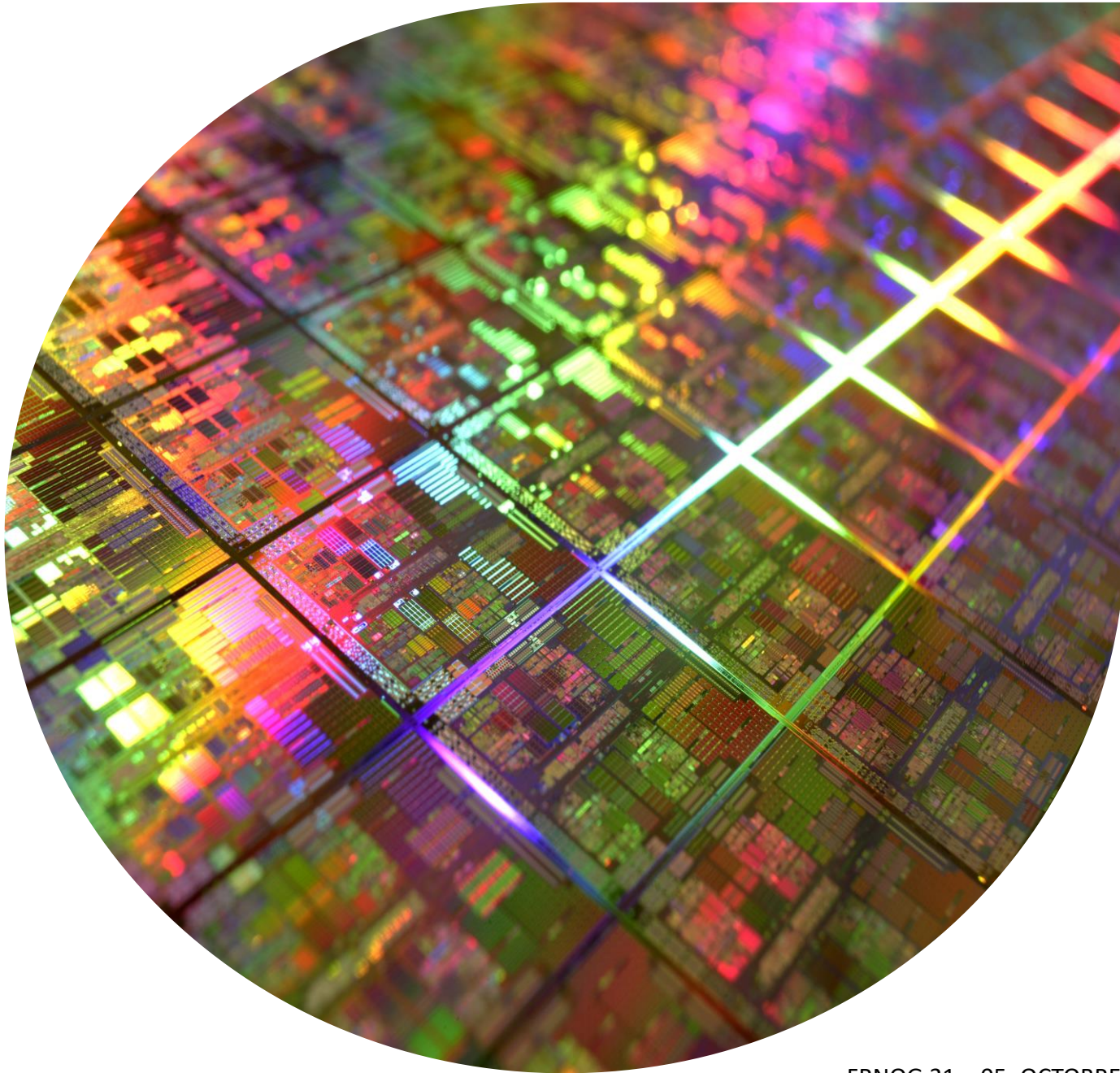




Les CPUs au cœur des transferts

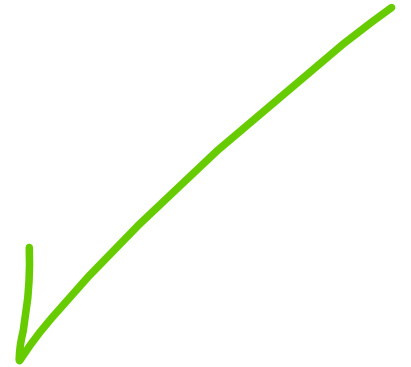


Qui sommes-nous ?

Fabriquant de serveurs pour le calcul et le stockage depuis 2004.
Basé principalement à Strasbourg et San Jose, CA.

