Catalyst 9500(H) & 9600-SUP1
UADP 3.0

Cisco UADP High Level

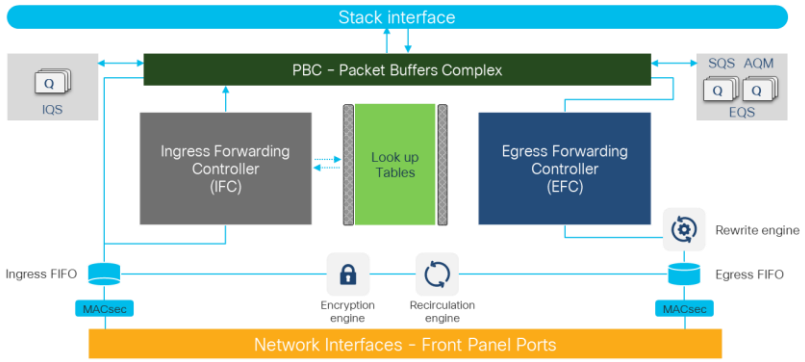


Cisco UADP – Flexible Lookups

Programmable Ingress and Egress Processing Stages



UADP Evolution



Catalyst 9400 Sup2
2022 – UADP 3.0sec

Catalyst 9500/9600
2019 – UADP 3.0

Catalyst 9300X
2022 – UADP 2.0sec

Catalyst 9400/9500
2018 – UADP 2.0XL

Catalyst 3850 mGig
2015 – UADP 1.1

Catalyst 3850
2013 – UADP 1.0

Catalyst 9200CX
2022 – UADP 2.0mini

Catalyst 9200
2019 – UADP 2.0mini

Catalyst 9300
2017 – UADP 2.0

Current UADP Family

Common ASIC Architecture for Switching in Access, Distribution & Core

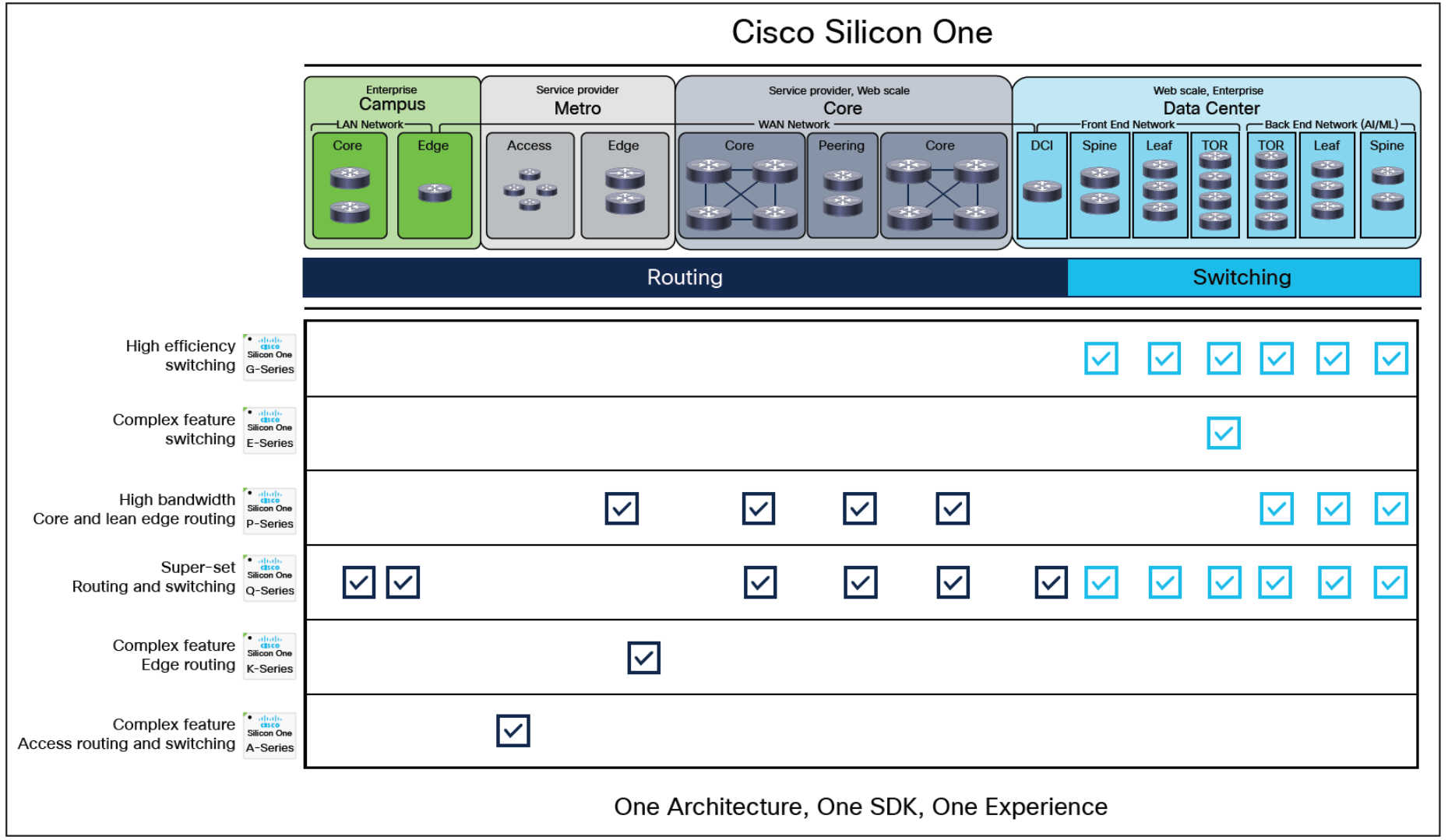
UADP 2.0m	UADP 2.0 XL	UADP 2.0sec	UADP 3.0	UADP 3.0sec
100 Gbps	240 Gbps	480 Gbps	1.6 Tbps	1.6 Tbps
16nm FinFET	28nm FinFET	16nm FinFET	16nm FinFET	16nm FinFET
1 Core + ARM CPU	2 Core	1 Core + SEC	2 Core	2 Core + SEC

- Multiple generations and formats, same architecture
- Rich flexible forwarding & services memories
- First fully programmable microcode network silicon

- **Multiple functions**: system-on-chip or line-card
- **Multiple form factors**: fixed or modular
- **Multiple places:** Access, Distribution and Core

Cisco Silicon One

Cisco Silicon One – Convergence without compromise



www.cisco.com/c/en/us/solutions/silicon-one.html

Cisco Silicon One in Campus

A100

1.3 Tbps
16nm FinFET
1 Slice SOC
SMS + DDR4

E100

6.4 Tbps
7nm FinFET
2 Slice SOC
SMS

K100

6.4 Tbps
7nm FinFET
2 Slice SOC
SMS + HBM2E

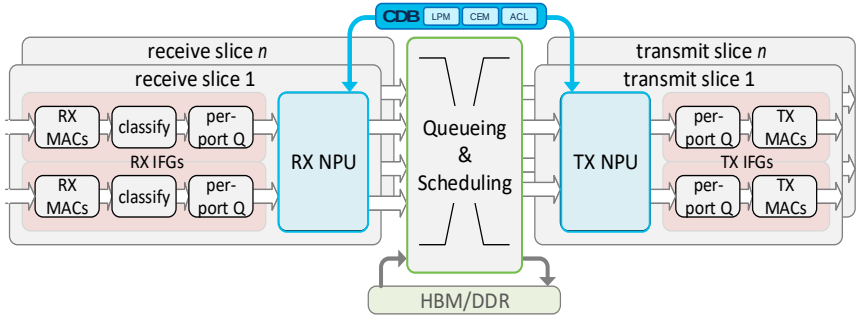
Q200

12.8 Tbps
7nm FinFET
6 Slice SOC
SMS + HBM2

- Comprehensive routing, with switching efficiency
- Onboard and expandable memories
- Flexible P4 NPL programmable packet processing

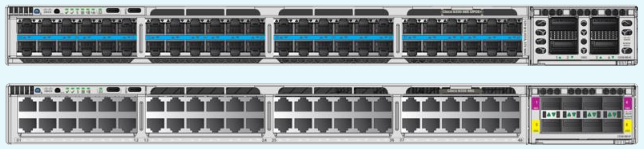
- Multiple functions: system-on-chip, line-card or fabric
- Multiple form factors: fixed or modular
- Multiple networks: Enterprise, Data Center and SP

Cisco C9000 with Silicon One



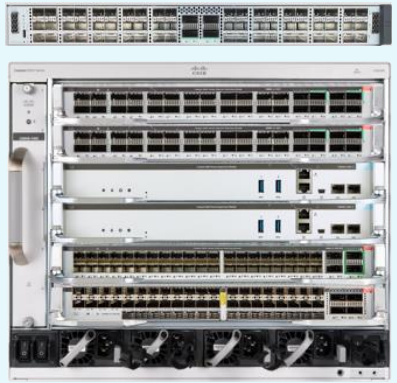
Cisco C9350 Series

Powered by Cisco Silicon One **A100**



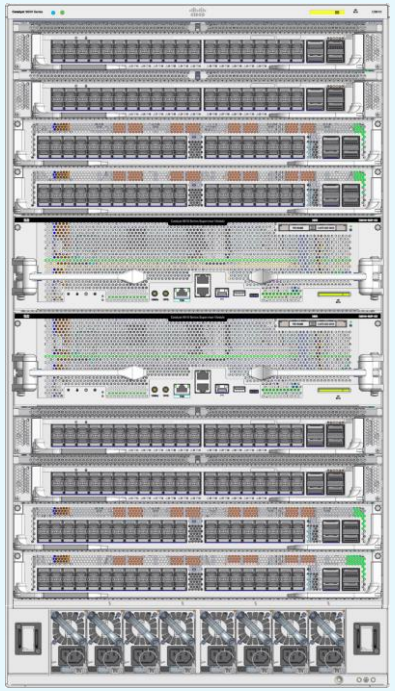
Catalyst 9500X & 9600X-SUP-2

Powered by Cisco Silicon One **Q200**



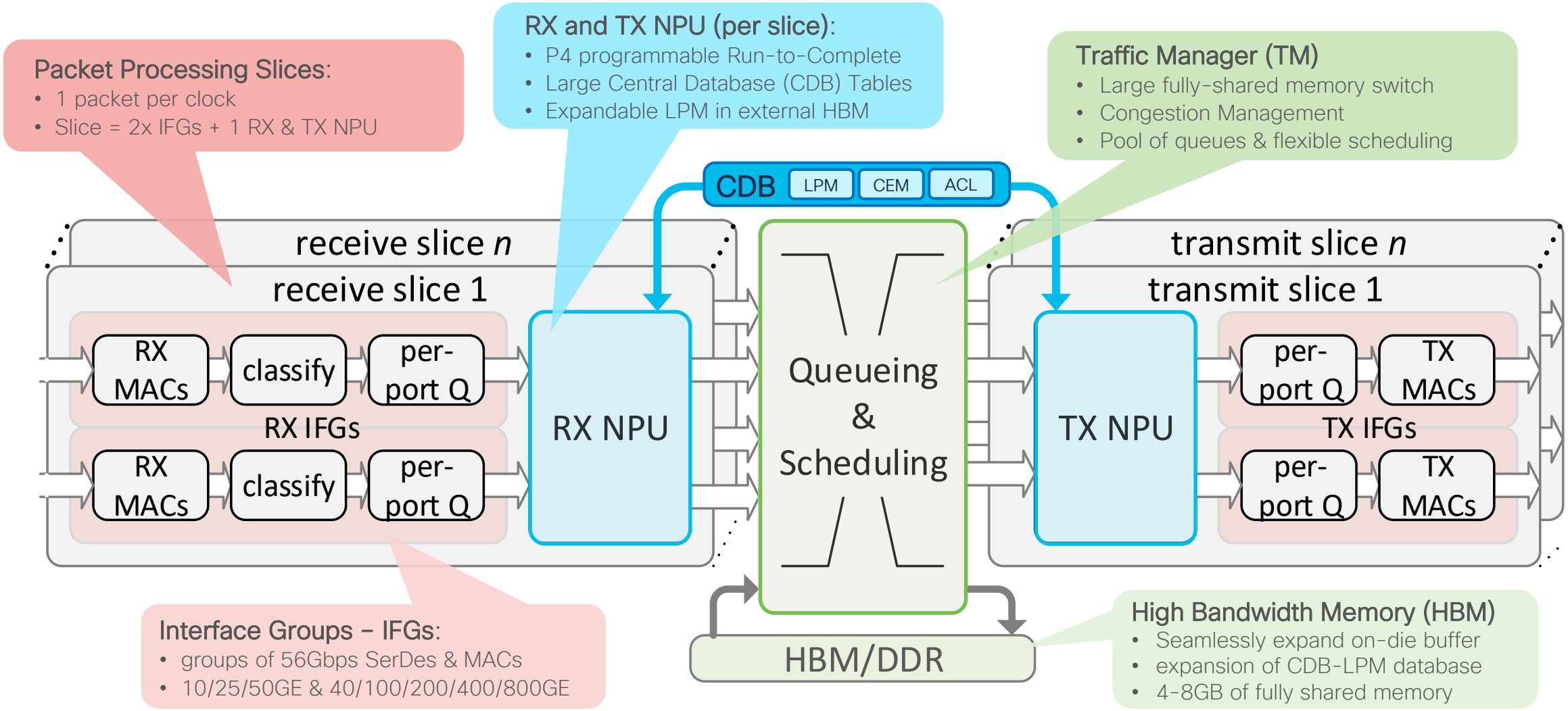
Cisco C9610-SUP-3 & 3XL

Powered by Cisco Silicon One **E100** & **K100**



Cisco Silicon One

Top Level



Cisco Silicon One – Central Databases

Onboard LPM, CEM & ACL memory

The CDB includes the Central L2/L3 Forwarding and ACL databases:

LPM – SRAM database for IP/mask routing implemented by Longest Prefix Match algorithm

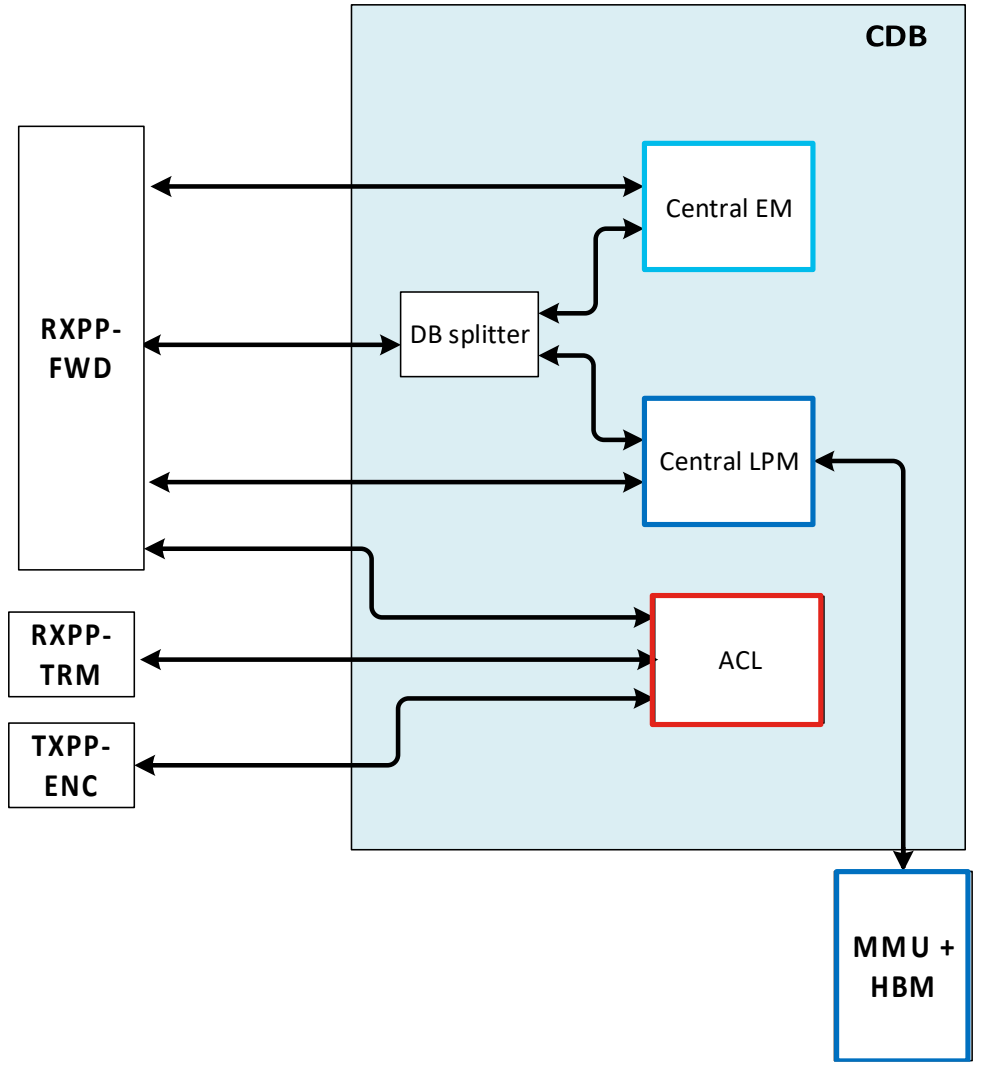
- Primarily used by IPv4 and IPv6 unicast routing
- LPM can be extended (from CDB) to HBM

CEM – SRAM database for MAC & Host (/48, /32 or /128), Multicast & Labels implemented by Exact Match algorithm

- For features using an exact match (every bit, no mask)
- CEM can be flexibly reallocated for different tables

ACL – TCAM classification database, contains Security, QoS and Services Access Control List entries

- For features that use (match criteria + action) policies
- Includes Security ACL (permit or deny)



* Exact scale depends on IP/mask distribution (contiguous vs. random) and hash efficiency.

What's in A100, E100, K100?

Shared innovative S1 Architecture

- Multi-Slice Run-To-Complete NPU
- Low Power with High-Performance
- P4-based NPL and Common SDK
- VOQ-based SMS and HBM/DDR
- Standalone, Module or Fabric mode

Newer Longest Prefix Match (LPM)*

- High-scale IP/mask with HBM extension
- Improved efficiency with embedded Hash

Newer & larger Exact Match (CEM)*

- High-scale MAC, ARP, Multicast, SGT, NAT, etc.
- More flexible for various 'exact match' features

New & larger Policy Match (HCAM)*

- High-scale Algorithmic TCAM + Hash SRAM
- Robust support for ACL, QoS & FNF features
- Better Ingress and Egress policy support

Hardware Flexible NetFlow*

- FNF records use HCAM (wide-entries)
- Support for AVC/NBAR and ETA/XDR

Hardware Data Encryption*

- Embedded AES-GCM 256-Bit Crypto Engines
- Line-rate (LAN) MACsec
- Hardware WAN MACsec and IPsec

Hardware Precision Time*

- Hardware PTP 1588, AVB, AES67 & G8275
- Support for IT/OT Frame Preemption

Advanced QoS & Buffering*

- More VoQs, Policers and Shapers
- Enhanced WRED and HQoS

Flexible Telemetry Counters*

- Millions of Programmable counters
- Configurable Feature allocation

Lower & Higher-Speed Ports*

- 10/100/1000M and 2.5/5/10G mGig RJ45
- 1/10/25G/50G SFP and 40/100/400G QSFP

... and much more

What's in A100, E100, K100?

Shared innovative S1 Architecture

- Multi-Slice Run-To-Complete NPU
- Low Power with High-Performance
- P4-based NPL and Common SDK
- VOO
- Sta

Newer

- Hig
- Imp

Newer

- Hig
- More

New & larger Policy Match (HCAM)*

- High-scale Algorithmic TCAM + Hash SRAM
- Robust support for ACL, QoS & FNF features
- Better Ingress and Egress policy support

Lots of things!

Hardware Flexible NetFlow*

- FNF records use HCAM (wide-entries)
- Support for AVC/NBAR and ETA/XDR

Hardware Data Encryption*

-Bit Crypto Engines

IPsec

ne*

ES67 & G8275

emption

ring*

pers

Flexible Telemetry Counters*

- Millions of Programmable counters
- Configurable Feature allocation

Lower & Higher-Speed Ports*

- 10/100/1000M and 2.5/5/10G mGig RJ45
- 1/10/25G/50G SFP and 40/100/400G QSFP

... and much more

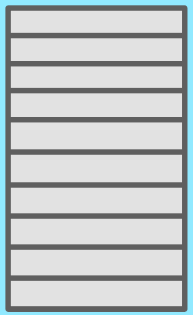
What's HCAM?