~~AS~~ ☹️

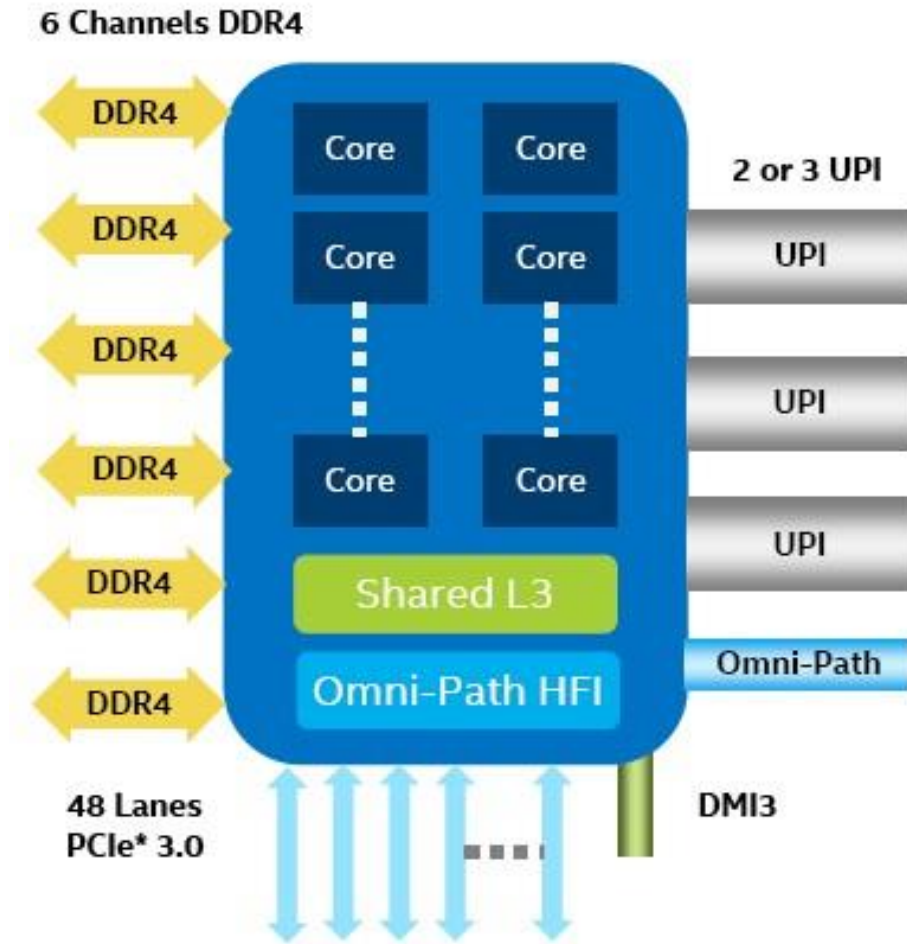
3C:3F:51:xx:xx:xx 😊



Adrien Badina – Directeur de l'Innovation

Frédéric Mossmann – Directeur Technique

Les CPUs au cœur des transferts



Réseau « très local »