Algorithmic Hash-based TCAM for High-Scale ACL/NetFlow

8K TCAM

- Key based – variable mask
- Supports flexible matching logic
- Complex & versatile searches
- Less-Dense (fewer bits in same area)
- Faster lookup speed
- Power & Heat intensive
- Expensive to scale

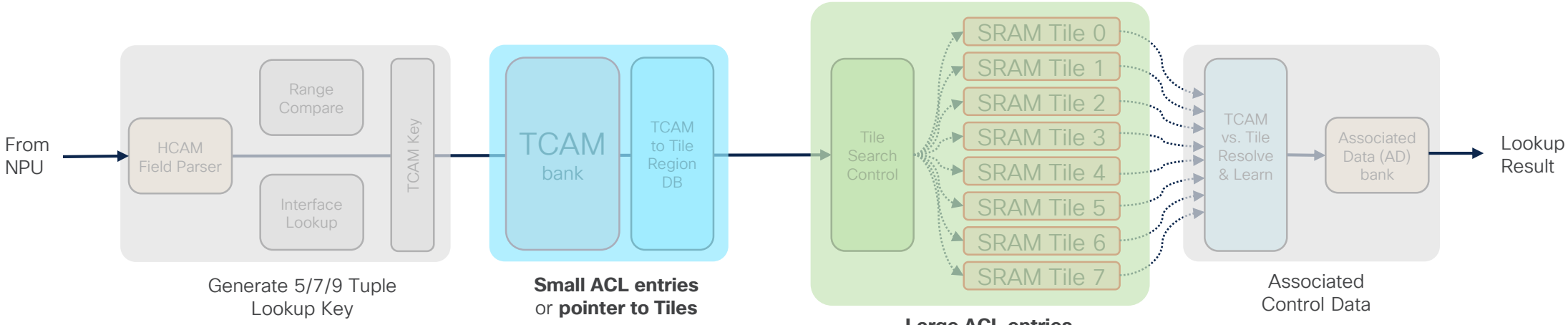
TCAM



128K SRAM (8x Tiles of 16K)

SRAM
(Tiles)

- Hash based – full mask
- Simple, specific searches
- Rigid exact matching logic
- More-Dense (more bits in same area)
- Slightly slower speed
- Lower Power & Heat
- Less costly to scale



Generate 5/7/9 Tuple Lookup Key

Small ACL entries or **pointer to Tiles**

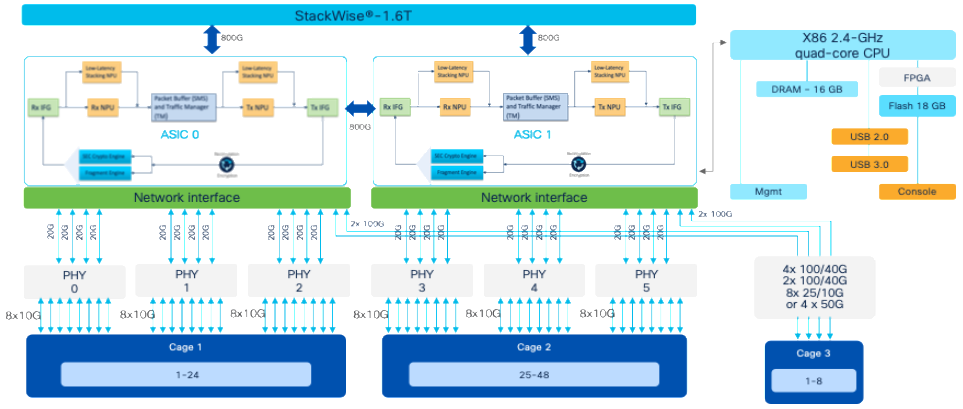
Large ACL entries or **NetFlow entries**

Associated Control Data

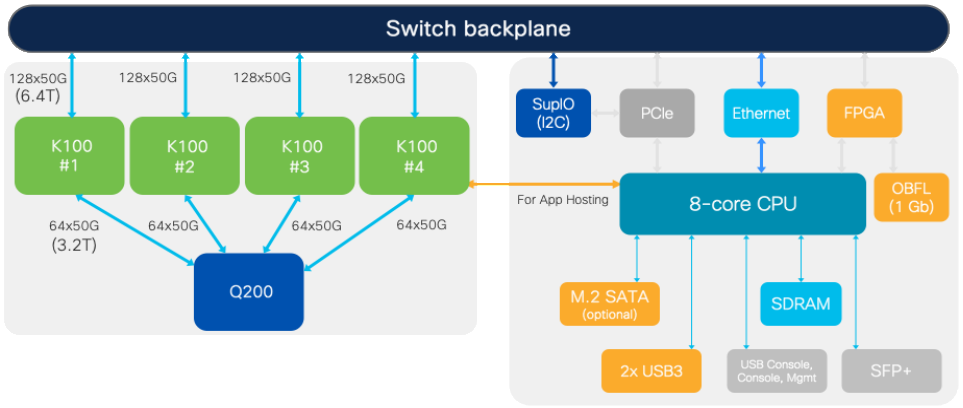
NOTE: IPv6 takes 2 entries in both TCAM and SRAM

Cisco Silicon One on Cisco C9000 Switches

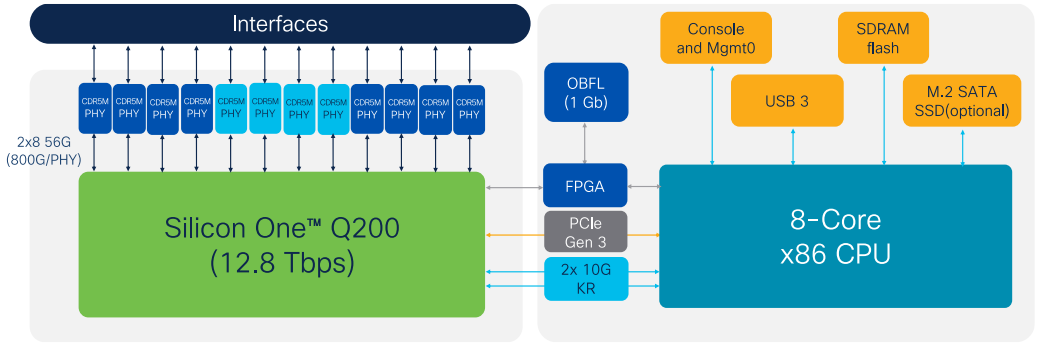
C9350-48HX & TX



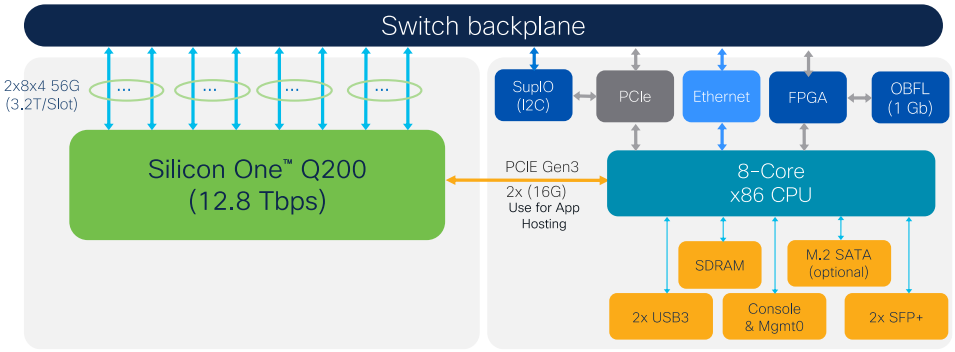
C9610-SUP-3/XL



C9500X-28C8D & 60L4D



C9600X-SUP-2

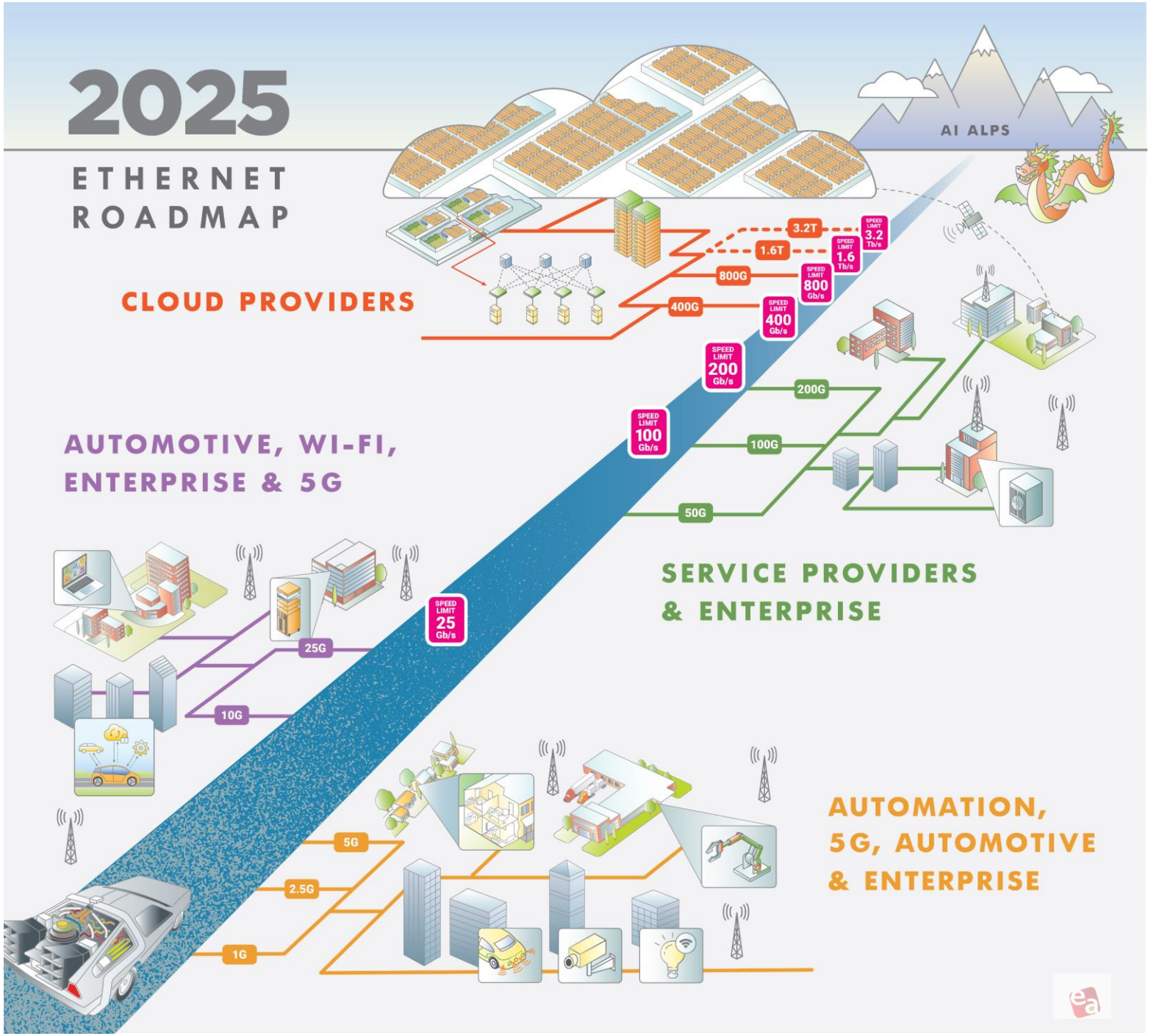


Latest in Ethernet

What's Up with Ethernet?

The Ethernet Alliance 2025 10th Anniversary Roadmap is here!!

ethernet alliance

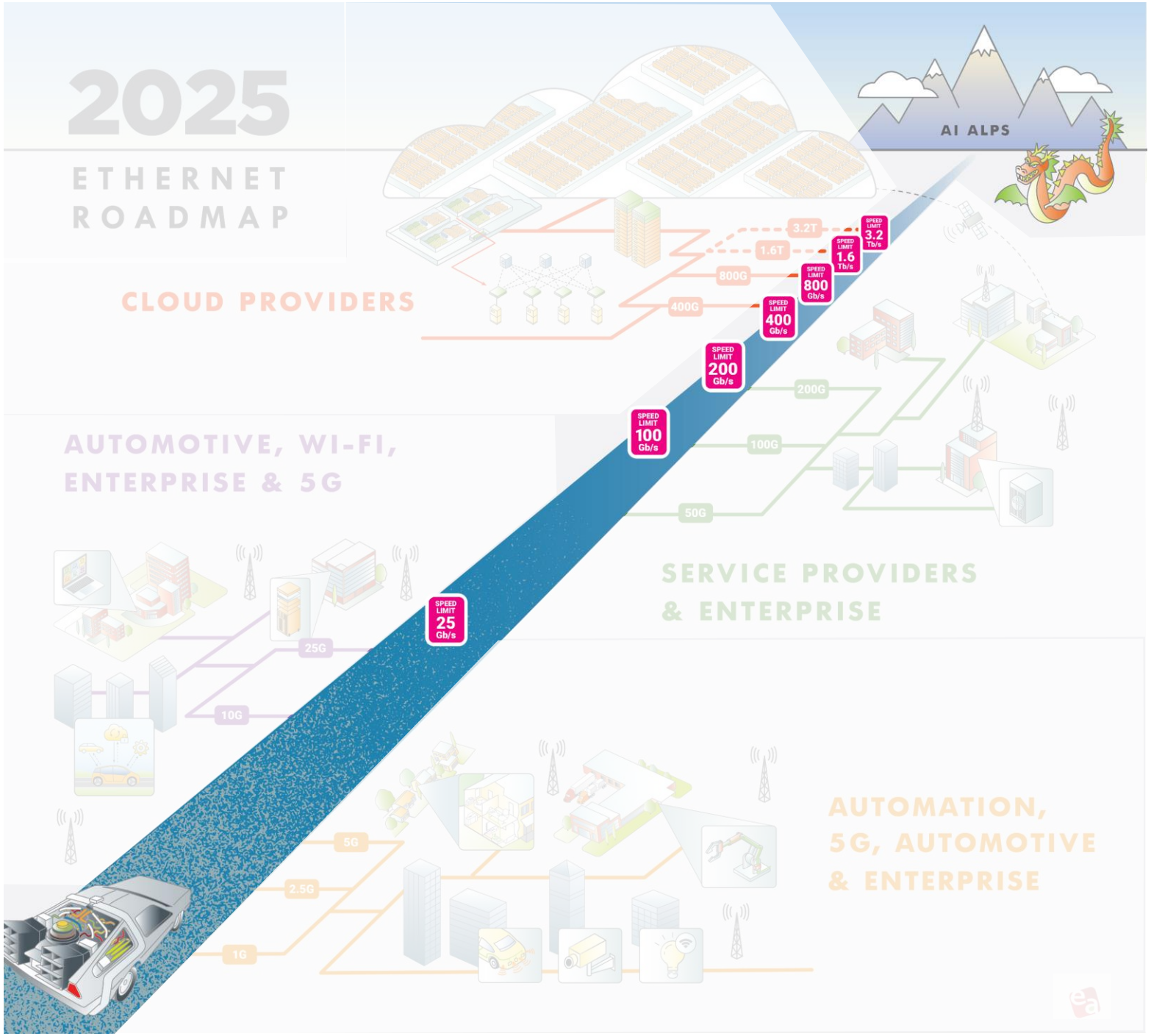


cisco

What's Up with Ethernet?

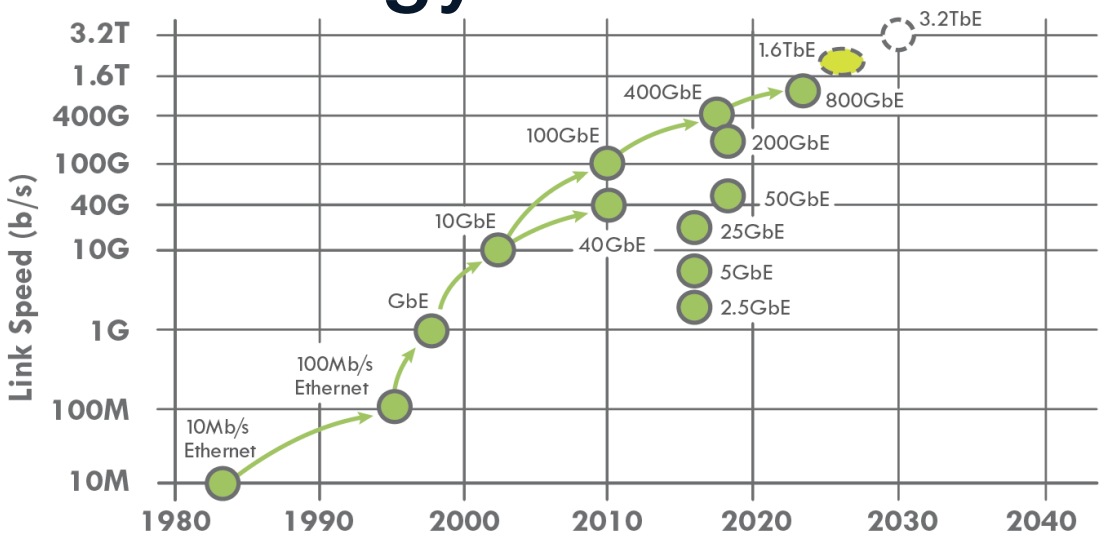
The Ethernet Alliance 2025 10th Anniversary Roadmap is here!!

ethernet alliance



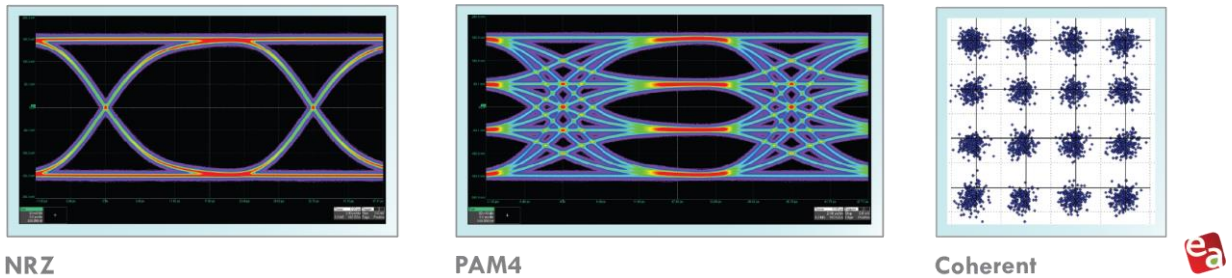
cisco

Technology Evolution



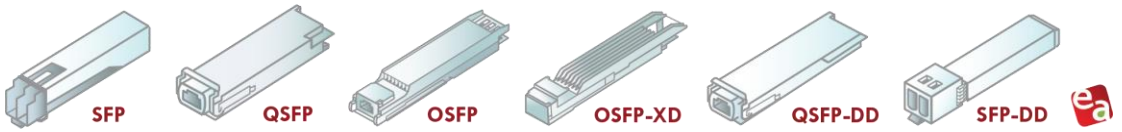
Initial Standard Completed

Ethernet Speed Speed in Development Future Speed

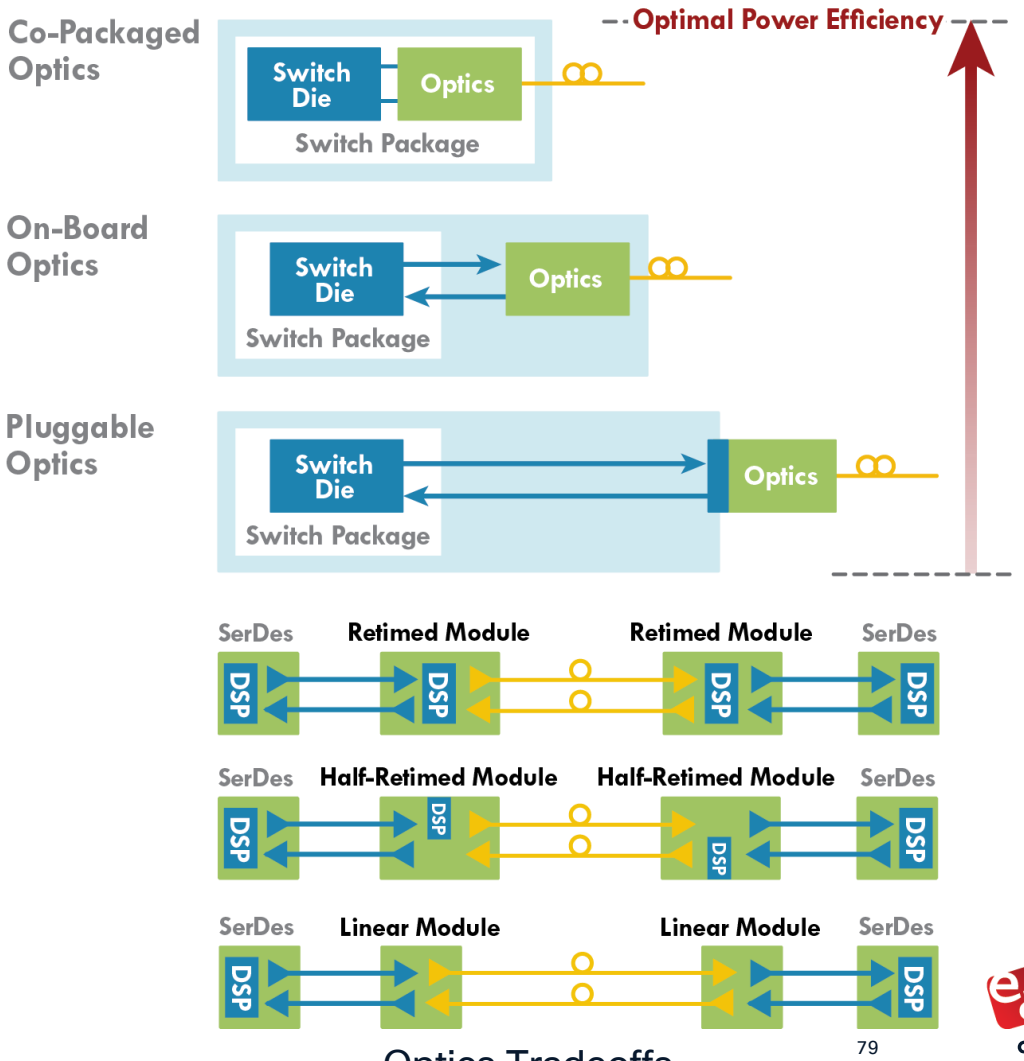


Optical Signaling

- Non-Return-to-Zero (NRZ) used for <= 25Gb/s per lane
- Four level Pulse-Amplitude Modulation (PAM4) for 50Gb/s per lane
- Coherent signaling (both in-phase and quadrature modulation) for >= 100Gb/s per lane



Pluggable module form factors



Optics Tradeoffs

From Ethernet Alliance 2025 Roadmap

Interface Magic Decoder

Image by Freeimages.com

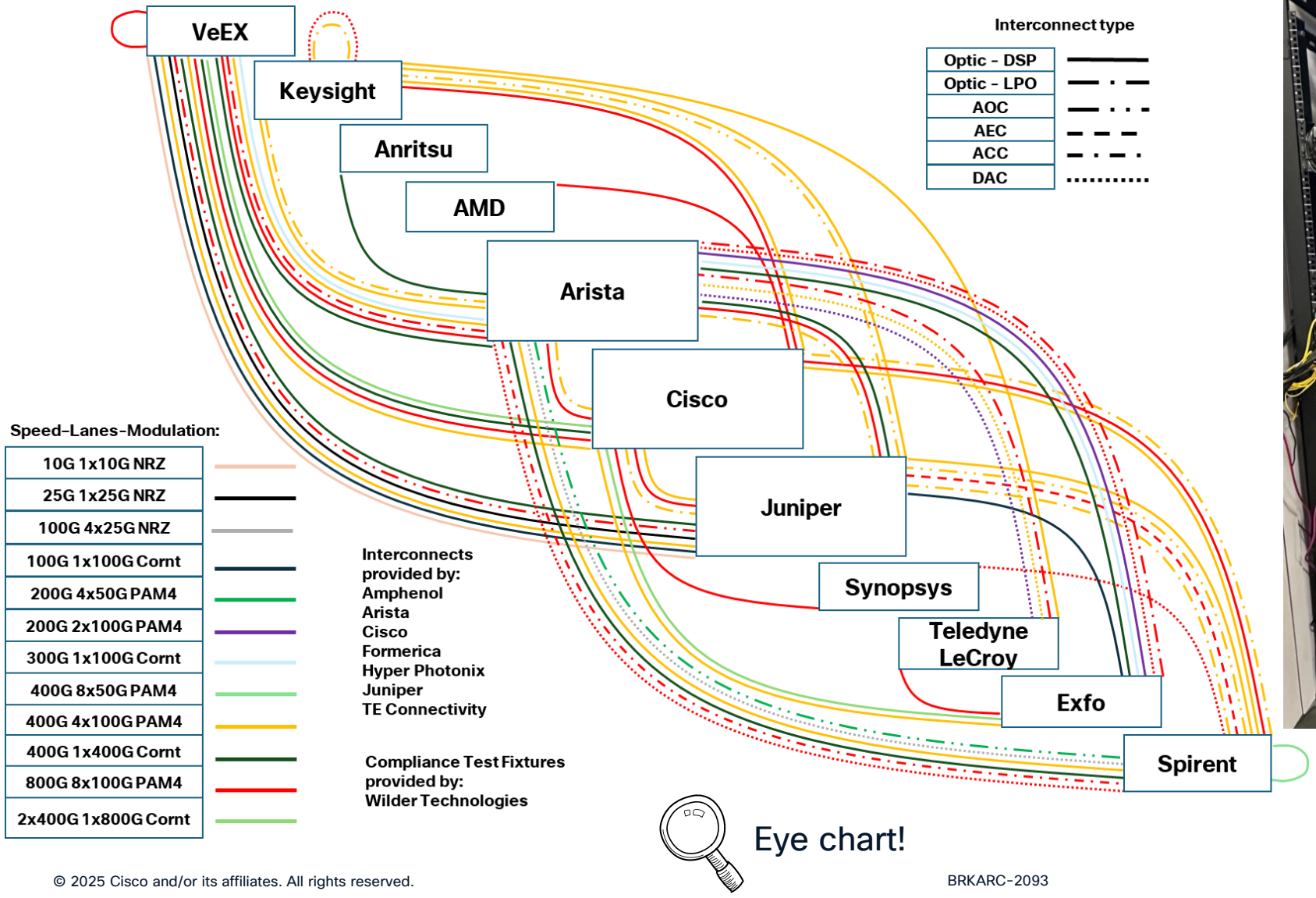
For your reference!

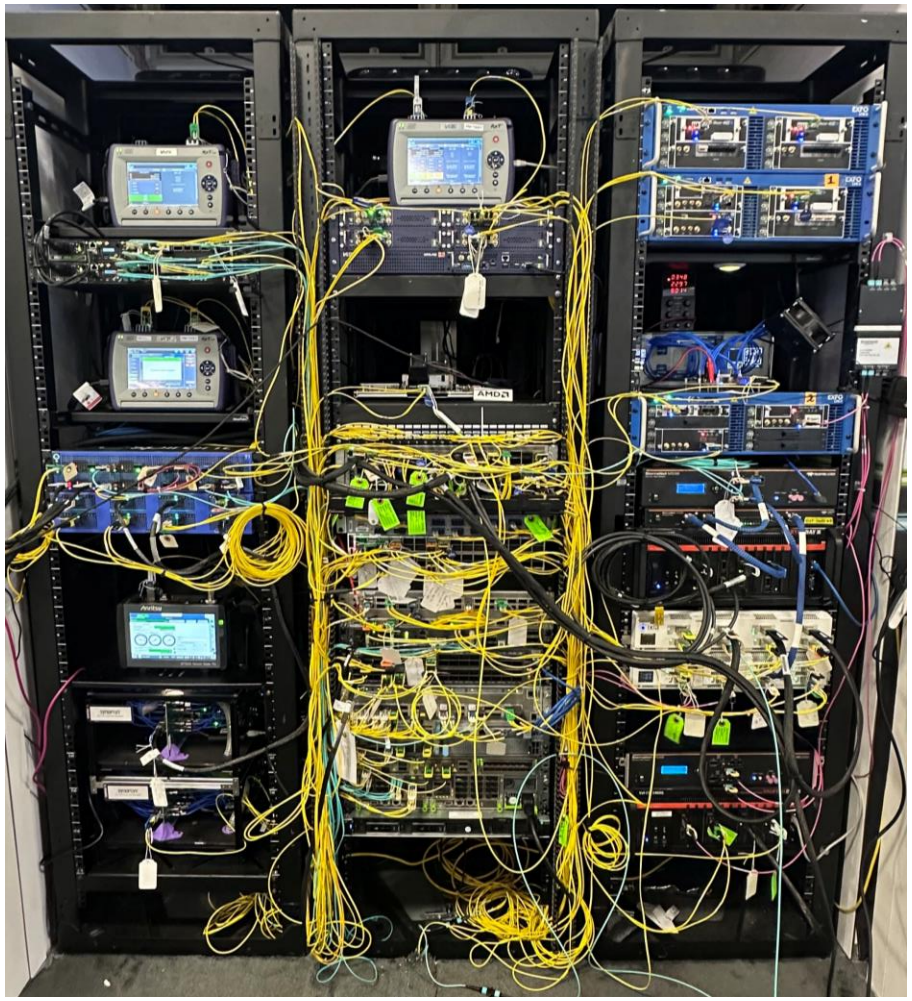
Gray Text: IEEE Standard
Red Text: In Task Force
Green Text: In Study Group
Blue Text: Non-IEEE standard but complies to IEEE electrical interfaces
Orange Text: LPO MSA specification in early stages of standardization, not compliant with IEEE electrical interfaces.

Warning! The Ethernet landscape is evolving rapidly – technologies listed here are subject to change.

	Backplane	Twinax Cable	15-40m(OT) Single Twisted Pair	>100m (OT) Single Twisted Pair	100m (IT) Twisted Pair (2/4 Pair)	MMF	500m SMF	2km SMF	10km SMF	20km SMF	30 km SMF	40km SMF	80km SMF	Electrical Interface	Pluggable Module
10BASE–	T1S		T1S	T1L	T										
100BASE–			T1	T1L	T										
1000BASE–			T1		T										SFP
2.5GBASE–	KX		T1		T										SFP
5GBASE–	KR		T1		T										SFP
10GBASE–			T1		T	SR			LR BR10-D/U	BR20-D/U		ER BR40-D/U			SFP
25GBASE–	KR1 KR	CR1 CR/CR-S	T1		T (30m)	SR			LR EPON BR10-D/U	EPON BR20-D/U		ER BR40-D/U		25GAUI	SFP
40GBASE–	KR4	CR4			T (30m)	SR4/eSR4	PSM4	FR	LR4			ER4		XLAUI XLPPI	QSFP
50GBASE–	KR2 KR	CR2 CR				SR		FR	LR EPON BR10-D/U	EPON BR20-D/U		ER BR40-D/U		LAUI-2/50GAUI-2 50GAUI-1	SFP/QSFP
100GBASE–	KR4 KR2 KR1	CR10 CR4 CR2 CR1				SR10 SR4 SR2 VR1/SR1	PSM4 DR	CWDM4 FR1	LR4/ 4WDM-10 LR1	4WDM-20 LR1-20	ER1-30	ER4/4WDM-40 ER1-40	ZR	CAUI-10/CPPI CAUI-4/100GAUI-4 100GAUI-2 100GAUI-1	SFP/SFP-DD QSFP/QSFP-DD OSFP
100G–							DR1-LPO							LEI-100G-PAM4-1	
200GBASE–	KR4 KR2 KR1	CR4 CR2 CR1				SR4 VR2/SR2	DR4 DR1	FR4 DR1-2	LR4			ER4		200GAUI-4 200GAUI-2 200GAUI-1	QSFP/QSFP-DD SFP-DD
200G–							DR2-LPO							LEI-200G-PAM4-2	
400GBASE–	KR4 KR2	CR4 CR2				SR16 SR8/SR4.2 VR4/SR4	DR4 DR2	FR8 FR4 DR4-2 DR2-2	LR8 LR4-6/LR4-10		ER4-30	ER8	400ZR	400GAUI-16 400GAUI-8 400GAUI-4 400GAUI-2	QSFP/QSFP-DD OSFP
400G–							DR4-LPO							LEI-400G-PAM4-4	
800GBASE–	ETC-KR8/KR8 KR4	ETC-CR8/CR8 CR4				VR8/SR8 VR4.2/SR4.2	FR4-500 DR8 DR4	FR4 DR8-2 DR4-2	LR4 LR1	ER1-20		ER1	800ZR-A 800ZR-B 800ZR-C	800GAUI-8 800GAUI-4	QSFP-DD OSFP/OSFP-XD
800G–							DR8-LPO							LEI-800G-PAM4-8	
1.6TBASE–	KR8	CR8				VR8.2/ SR8.2	DR8	DR8-2						1.6TAUI-16 1.6TAUI-8	QSFP-DD OSFP/OSFP-XD

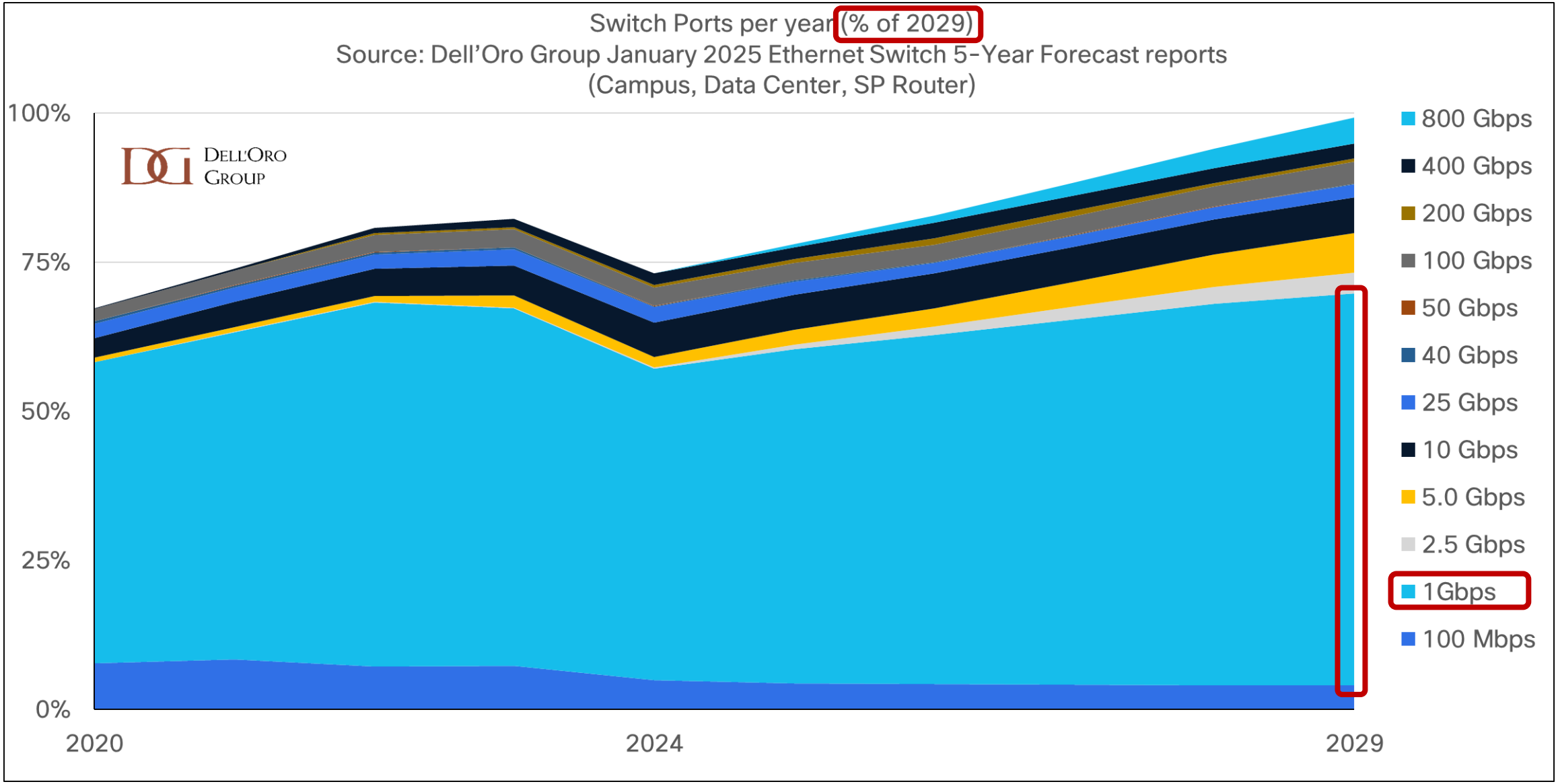
Ethernet Alliance Interop Demo @ OFC2025



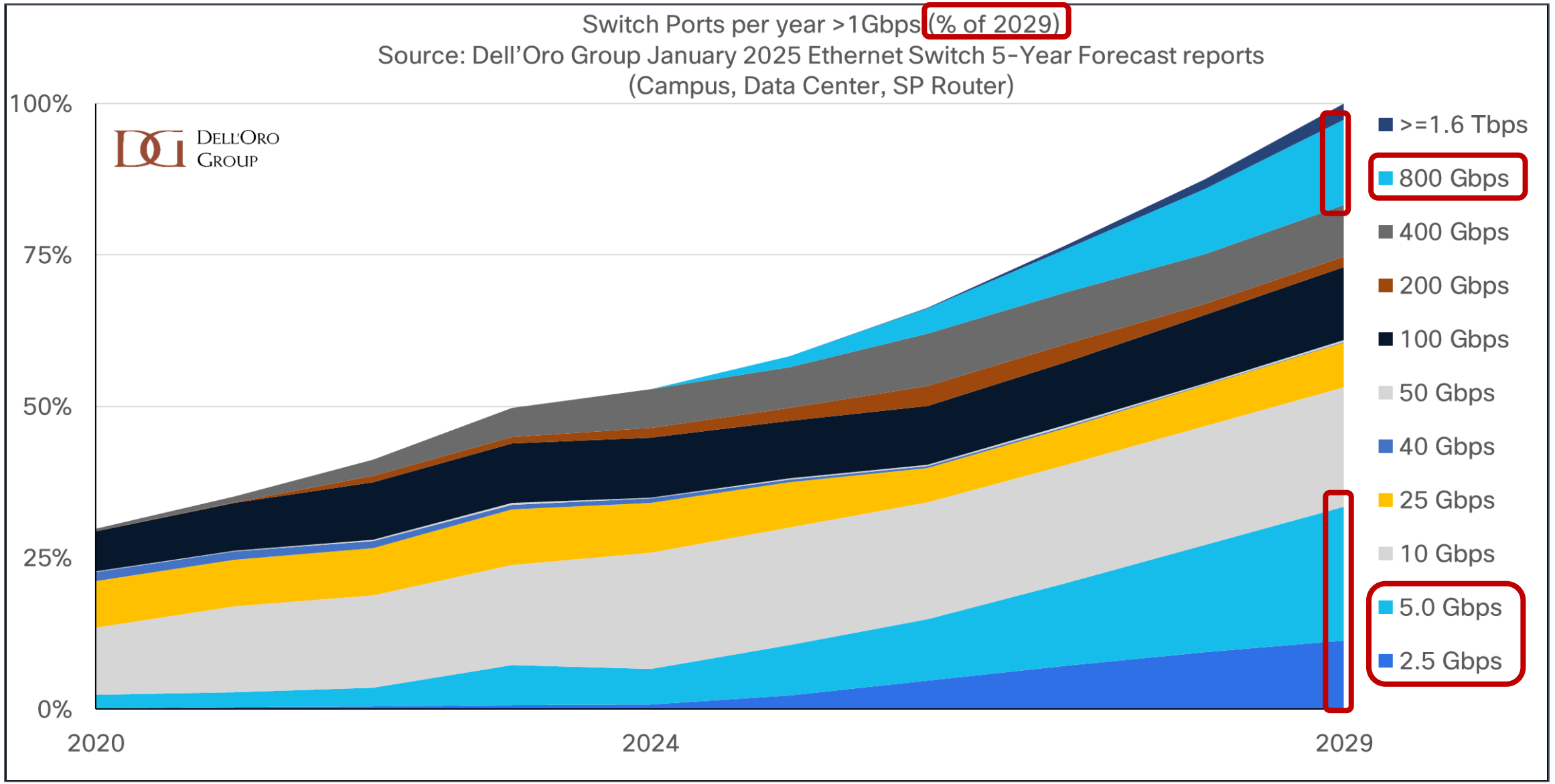


Eye chart!

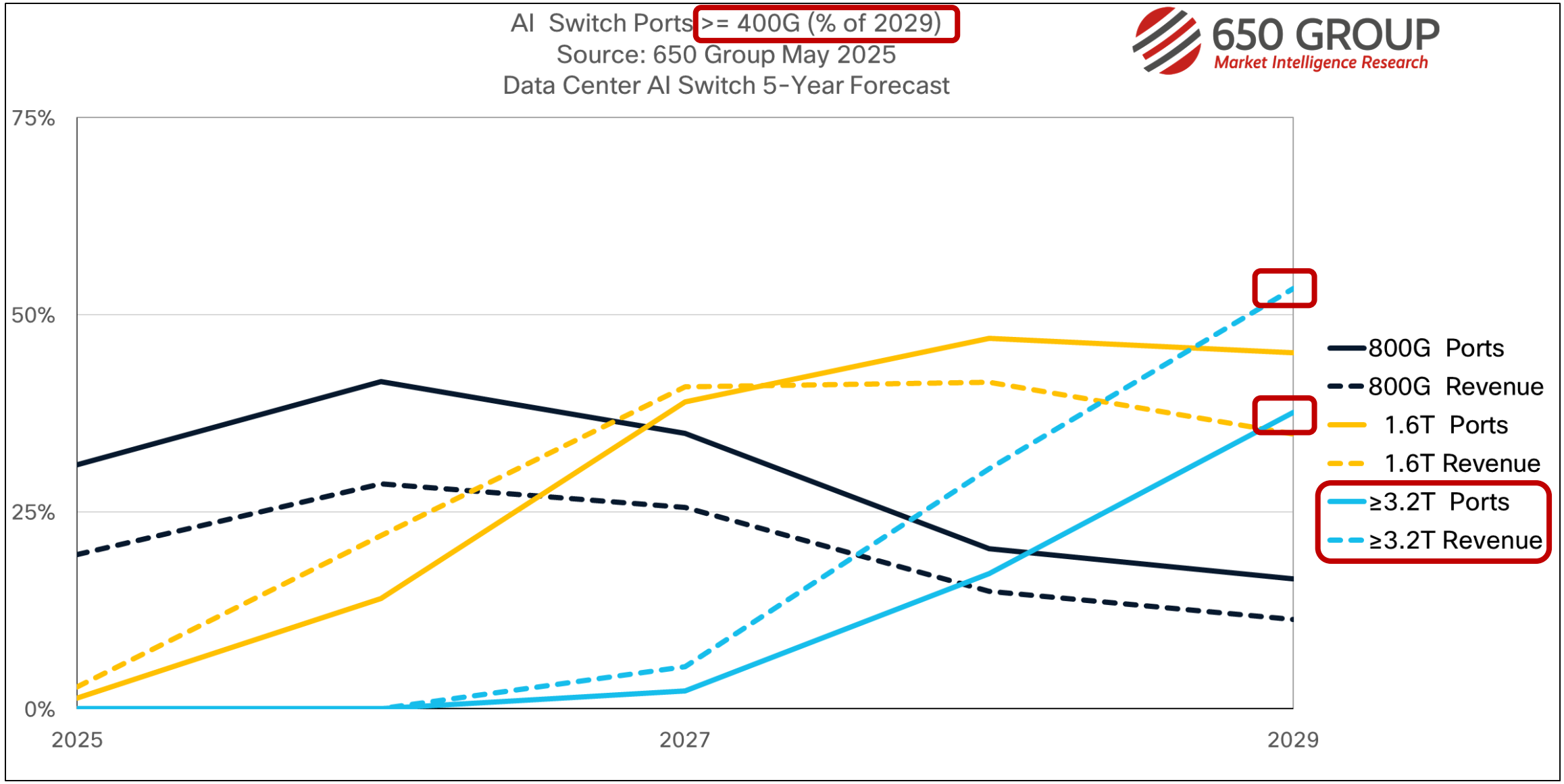
Speed Trends – Switch Ports per Year



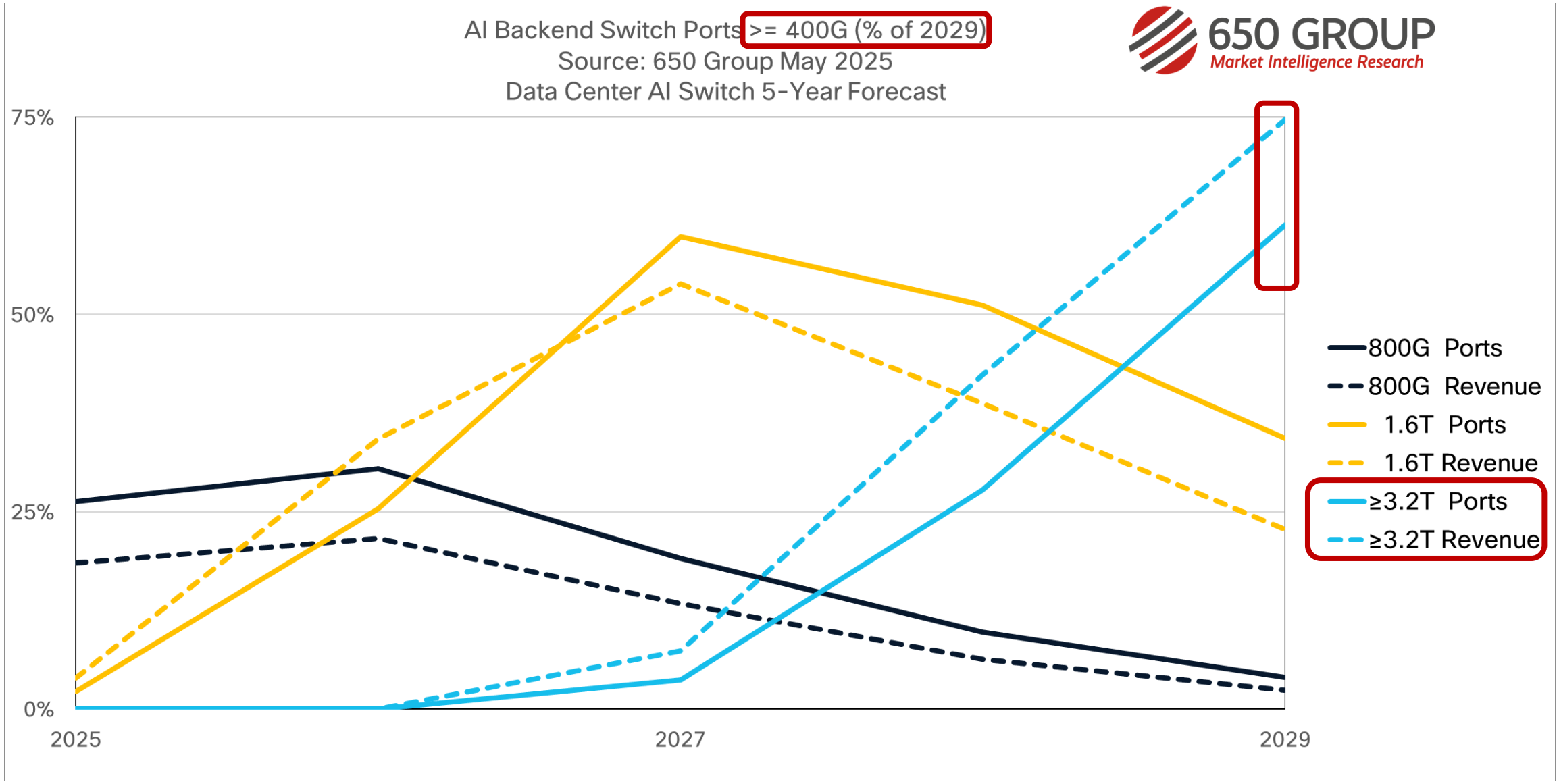
Speed Trends – Switch Ports per Year (> 1Gbps)