- Bus mémoire
- Bus d'interconnexion CPU
- Bus PCIe

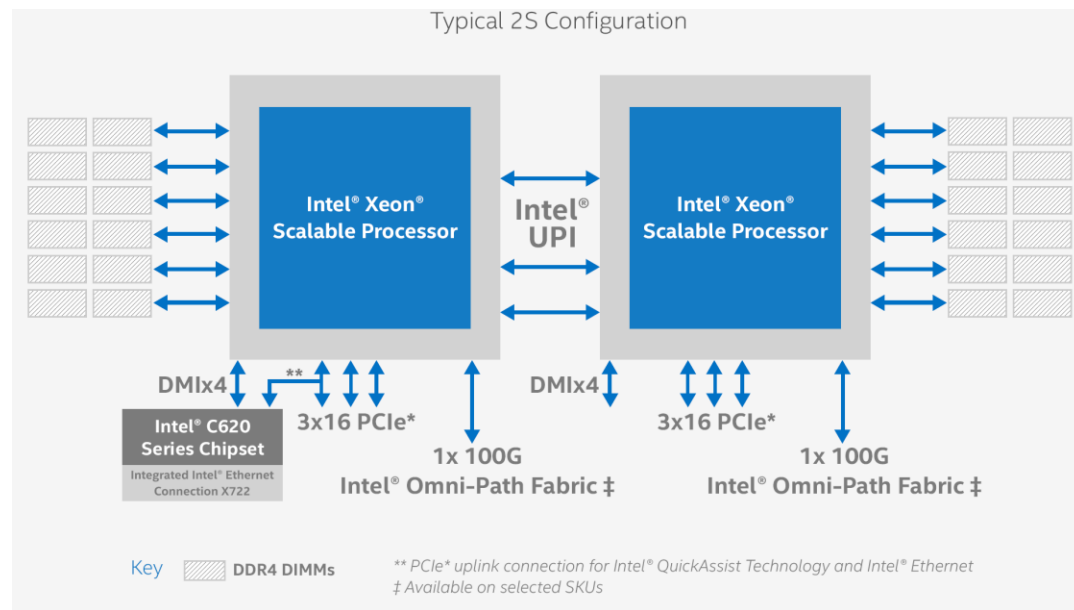
Bus mémoire

Type	Clock speed	Débit (par channel)		2018	2019	2020	2021	2022	2023	2024
DDR3	1066 MHz	17 Go/s								
DDR4	1600 MHz	25 Go/s								
DDR5	2666 MHz	50 Go/s			Adopt.					

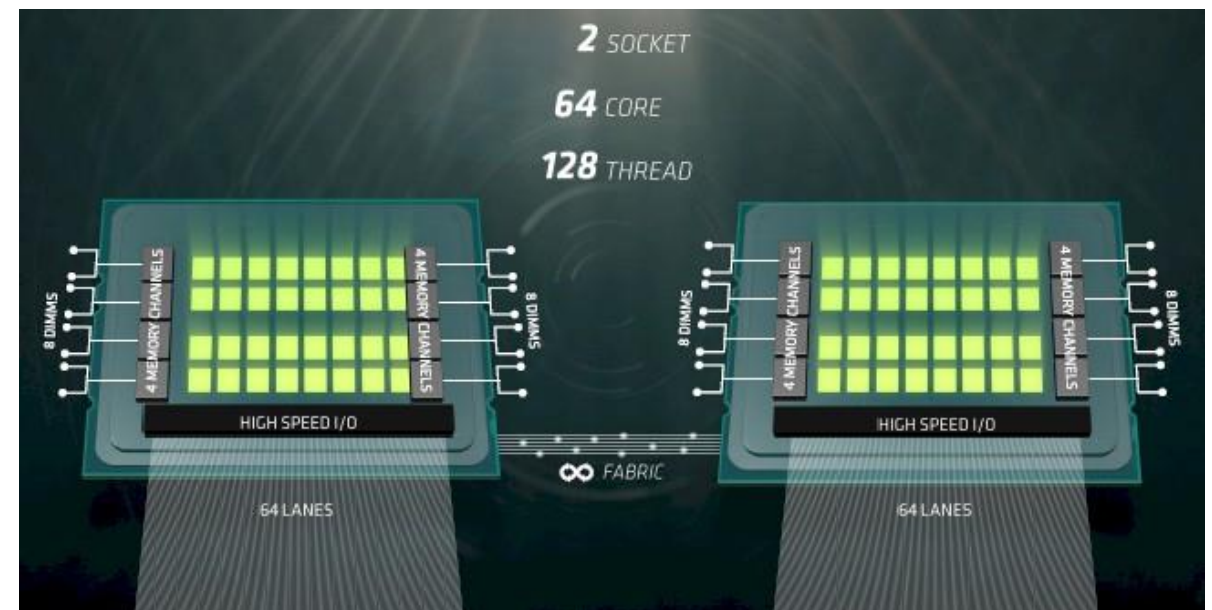
Vendor	Platform	Nb channels	Max speed	Bandwidth
Intel	Grantley	4	19,2 Go/s (DDR4-2400)	77 Go/s
Intel	Purley	6	21,3 Go/s (DDR4-2666)	128 Go/s
Intel	Purley refresh	6	23,5 Go/s (DDR4-2933)	140 Go/s
Intel	Whitley	8	25,6 Go/s (DDR4-3200) ?	205 Go/s
AMD	Naples / Zen	8	23,5 Go/s (DDR4-2933)	188 Go/s
AMD	Rome / Zen2	8	25,6 Go/s (DDR4-3200) ? 35,2 Go/s (DDR5-4400) ?	205 Go/s 282 Go/s

QPI UPI IFIS

CPU		Lien	Nbre lien	DEBIT Max / lien	
INTEL-Grantley	Broadwell-EP + Wellsburg	QPI	2	9,6 GT/s - 38,4 Go/s	Quick Path Interconnect
INTEL-Purley	Skylake-SP and Lewisburg	UPI	3	10,4 GT/s – 41,6 Go/s	Ultra Path Interconnect
AMD	Zen / Vega	IFIS	4	> 130 Go/s	Infinity Fabric Inter Socket