Ethernet Speed Trends – AI Network Switch Ports

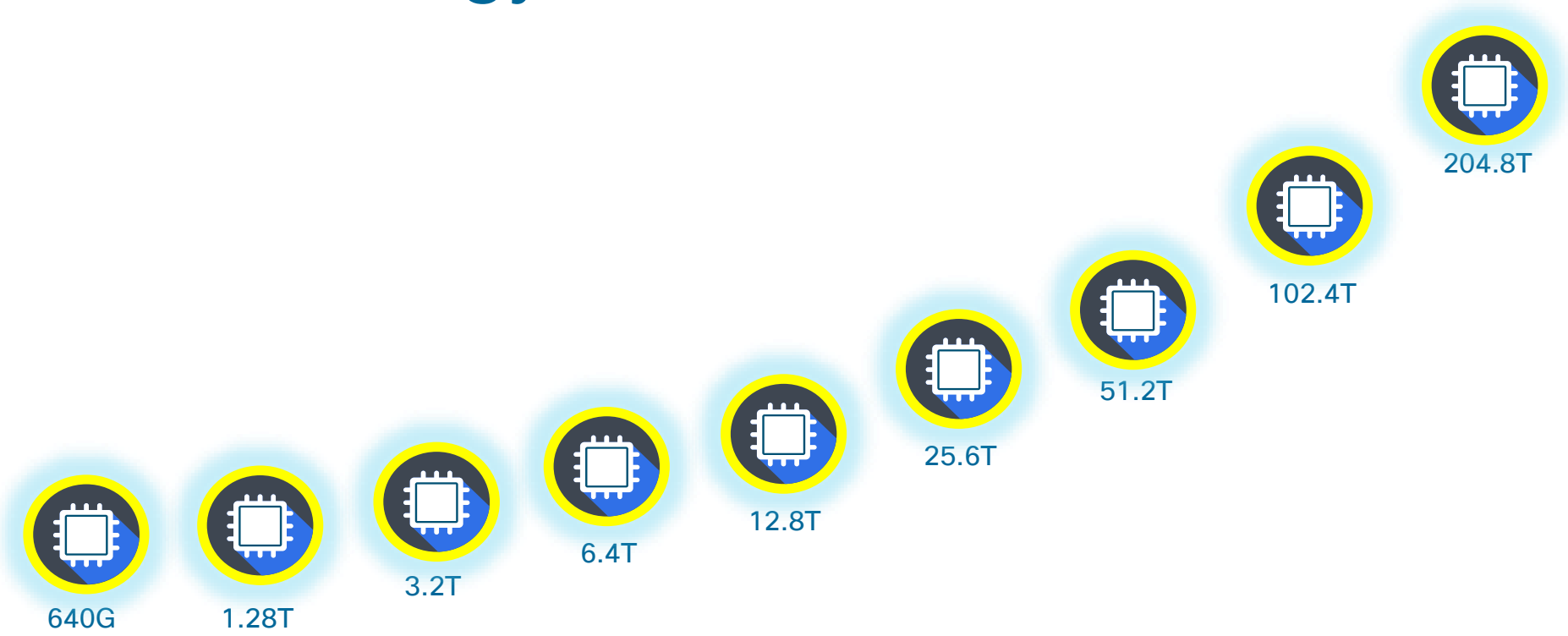


Ethernet Speed Trends – AI Backend Network Switch Ports



Review: "BRKOPT-2699 400G, 800G, and Terabit Pluggable Optics: What you need to know"

Ethernet's technology must match market needs



← 204.8T is 16x the bandwidth of 12.8T

320x increase in switching bandwidth over 16 years

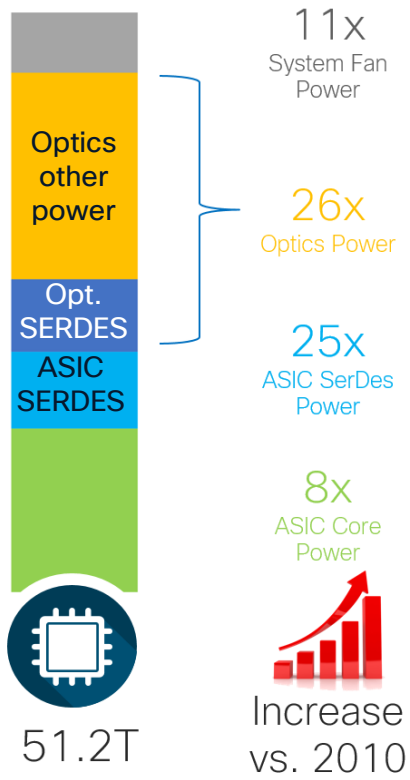
Switching Capacity	640G	1.28T	3.2T	6.4T	12.8T	25.6T	51.2T	102.4T	204.8T	
Year	2010	2012	2014	2016	2018	2020	2022	~2025	~2027?	in 17 years ...
ASIC # SerDes	64	128		256		512			1024	16x
Serdes Gbps	10G NRZ		25G NRZ		50G NRZ		100G PAM4	200G PAM4		20x
Port Gbps	40G		100G		400G		800G	1.6T		40x
Ports/Router	16	32		64	32	64	64	64	128	

ASIC density continues to redefine how products are built.
Gates & GHz. SerDes & Interconnect. Optics & wavelengths.

Credit: adapted from https://blogs.cisco.com/sp/co-packaged-optics-and-an-open-ecosystem

Review: "BRKOPT-2699 400G, 800G, and Terabit Pluggable Optics: What you need to know"

Interconnect and power



11x
System Fan Power

26x
Optics Power

25x
ASIC SerDes Power

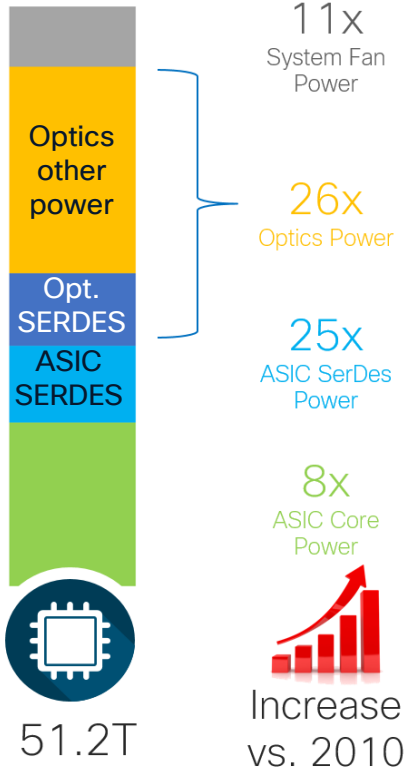
8x
ASIC Core Power

Increase vs. 2010

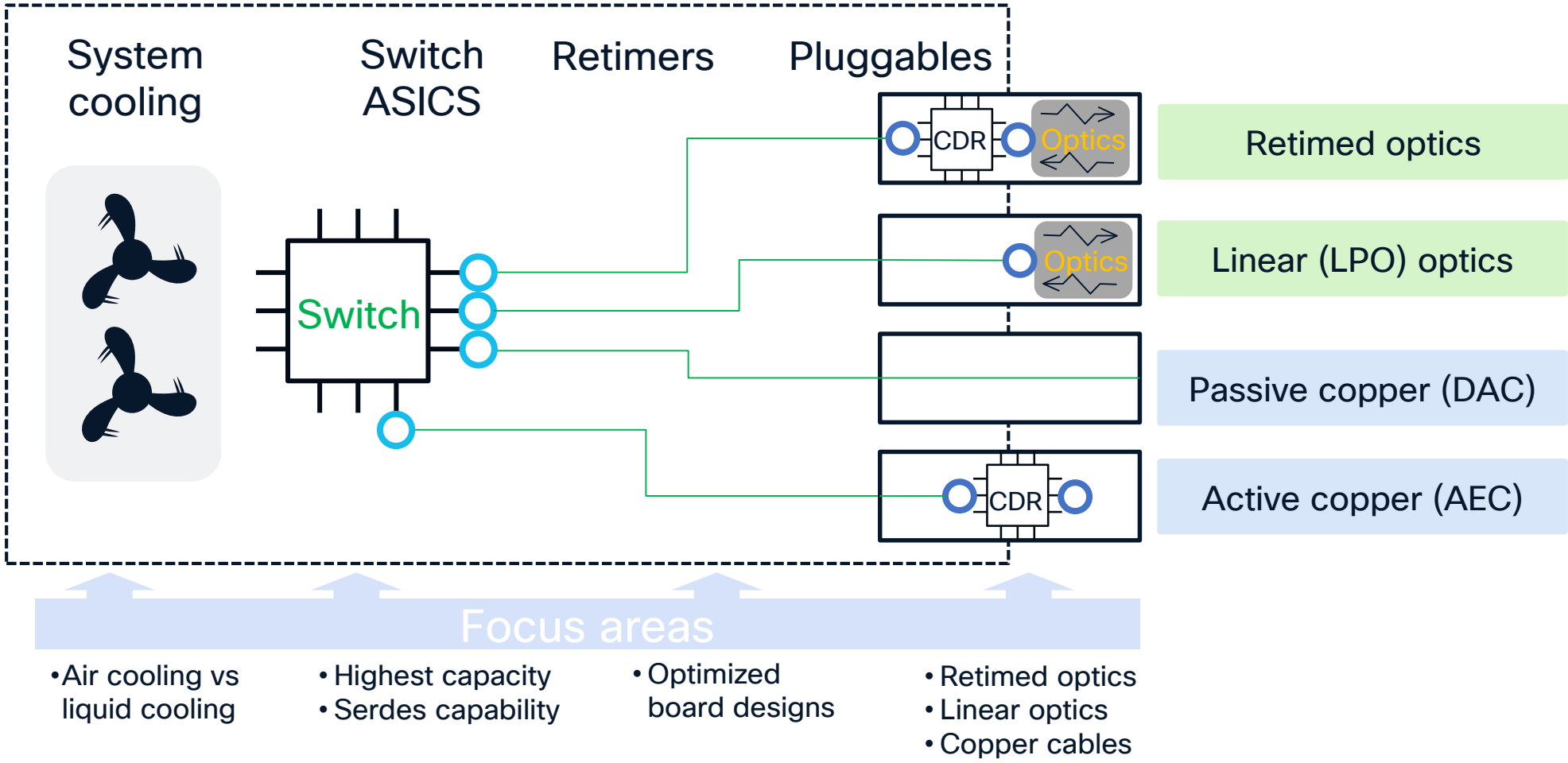
Power increase for 80x bandwidth increase

Review: "BRKOPT-2699 400G, 800G, and Terabit Pluggable Optics: What you need to know"

Interconnect and power



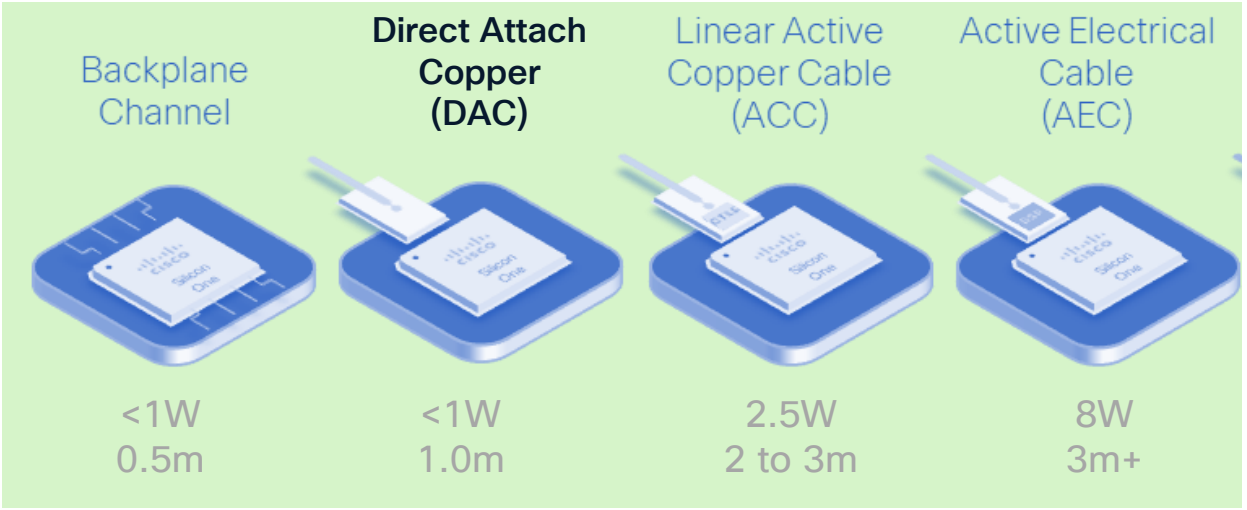
Power increase for 80x bandwidth increase



High-Density Interconnect options: Power vs. Reach Tradeoff

Review: "BRKOPT-2699 400G, 800G, and Terabit Pluggable Optics: What you need to know"

Power and Reach @ 800G

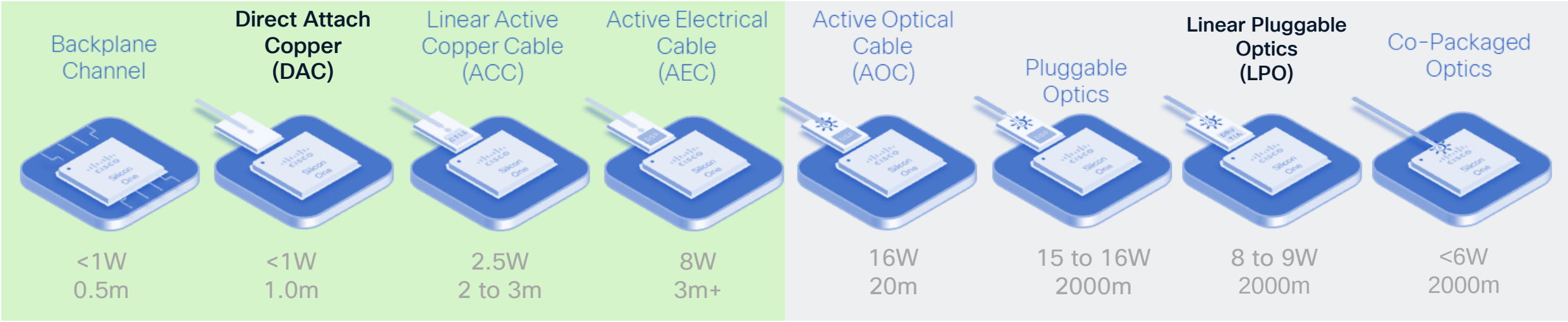


Mostly within rack

High-Density Interconnect options: Power vs. Reach Tradeoff

Review: "BRKOPT-2699 400G, 800G, and Terabit Pluggable Optics: What you need to know"

Power and Reach @ 800G



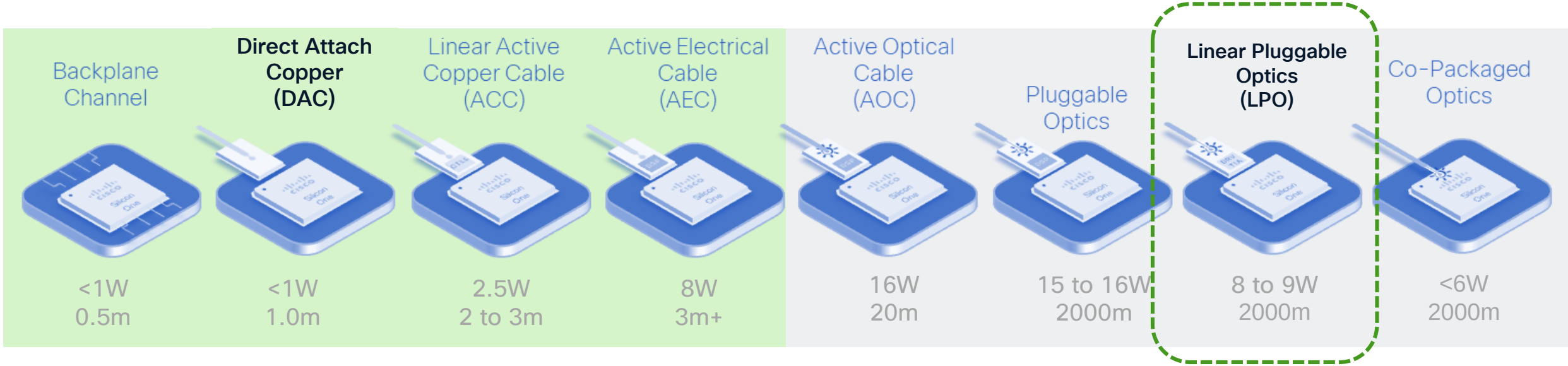
Mostly within rack

Mostly between racks

High-Density Interconnect options: Power vs. Reach Tradeoff

Review: "BRKOPT-2699 400G, 800G, and Terabit Pluggable Optics: What you need to know"

Power and Reach @ 800G



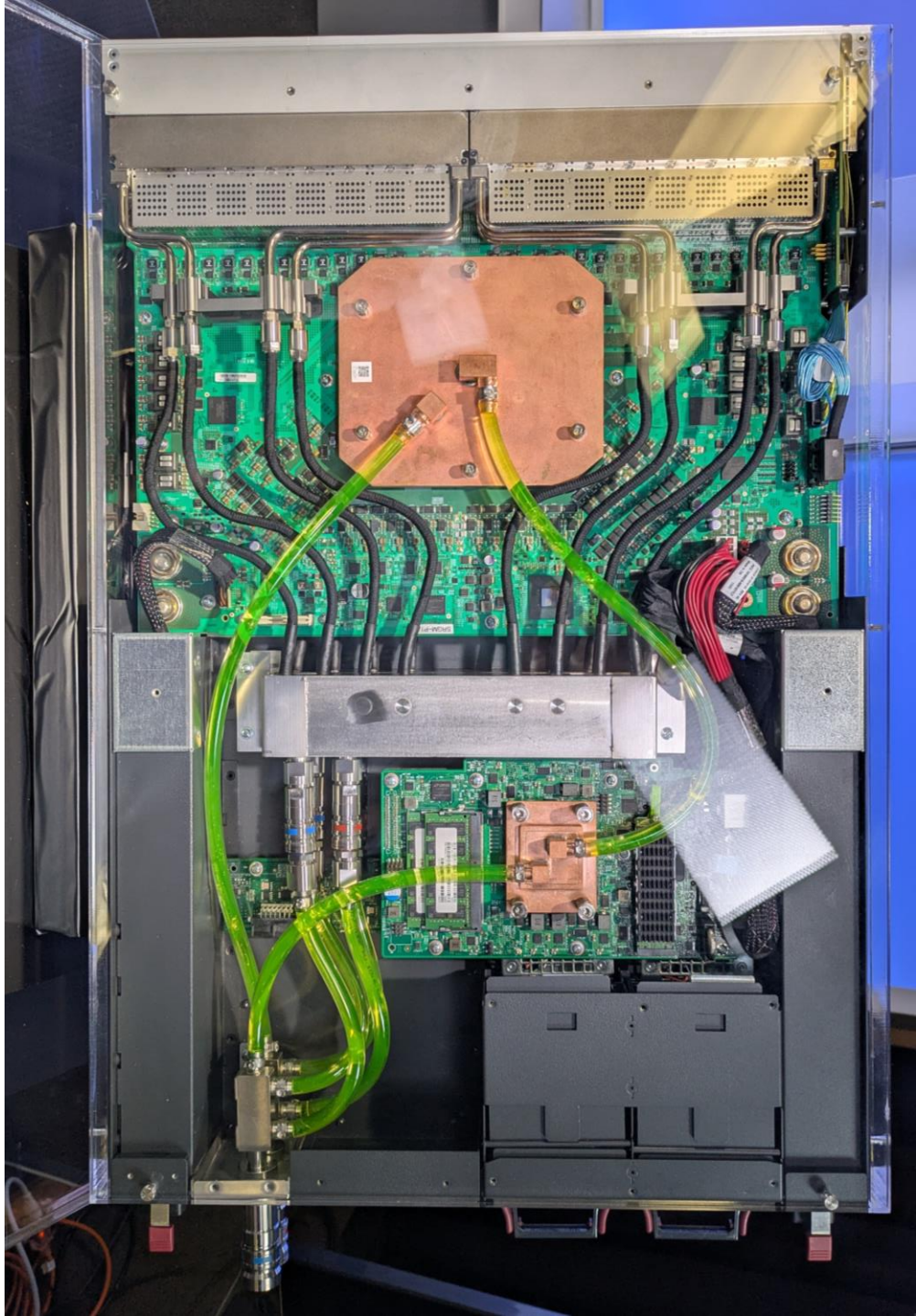
LPO – Linear Pluggable Optics
No DSP

LRO – Linear Receive Optics
DSP on transmit side

Linear Optics
- Excellent power vs reach
- Interoperability not there yet.

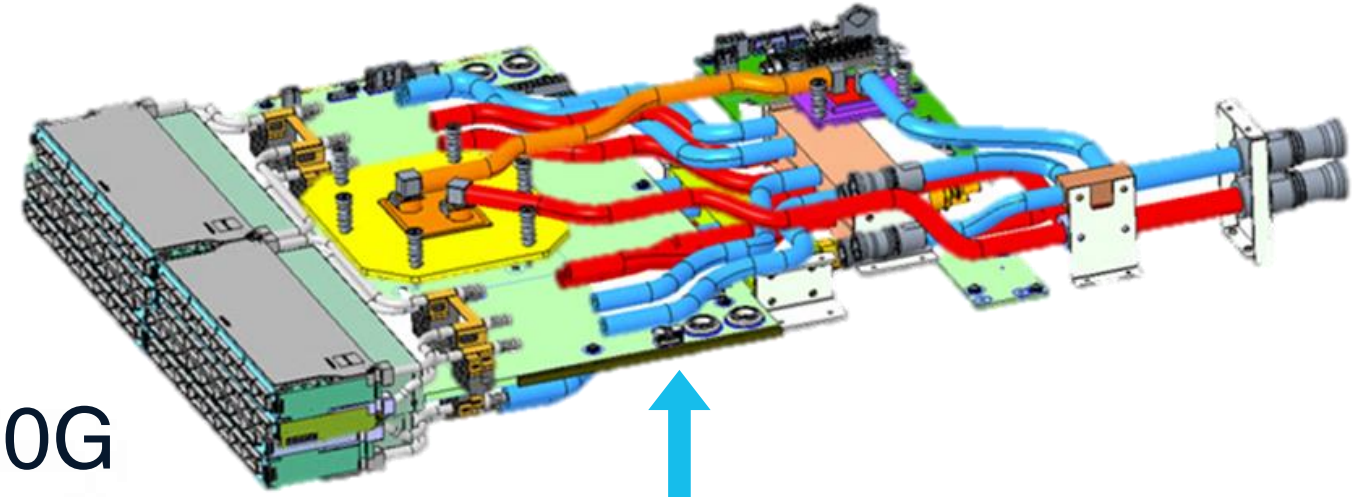
Cold Plate Liquid Cooling

- Power density/cooling is becoming the limiting constraint
- NVIDIA GB200 NVL72 is ***~1.2kW per GPU*** and ~***120kW per rack***[1]
- Microsoft and Meta Mount Diablo design uses ***400Vdc***[2] into the rack
- Google is planning for racks up to ***1MW*** [3]
- Power savings
  - ~10-15% from system fans
  - ~60% facility power (chillers etc)
- Improves Power Usage Effectiveness(PUE)[4] ~20%



1. NVIDIA GB200 NVL72: https://training.continuumlabs.ai/infrastructure/servers-and-chips/nvidia-gb200-nvl72
2. Mount Diablo: https://www.datacenterdynamics.com/en/news/microsoft-and-meta-reveal-open-ai-rack-design-with-separate-power-and-compute-cabinets/
3. Google 1MW rack plans: https://cloud.google.com/blog/topics/systems/enabling-1-mw-it-racks-and-liquid-cooling-at-ocp-emea-summit
4. Power usage effectiveness: https://en.wikipedia.org/wiki/Power_usage_effectiveness

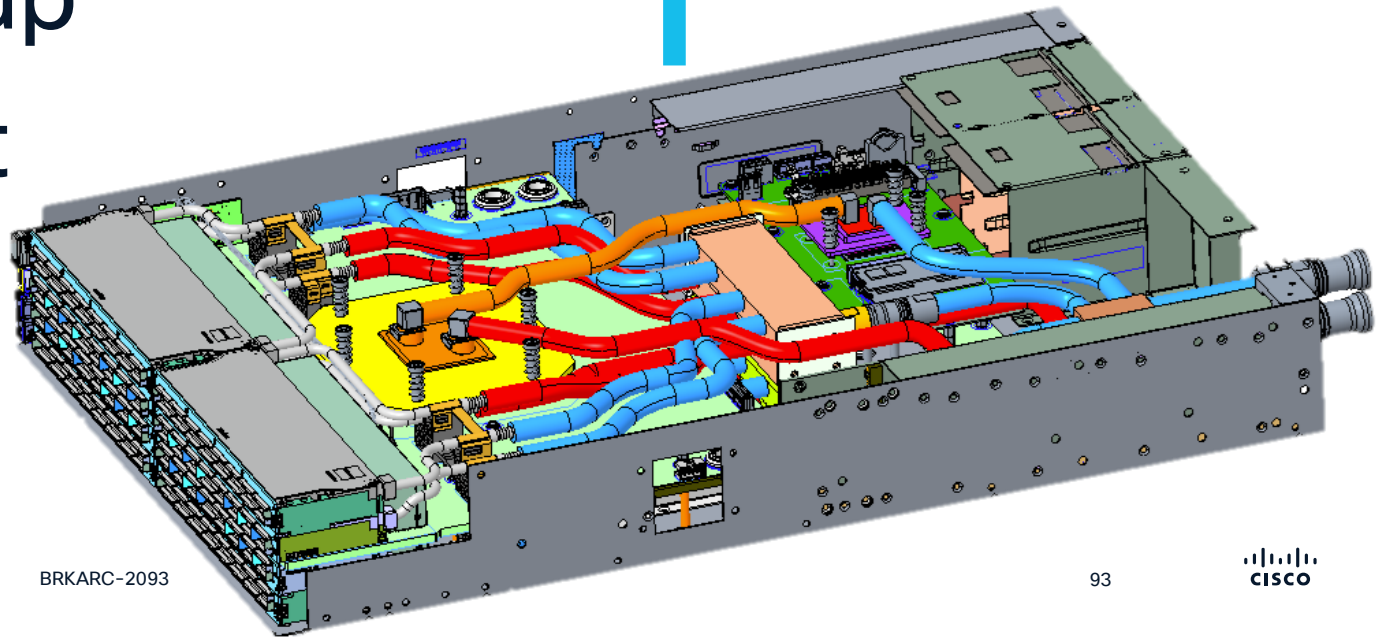
Liquid Cooling 51.2T Switch Technology Demonstration



Liquid cooled components:

ASIC, CPU, 64 x OSFP 800G

Liquid Cooling removes up to ***80%*** of system heat



25.6T Co-Packaged Optics(CPO) at OFC 2023

Retimed optics

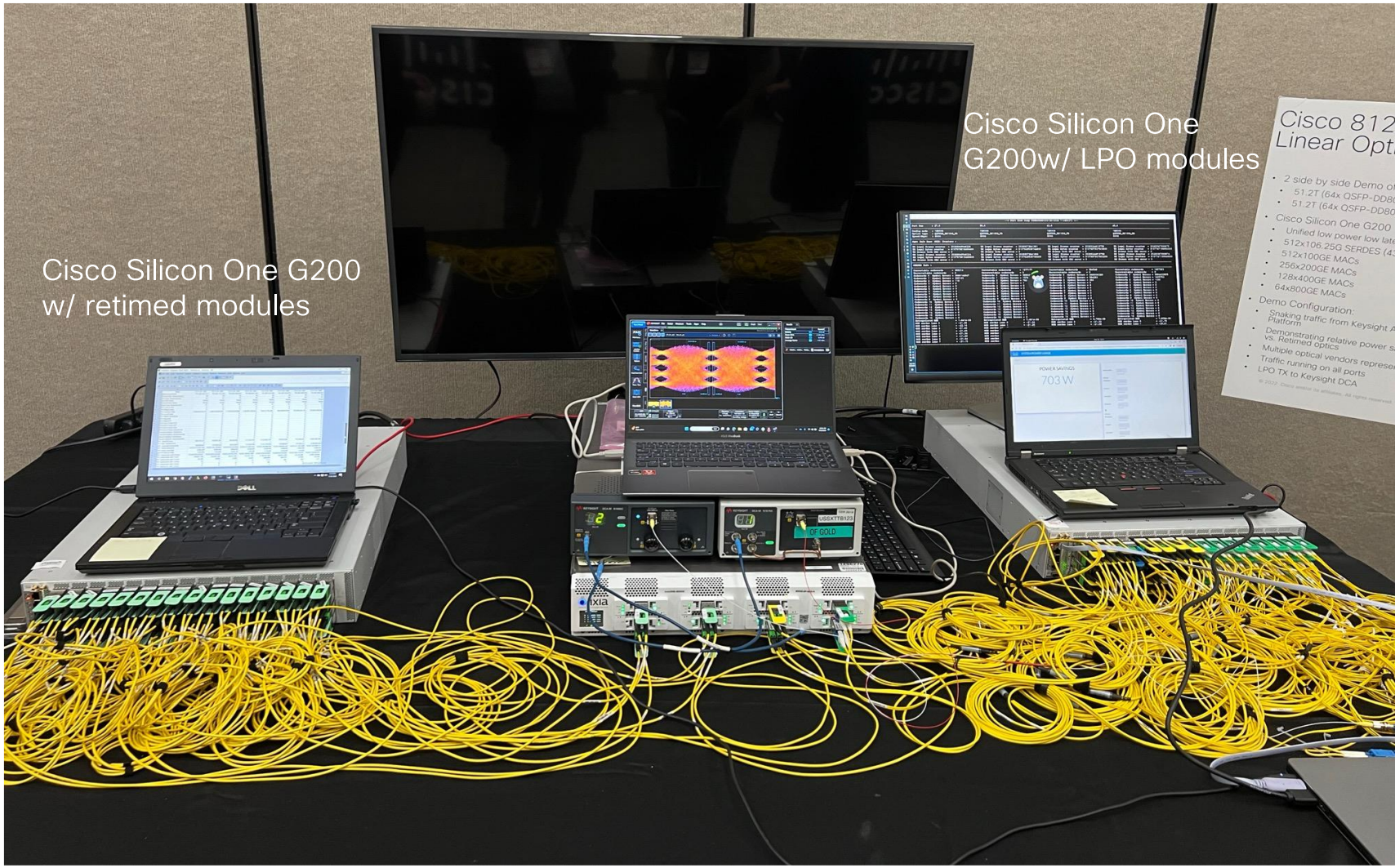
CPO

CPO power reduction: ~270W

Review: "BRKOPT-2699 400G, 800G, and Terabit Pluggable Optics: What you need to know"

51.2T Linear Pluggable Optics(LPO) at OFC 2024

LPO power reduction: ~700W



Fault Managed Power: Touch Safe High Voltage DC

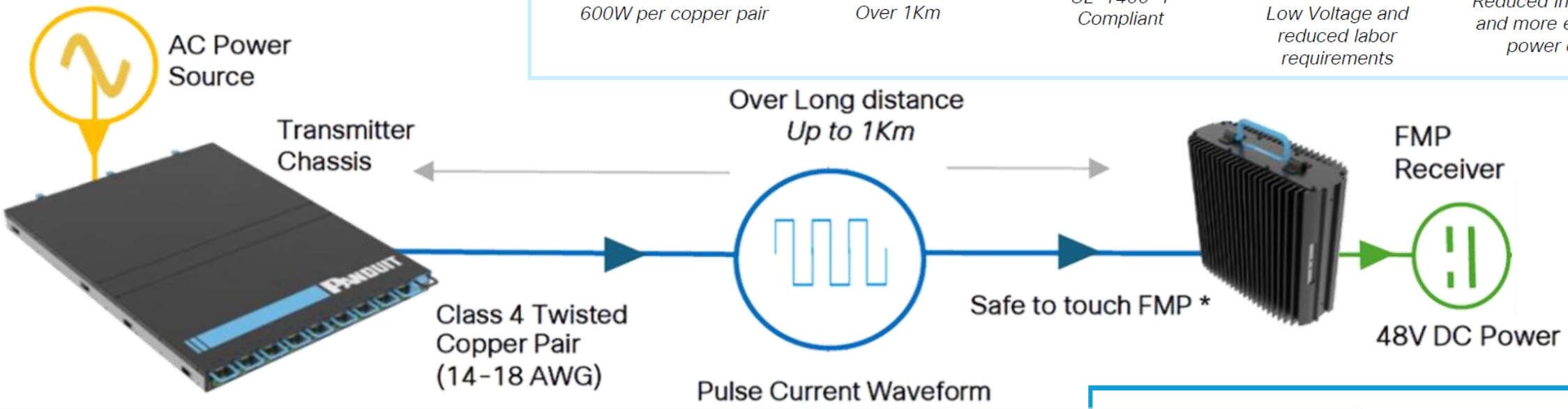
Significant Power
600W per copper pair

Long Distance
Over 1Km

Safety
UL-1400-1 Compliant

Speed to Deploy
Low Voltage and reduced labor requirements

Sustainable
Reduced Infrastructure and more efficient DC power delivery



5-30% increase in energy savings in buildings with widespread adoption of DC power
US Dept Energy

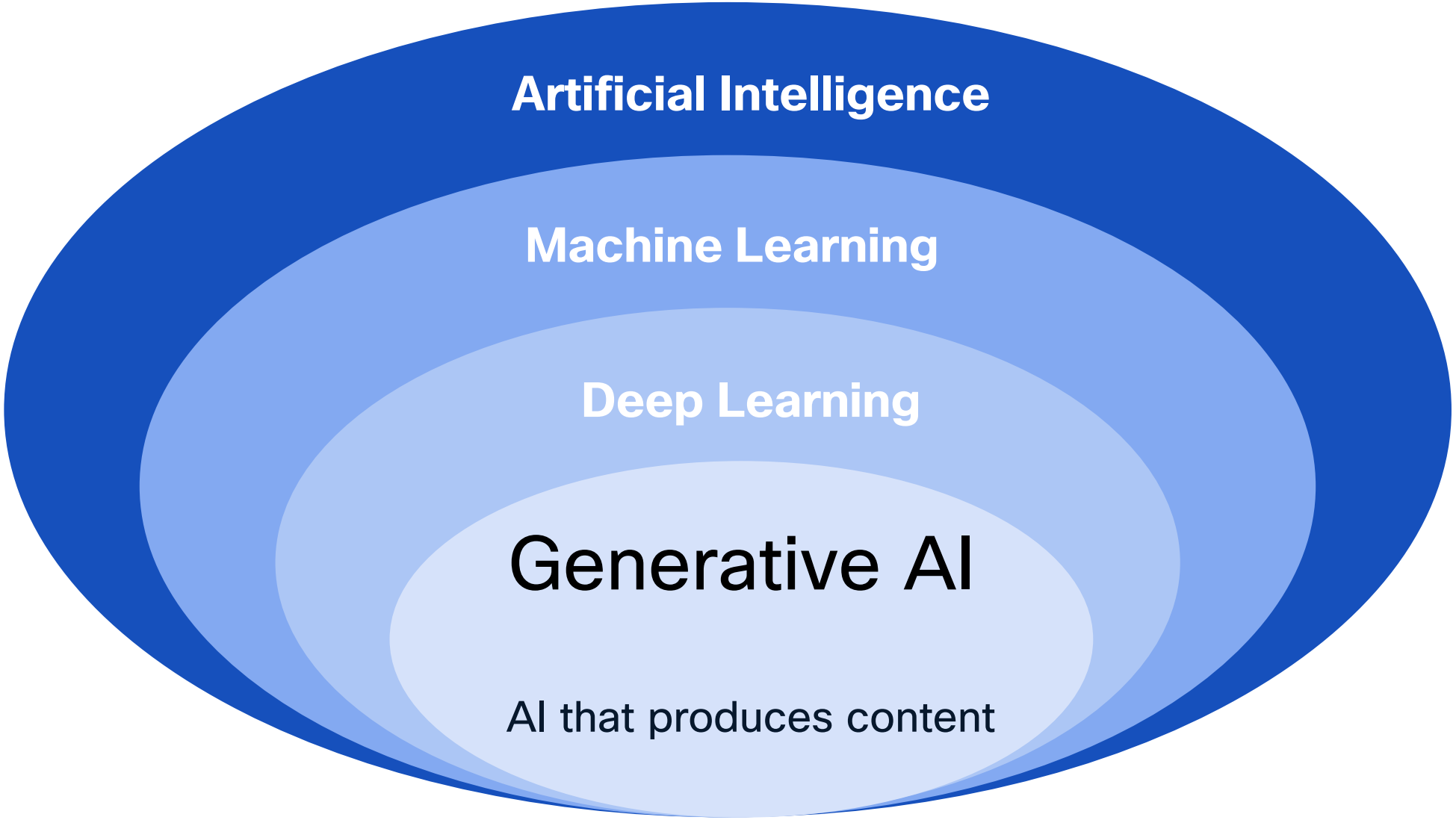
10-20% increase in energy efficiency by eliminating AC to DC conversion
www.energy.gov

https://www.panduit.com/en/products/featured-products/panduit-fault-managed-power-system.html
https://www.cisco.com/c/en/us/td/docs/engineering_alliances/panduit_fmps_and_cisco_implementation_guide.html

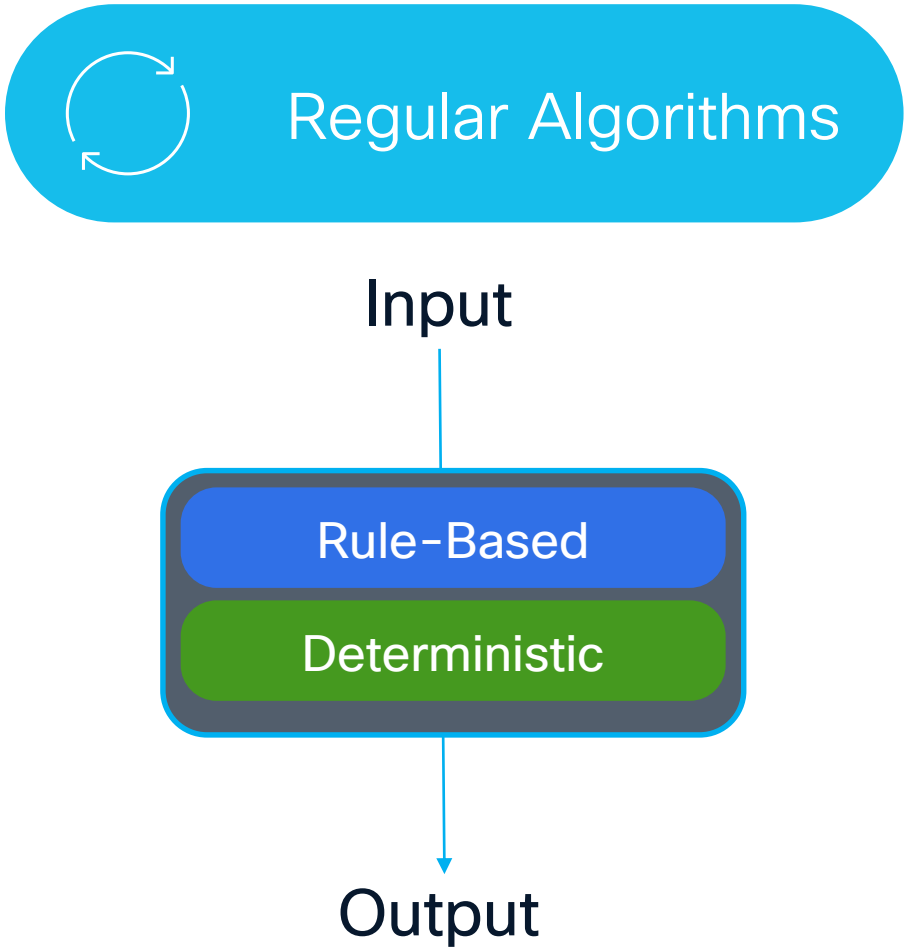
Cisco Networking

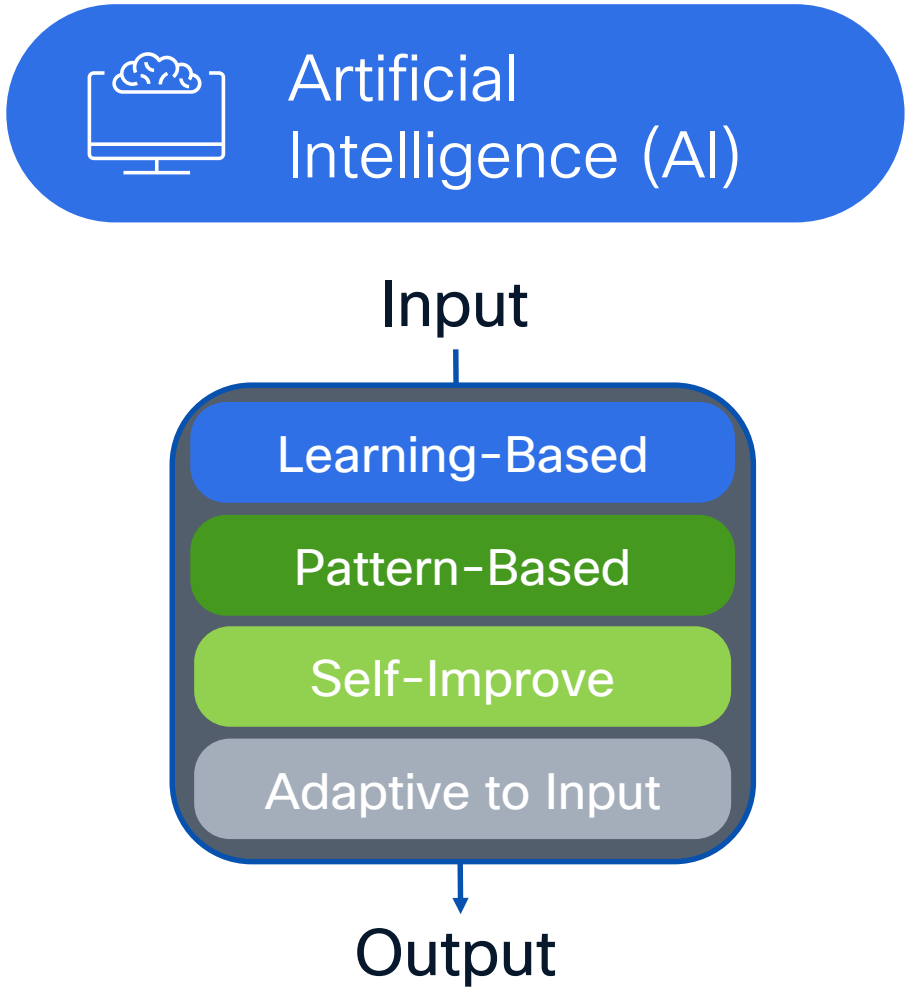
The AI/ML-based Future

The Breakdown of Artificial Intelligence

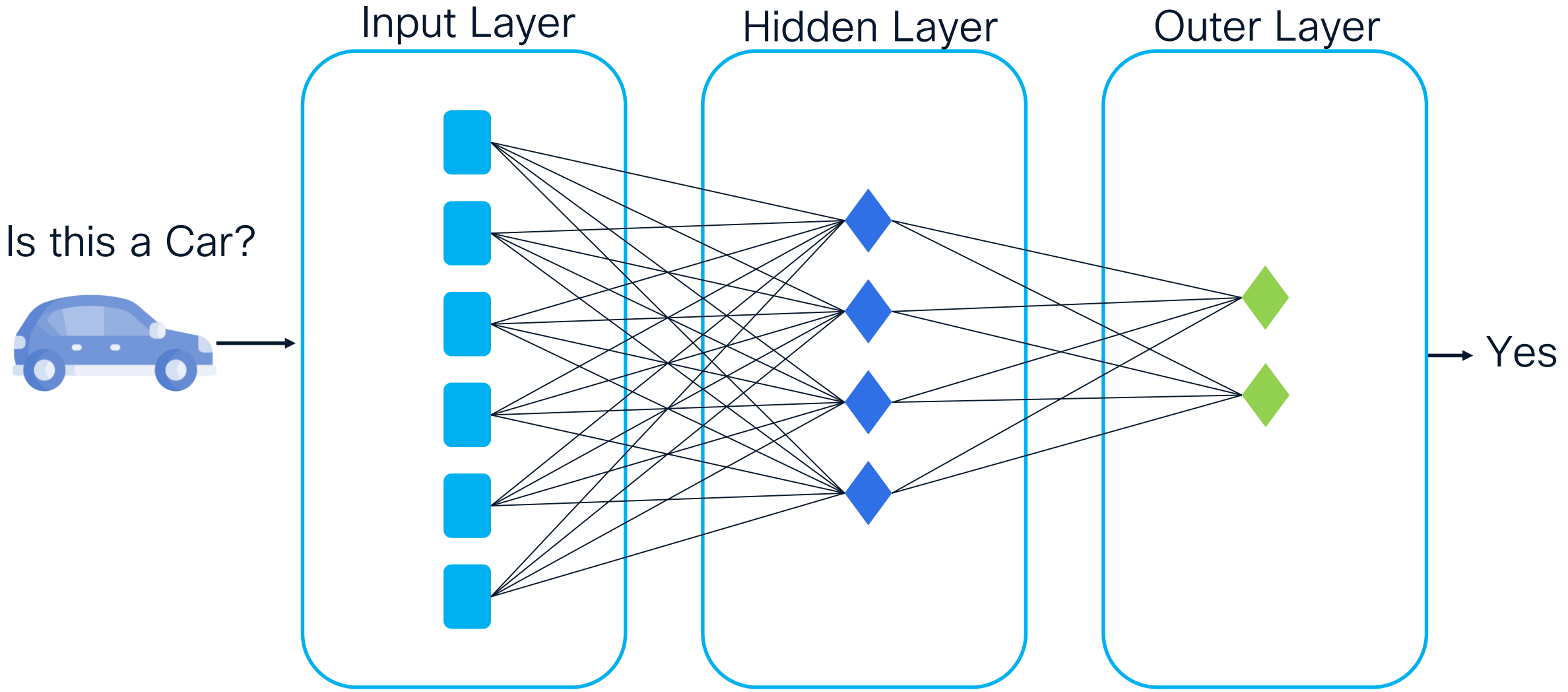


How Is AI Different From Regular Algorithms?





Neural Networks – Identify Patterns with Deep Learning
Divide and conquer large amounts of complex data

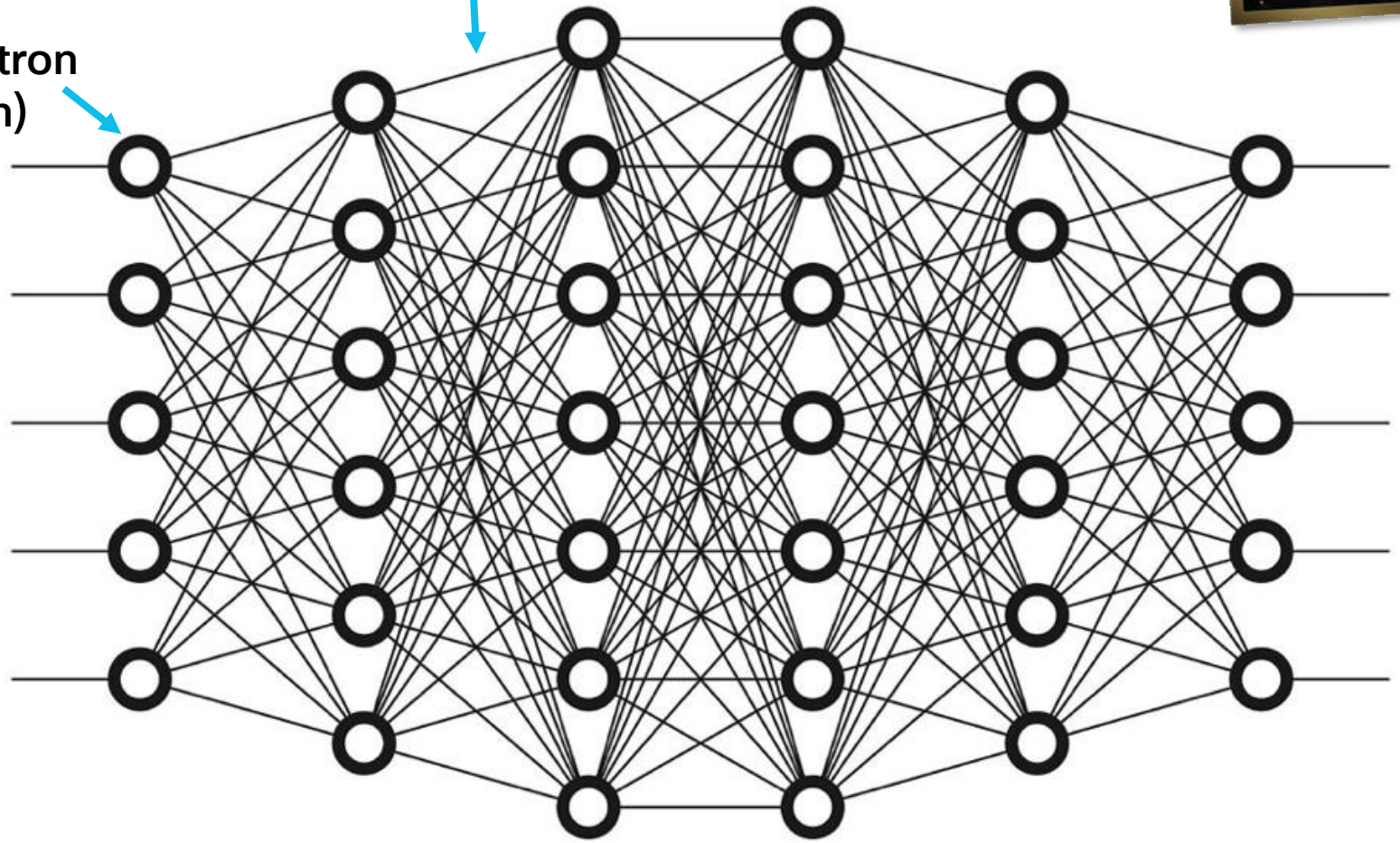


Why is this happening **now?**

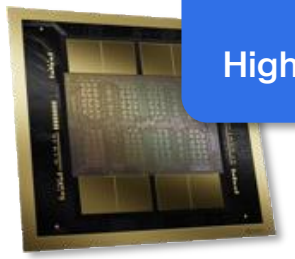
Scale

Parameter
(Synapses)

Perceptron
(Neuron)



NVIDIA
Blackwell



Advances in Silicon –
High-density, High-performance GPUs

NVIDIA Flagship Accelerator Specification Comparison

	B200	H100	A100 (80GB)
FP32 CUDA Cores	A Whole Lot	16896	6912
Tensor Cores	As Many As Possible	528	432
Boost Clock	To The Moon	1.98GHz	1.41GHz
Memory Clock	8Gbps HBM3E	5.23Gbps HBM3	3.2Gbps HBM2e
Memory Bus Width	2x 4096-bit	5120-bit	5120-bit
Memory Bandwidth	8TB/sec	3.35TB/sec	2TB/sec
VRAM	192GB (2x 96GB)	80GB	80GB

Geoffrey Hinton - the "Godfather" of Deep Learning

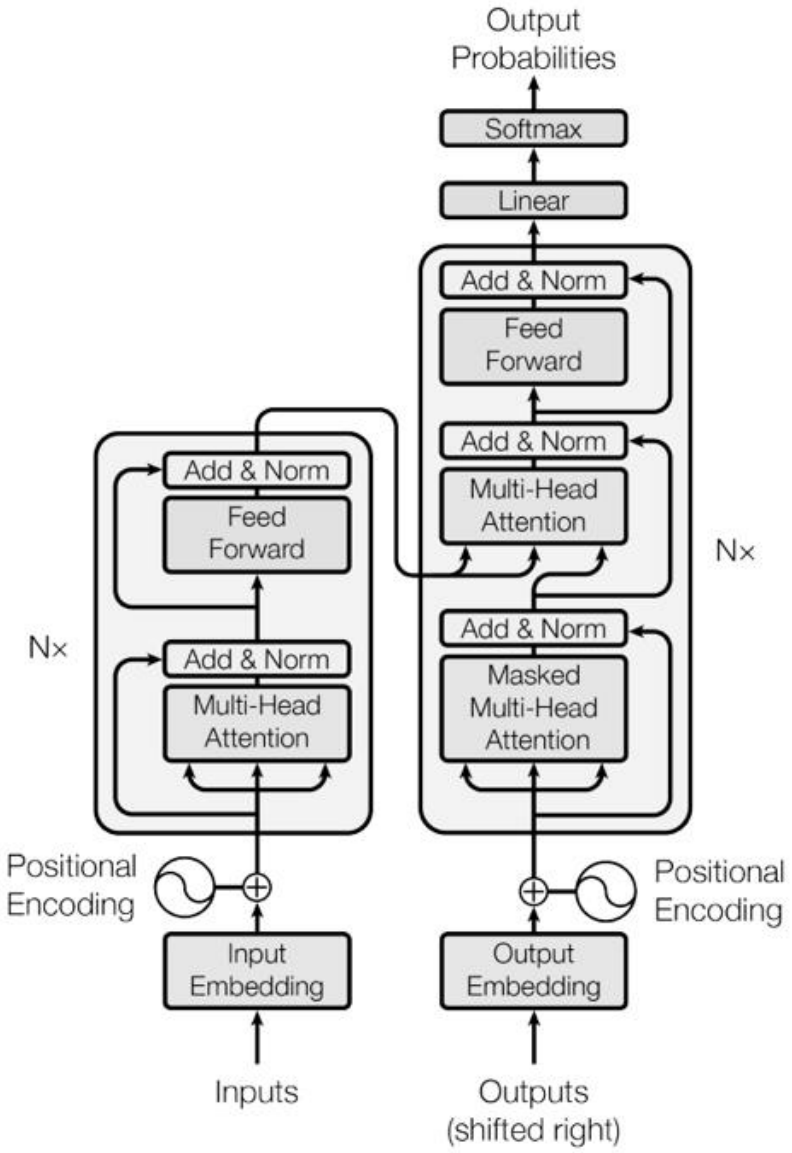


Figure 1: The Transformer - model architecture.

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

arXiv:1706.03762 [cs.CL]

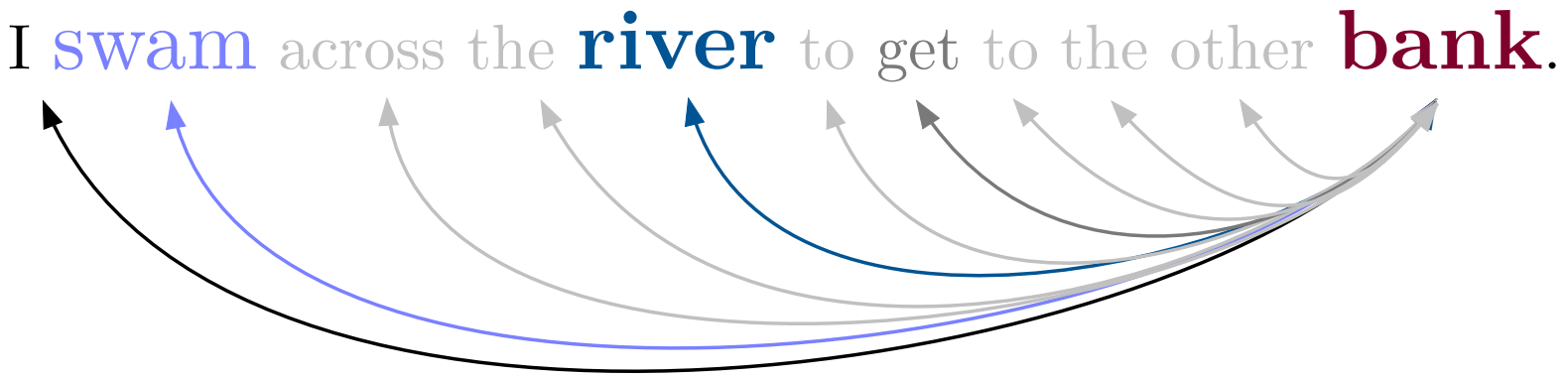
Attention Mechanism – Overview

Foundation for Transformer Models

You have no problem interpreting “bank” in the following sentence:

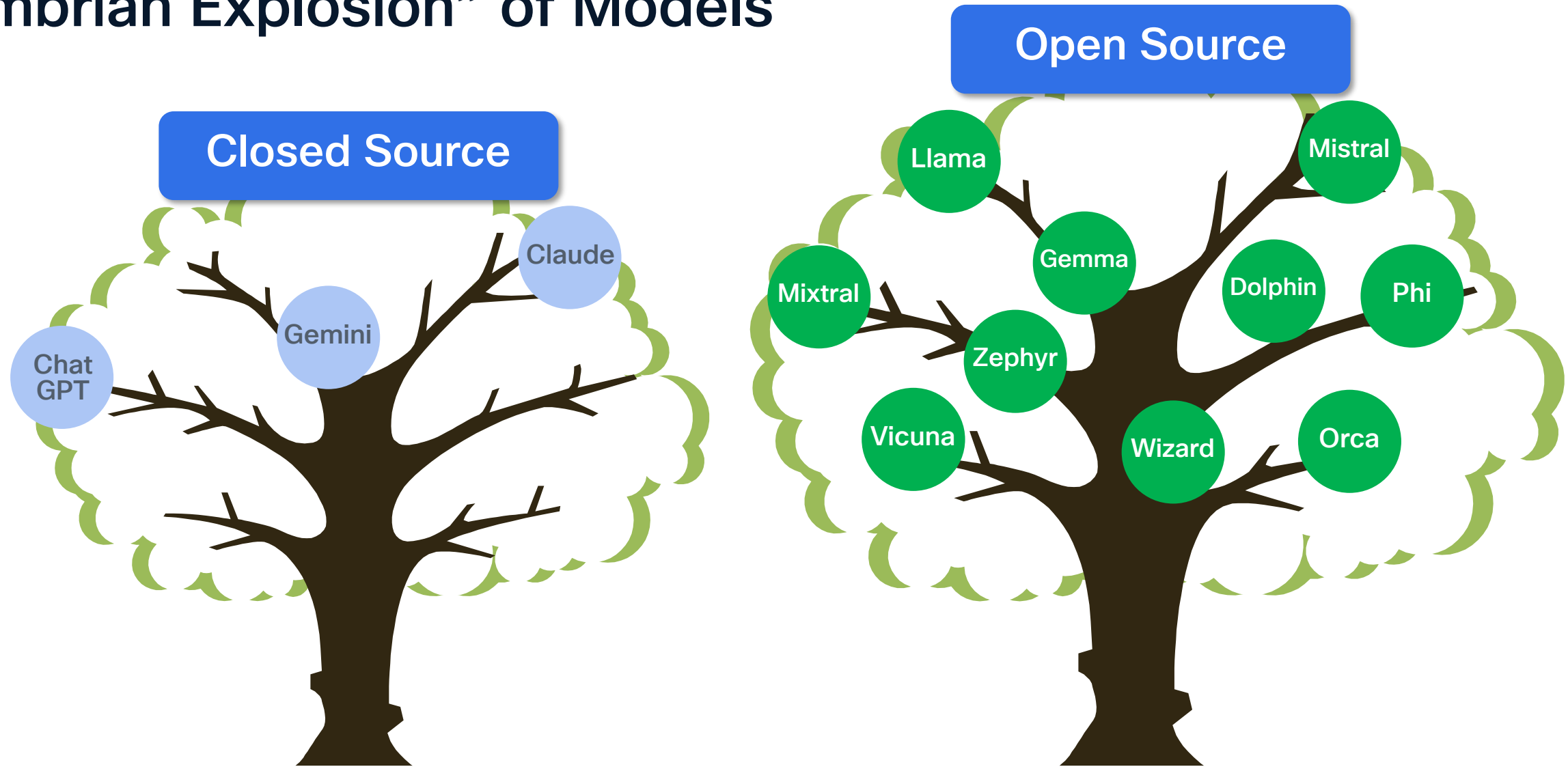
“I swam across the river to get to the other bank.”

A machine needs some help...



The goal of the attention mechanism is to add **contextual information** to words in a sentence.

“Cambrian Explosion” of Models

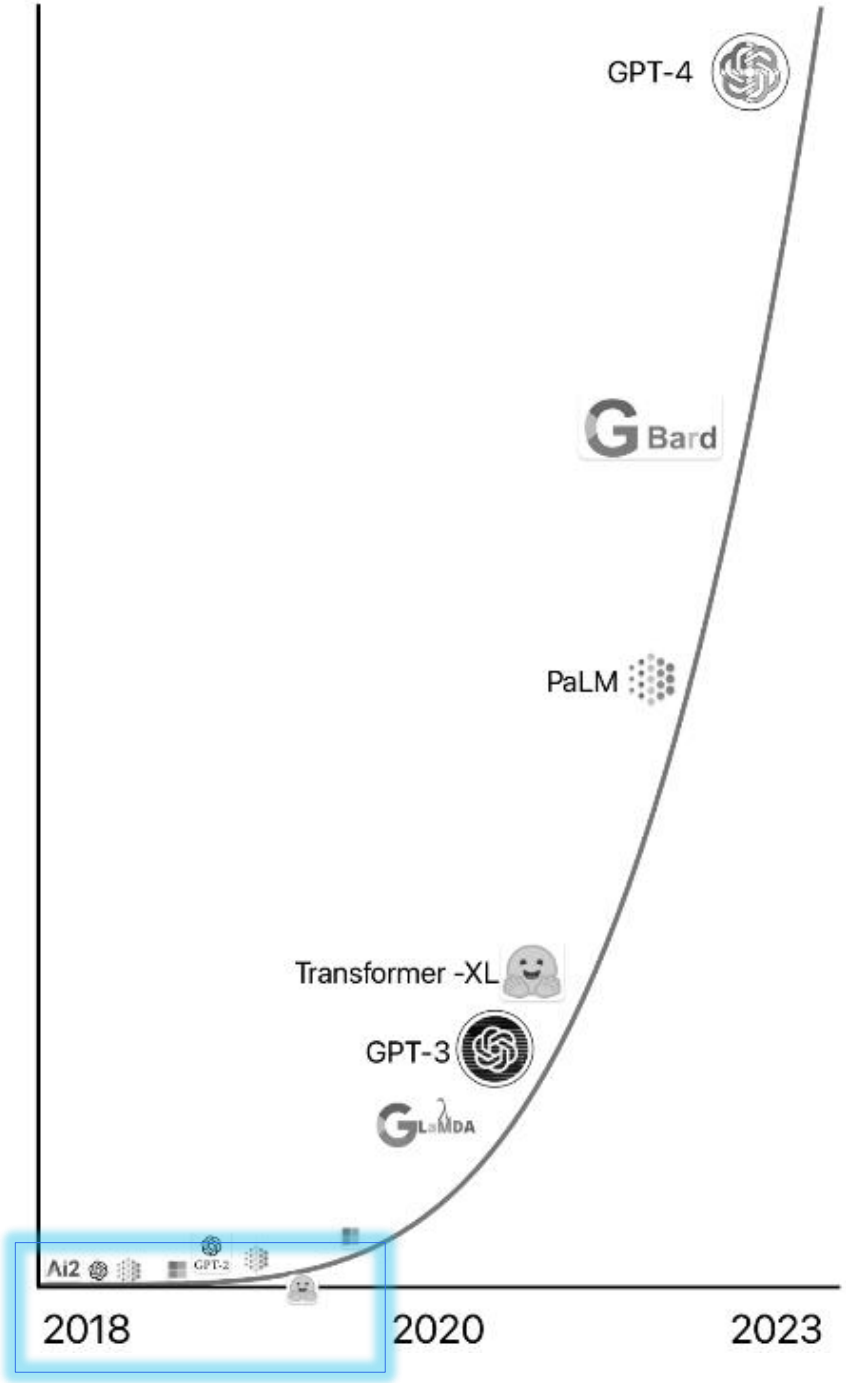


Models – Various types, sizes, focus, ...

From Billions to Trillions of Parameters ...

1,000,000,000,000

500,000,000,000



FUN FACT!

The human brain contains **86 billion neurons**, and over **100 trillion** synaptic connections

How are LLMs Trained for Text and Code?

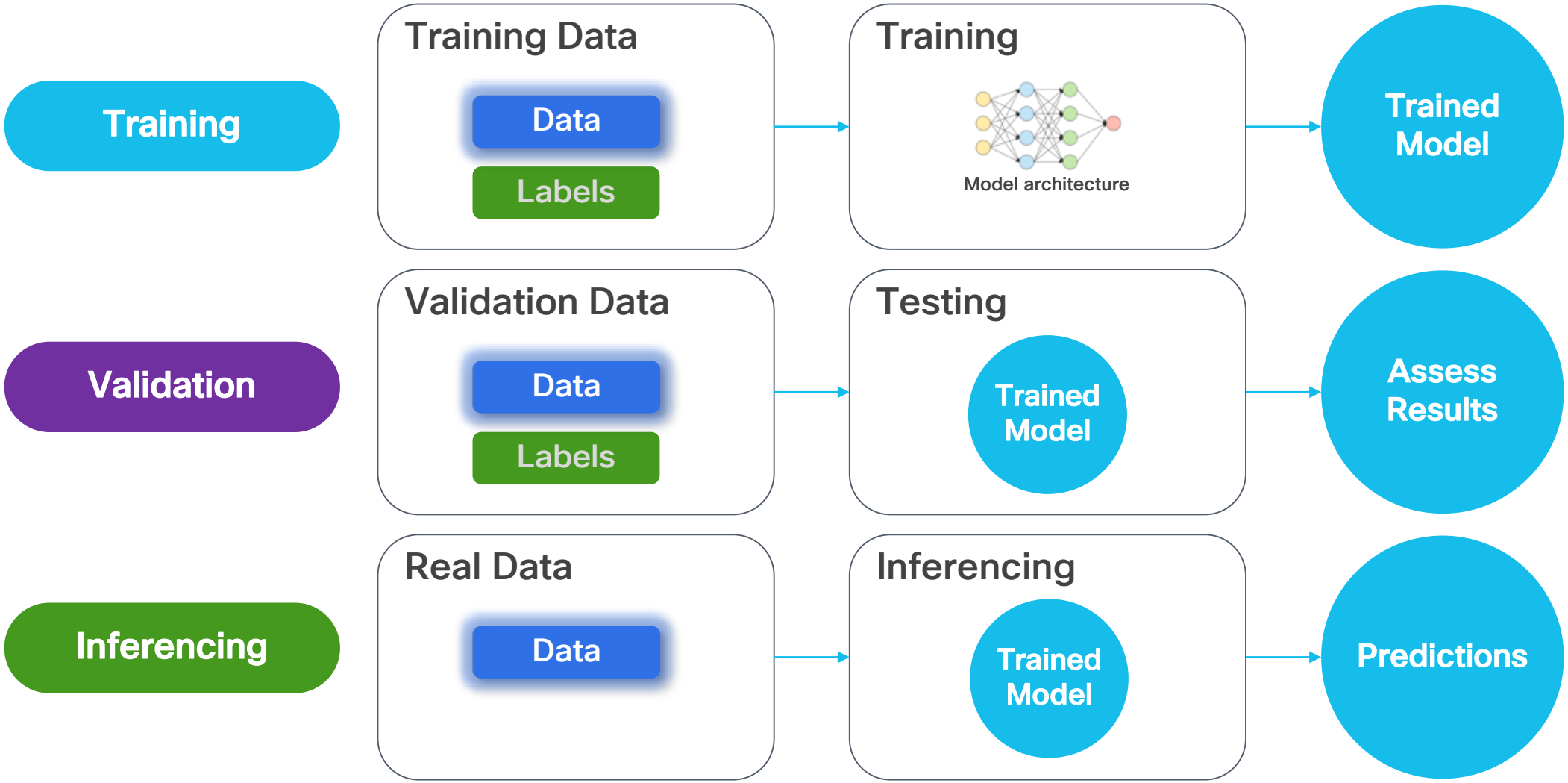
Step 1: Data Collection
(Feeding Knowledge)

Step 2: Tokenization
(Breaking It Down)

Step 3: Parameter Learning
(Storing Knowledge)

Step 4: Fine-Tuning
(Specialized Learning)

AI - Stages



Why Networking is Relevant to AI Deployments

LLMs are orders of magnitude more intensive than DLRM

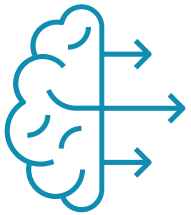
Deep Learning Recommendation Models

Search, Feed ranking. Ads & content recommendation

Inference needs a few Gigaflops for 100ms TTFT

Narrower scope, domain specific

Training: ~100 Gigaflop/ sentence



Large Language Models

Intricacies of human language

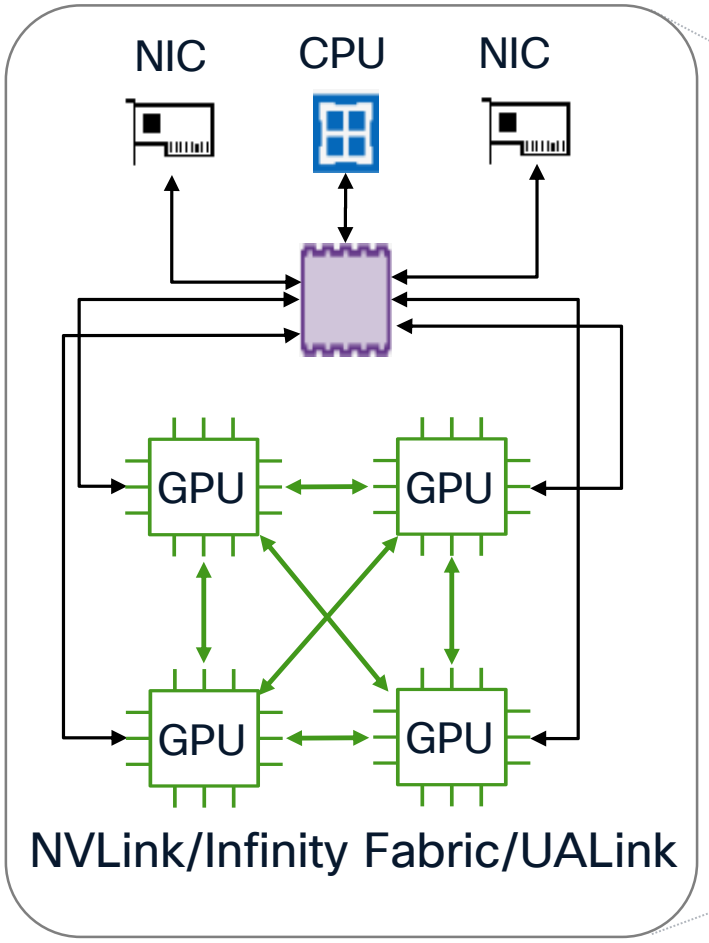
Inference needs 10s of Petaflops for 1 sec TTFT

Generate intZelligent, creative responses

Training : ~1 Petaflop/ sentence

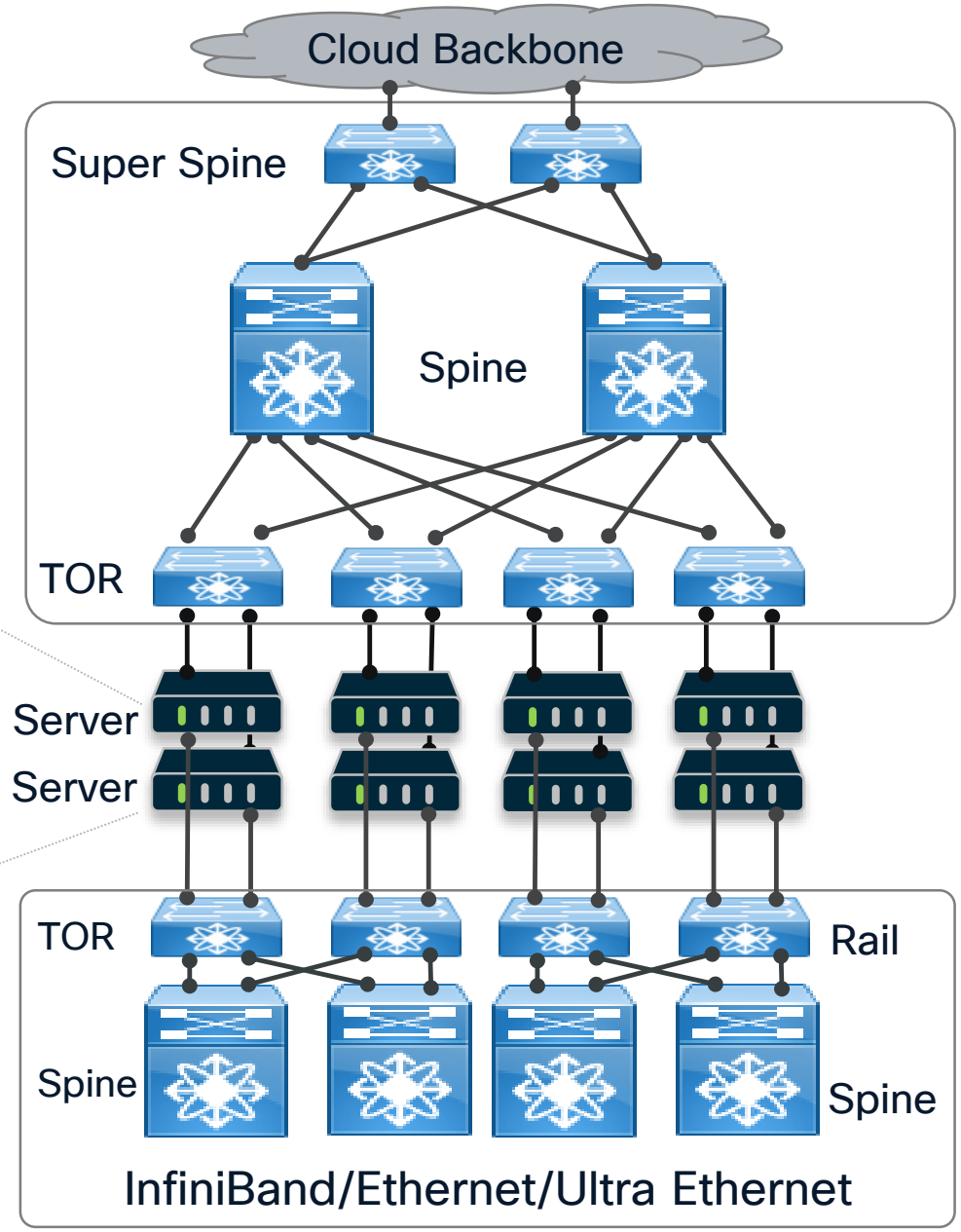
An Improved user experience means *a **faster time to first token,*** making ***distributed inference an imperative***

AI Network Fundamentals



Back-end
Scale-up
Network

Front-end
Network



Back-end
Scale-out
Network

AI Ethernet Fabric Options

	Ethernet	Enhanced Ethernet		Ultra Ethernet	Scheduled Ethernet
Load Balance	Stateless ECMP	Stateful Flow/ Flowlet	Spray & Re-order in SmartNIC	Endpoint Controlled adaptive packet spraying	Spray & Re-order in leaf
Congestion Management	Congestion Reaction with ECN/PFC	Adjust distribution based on congestion		Congestion Management	Congestion Avoidance
Link Failure	Software	Hardware		Hardware	Hardware
JCT	Good	Better		Even Better	*Best*
NIC and Fabric Coupled	No	No	Yes	Yes	No
Place in Network	Frontend, Backend	Frontend, Backend		Backend	Frontend, Backend

Performance *DEPENDENT* on Traffic Characteristics

Performance *NOT DEPENDENT* on Traffic Characteristics

Our AI strategy…
To connect and protect the AI era

Cisco connects and protects the AI era

AI infrastructure

Security for AI

Data

AI-native products

Services

Hypershield – Scalable hyper-distributed architecture

AI- Native

Built- in AI from inception.

Earns your trust, analysis- backed recommendations and interactions

Cloud- Native

Built on open source eBPF

eBPF powers default network for cloud- native workloads in hyperscalers

Hyper- Distributed

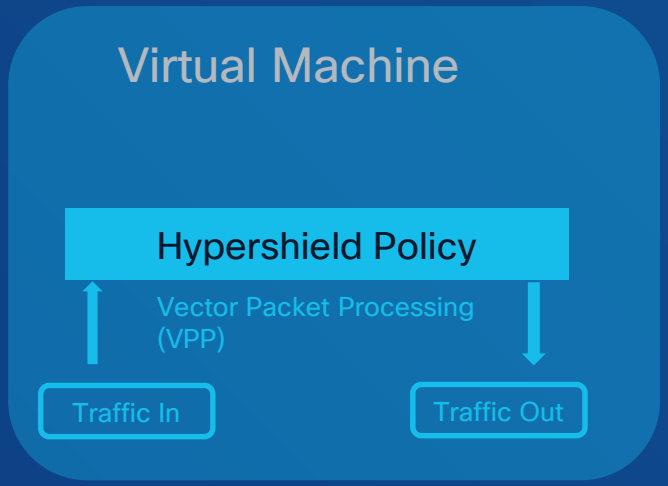
Distributed enforcement points across appliances, in the network, and the workload.

Managed as one system.

Network Enforcer VM

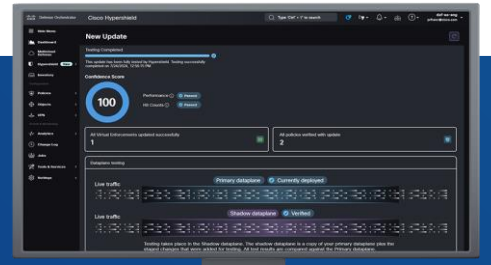
A platform to enable stateful services

Network



Security

Cisco Hypershield



Integrated security (license add-on)

- Intelligent security policy placement
- Self-qualifying policy updates
- Policy unified with workload/network enforcement, public and private clouds

Network Enforcer: N9300 Smart Switch

A platform to enable stateful services

Network

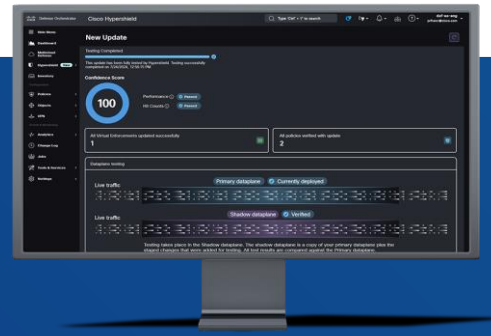
N9300 Series Smart Switches

Converge stateful services and network

- 800G stateful services throughput and scale
- 24-port 100G
- 4.8T Silicon One + 4 AMD DPU
- 1 RU

Security

Cisco Hypershield



Integrated security (license add-on)

- Intelligent security policy placement
- Self-qualifying policy updates
- Policy unified with workload/network enforcement, public and private clouds

Transform Your Network Security using Cisco Smart Switches integrated with Cisco Hypershield

N9324C-SE1U

24-port 100G

- Cloud Edge, Zone-Based segmentation, DCI, Top-of-Rack
- 2.4T switch throughput, 800G services throughput
- Silicon One E100 ASIC + AMD DPUs
- Shipping now, Hypershield Target Initial Product Readiness: end of July'25

NEW

N9348Y2C6D-SE1U

48-port 25G, 6-port 400G, 2-port 100G

- DC Top-of-Rack
- 3.8T switch throughput, 800G services throughput
- Silicon One E100 + AMD DPUs
- Target Limited Orderability: July ‘25*

NEW

Cisco C9350

48/24 ports 10G/mGig, network-modules, 90W UPoE

- Campus
- 1.3T (800G stacking, 500G for switch) throughput
- Silicon One E100 + Security co-processor
- Target Orderability: June ’25*

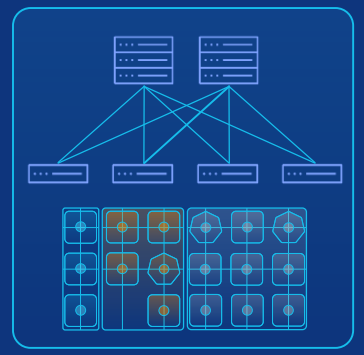
Policy Scale with Stateful Services

Pervasive Application Traffic Visibility

Zero Trust Security for All Workloads

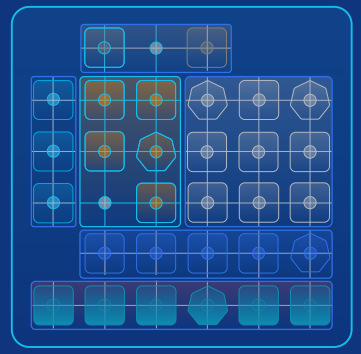
TCO Reduction

Cisco Hypershield use cases



L4 Zone-Based Segmentation

- Within and across data centers, cloud edge and top-of-rack
- Consistent policy enforcement
- Simplified architecture and lower costs



Autonomous Segmentation

- Deep understanding of app behavior
- Comprehensive inputs for policy creation
- Constantly adapting to changing apps

Distributed Exploit Protection

- Mitigate known and unknown vulnerabilities
- Surgical mitigating controls
- Protection within minutes, while app keeps running

AI in Hypershield

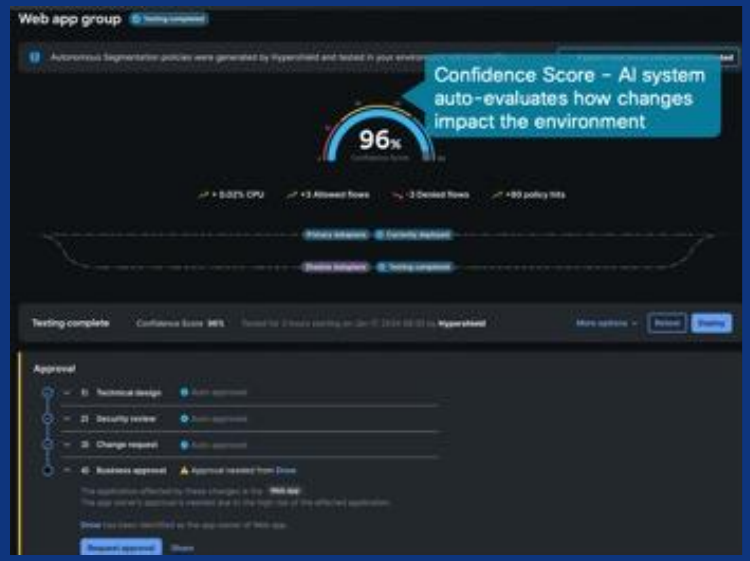
- AI- Native – AI modules as building blocks to deliver capabilities in Hypershield
- ML: Graph Engine
  - Structural learning and behaviral characteristics
- Use Cases
  - ML: Autonomous Segmentation
    - Pattern matching and machine learning
    - Continuous Policy recommendations
  - Gen AI: Distributed Exploit Protection
    - Generative AI use to generate compensating controls in a curated contruct
- Gen AI: AI Assistant

  Collateral

  Importing rules sets from existing FWs

…

Digital Twin :
Verify AI recommendation



Acceptance criteria for production test

Individual AI Assistants Are Integrated Across Cisco

Security	Firewall, Secure Access, Hypershield, Duo, Identity Intelligence, Splunk Enterprise Security, ISE
Networking	Meraki, Catalyst Center, Catalyst SD-WAN, ThousandEyes, Intersight, Mobility Services
Observability	Splunk Observability (Cloud, ITSI, AppDynamics)
Data	Splunk Platform
Collaboration	Webex Control Hub
Service Ops	Customer Experience

Native Skills Across Products Examples

1. Connection & Security logs
2. Policy inquiry
3. Policy creation

1. SPL generation
2. SPL querying
3. Data summarization

1. User activity timeline
2. Device info & compliance
3. Authentication logs

1. Client troubleshooting
2. Device troubleshooting
3. App troubleshooting

1. Internet outages
2. Network events
3. User to app troubleshooting

1. TAC case management
2. Field notices
3. Vulnerability & PSIRTs

What is an AI Agent?

- An Autonomous system “skilled” to accomplish specific task(s)
- LLM accompanied with:
  - Tools / Functions
  - Memory
- Core capabilities:
  - Planning and Reasoning

“Do This Task for Me, Please”

Assistant

Scheduling

Errands

Expenses

Famous Actor

“Accomplish This Goal, Please”

“Done! ☺”

Agent

Pitching

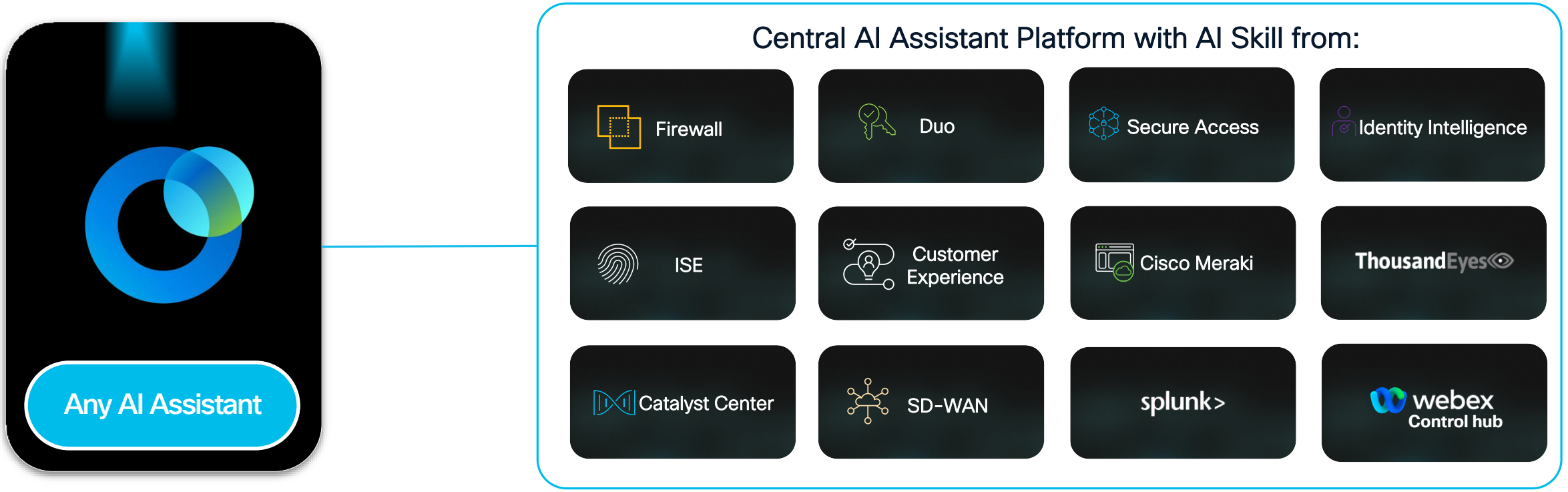
Reasoning
External access
Evaluation
Reflection

Relationship Building

Career Management

Contract Negotiation

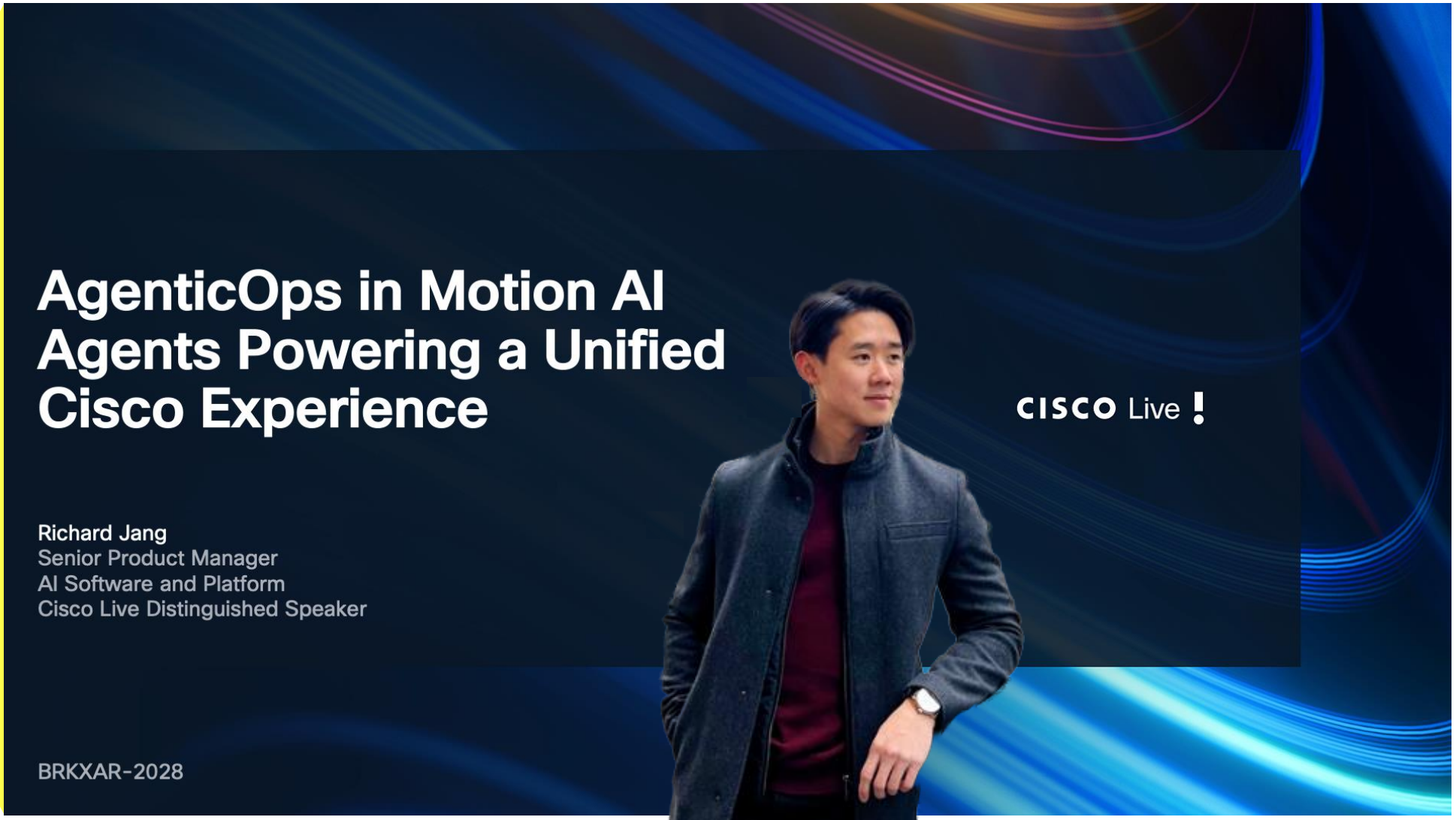
Unify Cisco AI Assistants to enable a network of AI Agents that can use cross-product AI Skills



More about Cisco Assistants, GenAI, ...

Tuesday, June 10th

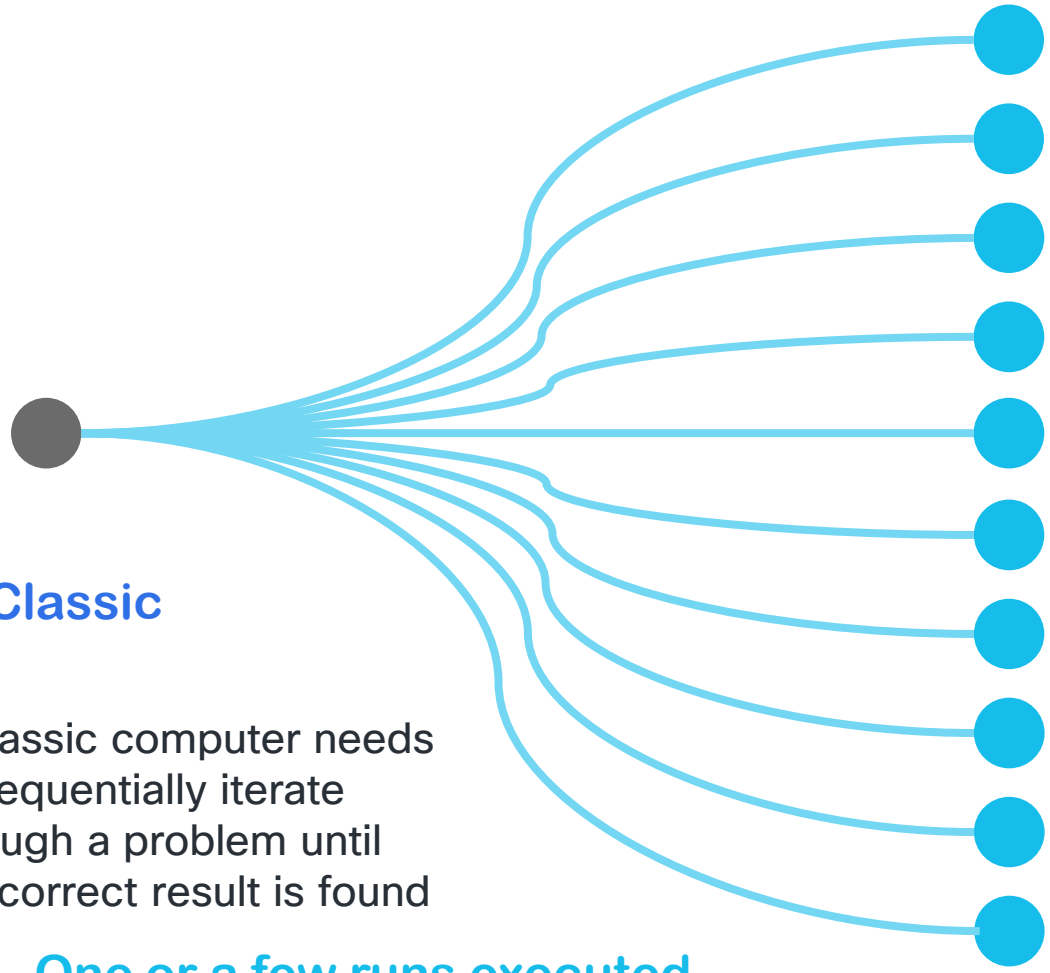
2:00 – 3:30pm



Late-breaking news!

Quantum @ Cisco

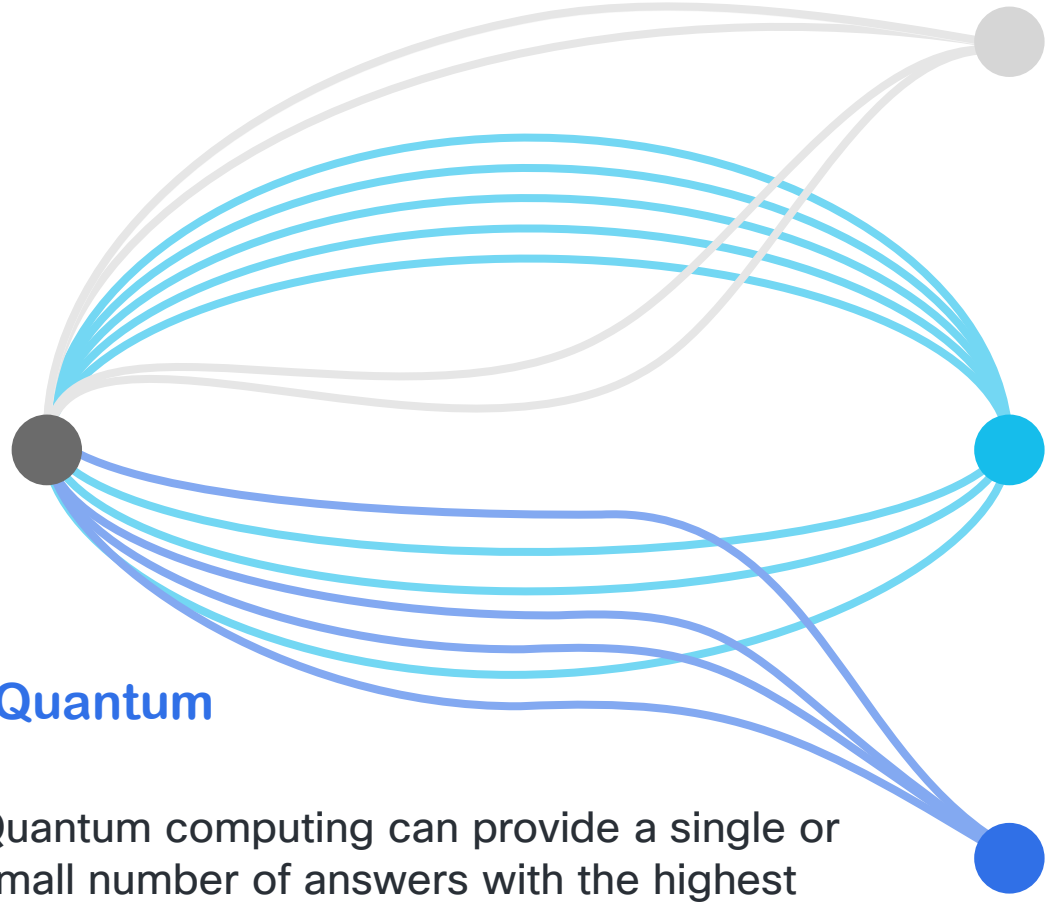
Quantum Solving vs Classic Solving



Classic

A classic computer needs to sequentially iterate through a problem until the correct result is found

One or a few runs executed



Quantum

Quantum computing can provide a single or small number of answers with the highest probability of being correct which narrows down the search for the correct solution

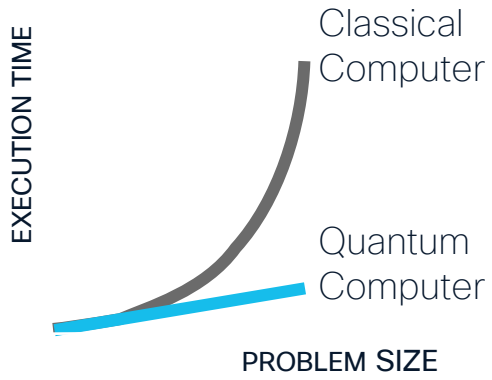
Multiple “shots” (could be 1000s) are executed to get a probability distribution of results

Based on an explanation from Australian Academy of Science and Kurzgesagt

Cisco's Focus in Quantum Computing

Use Cases Today:

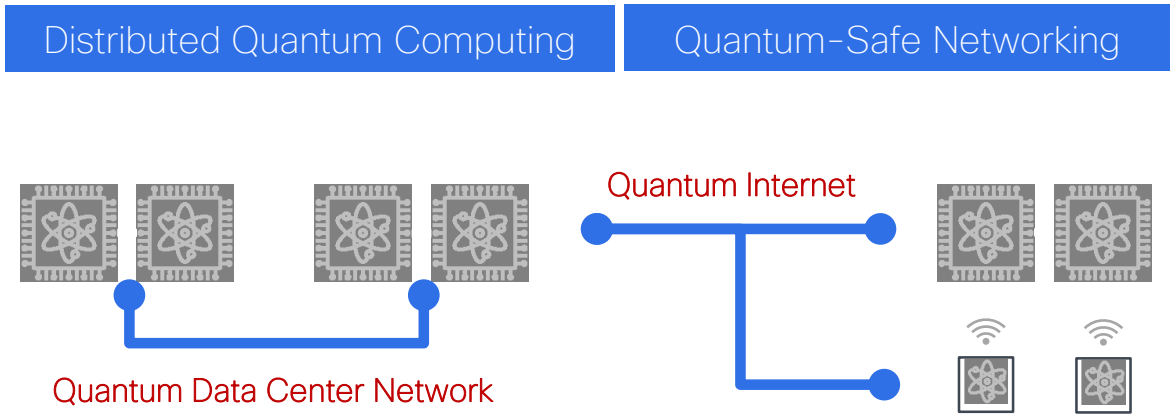
- Financial forecasting
- Drug discovery
- Logistics optimization
- Cryptography
- AI/ML models
- Weather forecasting

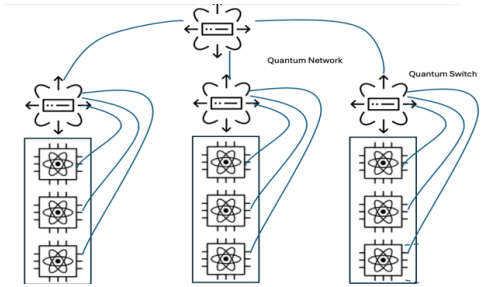


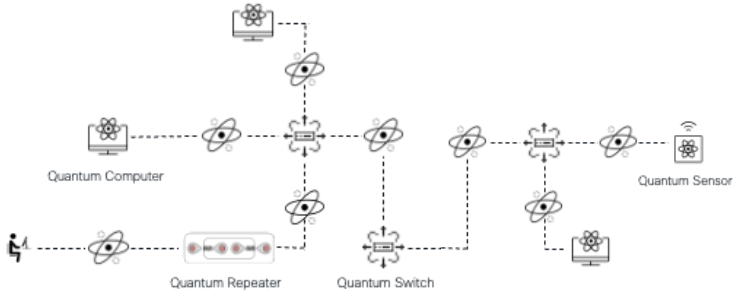
At-scale (fault tolerant) quantum computing will require 1M+ physical qubits

We focus on

- distributed quantum (+classical) computing
- quantum networking
- quantum-safe communications

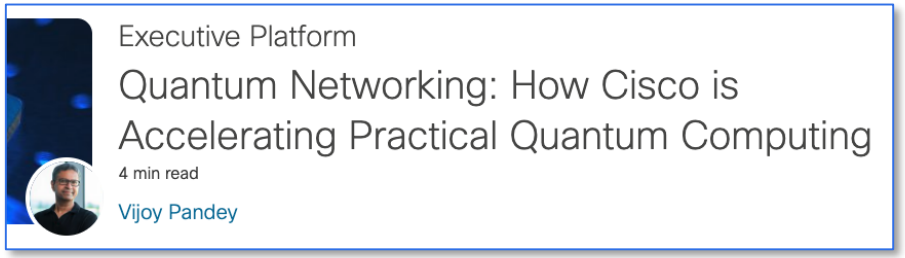


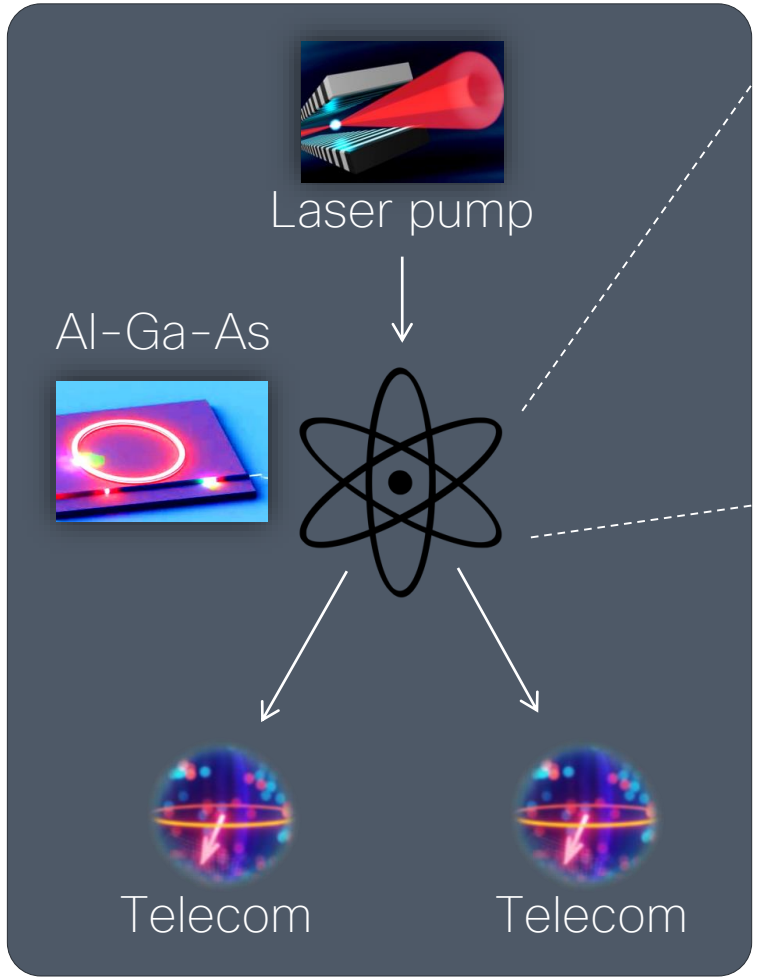


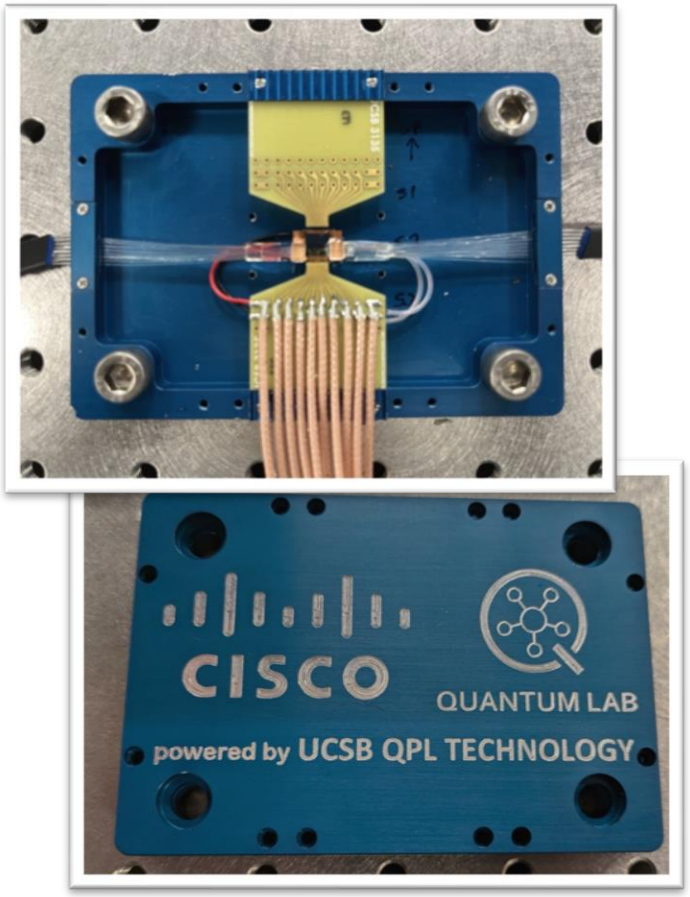


Quantum Entanglement Sources

(Designed by Cisco)







- **High rate:** >1M pair/sec per channel, 200M pairs/sec in-chip
- **High fidelity:** > 99%
- **Low power:** < 1mW
- **Small form factor:** 100µm on photonic integrated chip
- Operates at **room temperatures**

Keep your eye on this space!

https://blogs.cisco.com/news/quantum-networking-how-cisco-is-accelerating-practical-quantum-computing

Want to Know More?

An Introduction to Quantum Mechanics, Computing, and Networking

What is Cisco Doing in Quantum?

Tim Szigeti
Distinguished Engineer

cisco Live !

BRKETI-1401

Monday, Jun 9
8:00 AM - 9:30 AM

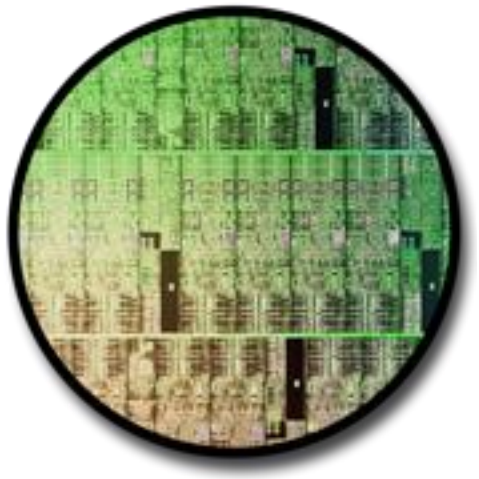
Available in the CISCO LIVE ON-DEMAND LIBRARY

Tim Szigeti

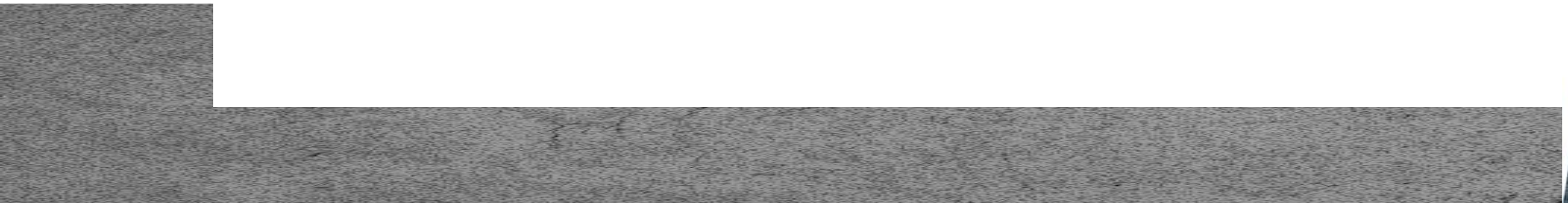
Summary

Why Hardware Still Matters in a Software-Defined World

ASICs

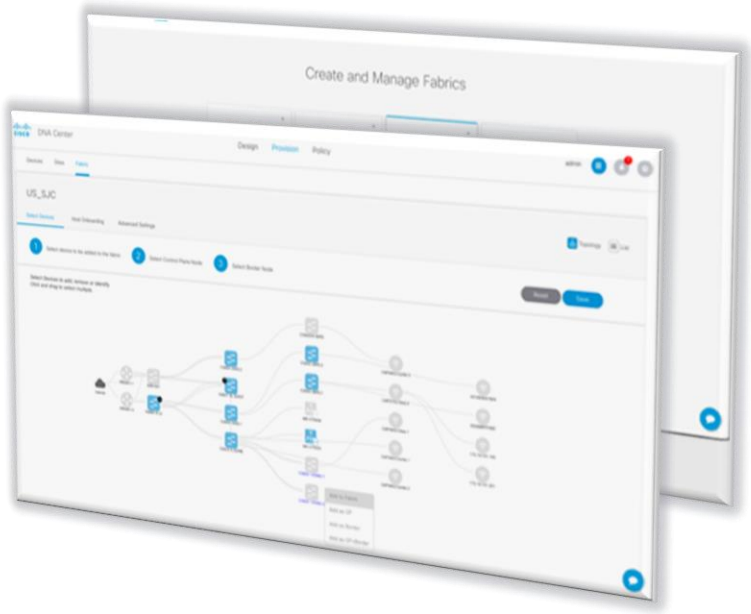


Critical Role of Flexible Silicon

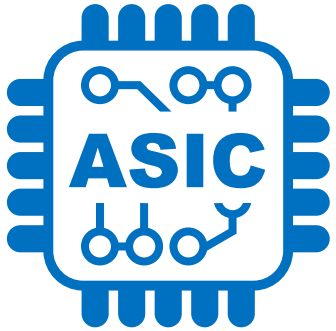


Innovation All The Way Up the Stack

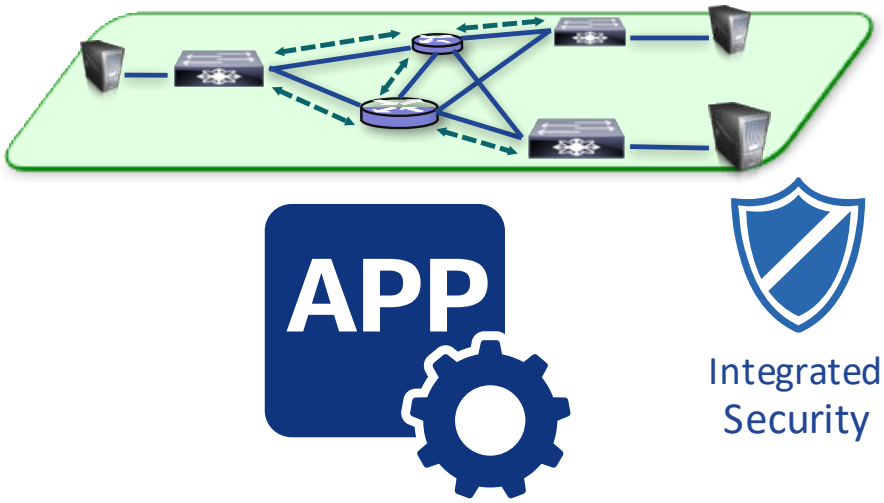
Hardware, Software, and Solutions



From the **Hardware ...**



... to the **Software and Protocols, with Integrated Security ...**



to the **Whole Solution ...**

Cisco Innovations – In Hardware, Software, and Solutions – **Tie It All Together**

How Did We Do?

Innovations in ~~Silicon~~ Hardware and Software–
Defining the Next Generation of Enterprise Technology

Do You Have a Better Understanding ...

... of how **ASICs and Network Hardware** are Designed and Built ...

... of why **Hardware Flexibility** is Key for Solutions ...

... and how You can **Leverage Cisco's Latest Flexible Hardware and Advanced Software** in Your Own Network Designs?

Complete your session evaluations

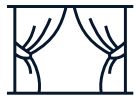
Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to win 1 of 5 full conference passes to Cisco Live 2026.

Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

Level up and earn exclusive prizes!

Complete your surveys in the Cisco Live mobile app.

Continue your education



Visit the Cisco Showcase for related demos

Book your one-on-one Meet the Engineer meeting

Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs

Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

What else to see

Silicon One

- Networking for AI - DEMCPA-09
- Networking for AI | Silicon One - DEMAIDC-04
- Redefine your AI/ML networks with Silicon One - PSODCN-1005
- Redefine your AI/ML networks with Silicon One - AIHUB-1004
- SILICON ONE & ULTRA ETHERNET FOR AI INFRASTRUCTURE - BRKNWT-2508
- Preparing for AI-Ready Infrastructure with Silicon One - ITLGEN-2065
- Silicon One - DEMCPA-10
- Ethernet Fabrics for AI clusters - Silicon One and Nexus - ultra high performance, scalable & non-blocking ethernet fabric. - BRKCOC-3005

Liquid Cooling

- WoS demonstration – Sustainability Booth
- Integrated Rack Design | Liquid Cooling for Networking, Linear Pluggable Optics, and Rack System Cooling - DEMAIDC-02
- The AI-Revolution – Cooling Technologies for the Data Center & Edge - WOSGEN-2100
- Improving Power Usage Effectiveness | Immersion Cooling and Energy Management - DEMAIDC-06
- Next generation power and cooling technologies in the datacenter - IBOCOM-2101

Optics

- Optics for AI Infrastructure - WOSGEN-2102
- Optics for AI Connectivity - DEMSGC-03
- Integrated Rack Design | Liquid Cooling for Networking, Linear Pluggable Optics, and Rack System Cooling - DEMAIDC-02
- 400G, 800G, and Terabit Pluggable Optics: What You Need to Know - BRKOPT-2699

What else to see (cont.)

Campus Switching

- Cisco Catalyst 9000 Series, Cisco Silicon One, and IOS XE Architecture and Innovation – BRKARC-2092
- Catalyst 9000 Series Switching Family: Access – BRKARC-2098
- Catalyst 9000 Series Switching Family: Core and Distribution – BRKARC-2099
- Cisco Catalyst 9000 Switching QoS with Silicon One ASICs Deep Dive – BRKARC-2039
- Catalyst 9000 Switching Innovations & Roadmap – CIUG-1109
- Packet Journey inside Catalyst 9000 Switches – BRKARC-2100

Thank you

cisco Live !


CISCO