INTEL



AMD

Bus PCI Express

Gen	Spec	Transf. rate	Line code	x1	x16
Gen1	2003	2.5 GT/s	8b/10b	250 Mo/s	4.0 GB/s
Gen2	2007	5 GT/s	8b/10b	500 Mo/s	8.0 GB/s
Gen3	2010	8 GT/s	128b/130b	1 Go/s	15.8 GB/s
Gen4	2017	16 GT/s	128b/130b	2 Go/s	31.5 GB/s
Gen5	(V0.7)	32 GT/s (25 GT/s ?)	128b/130b	4 Go/s	63.0 GB/s

Vendor	Platform	Nb links (SP)	Nb links (DP)
Intel	Grantley	40	80
Intel	Purley	48	96
Intel	Whitley	≥ 48 ?	≥ 96 ?
AMD	Naples / Rome	128	128 (*)

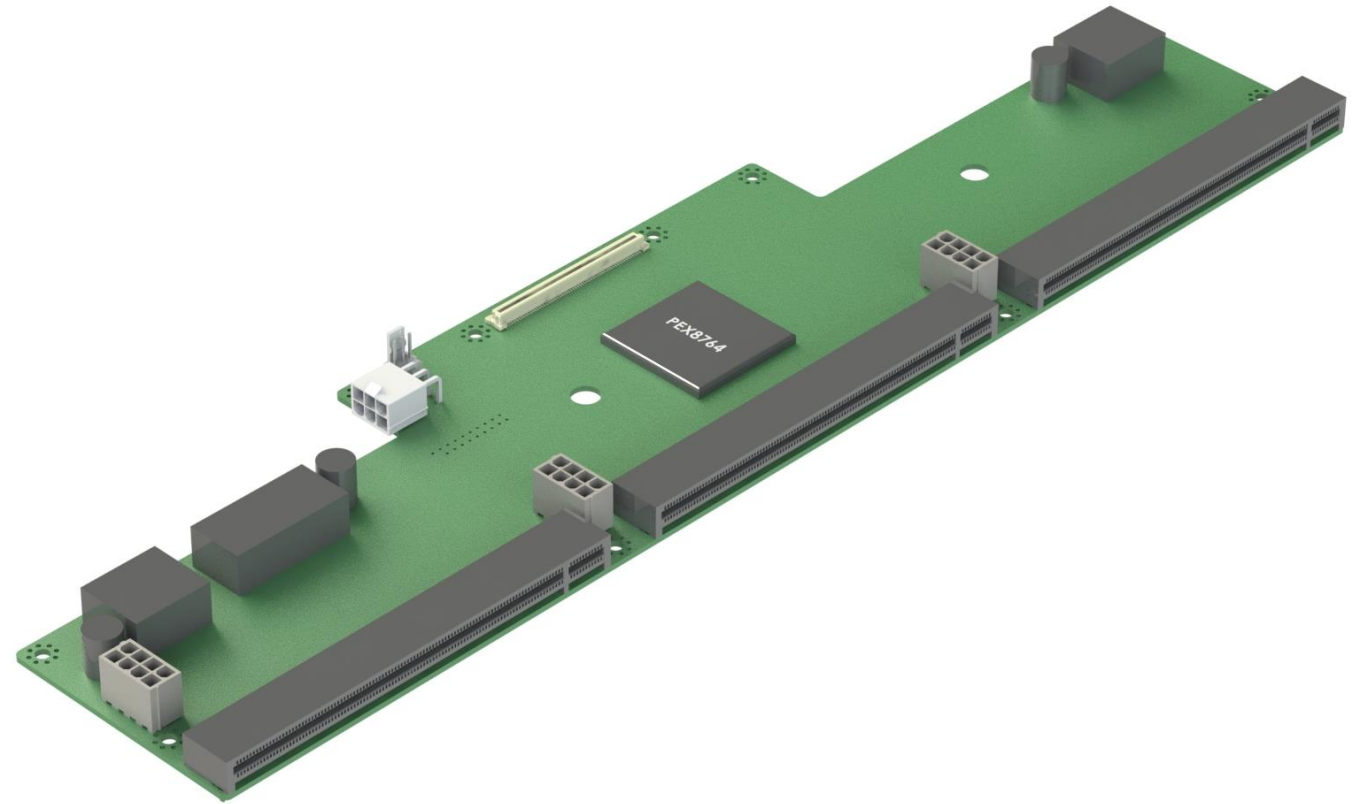
(*) 64 lanes PCIe par CPU utilisées pour l'Infinity Fabric

Contourner les limites du PCIe Gen3 ?

- Attendre le PCIe Gen4
- Contourner la limite de débit : NVLink
(300Go/s, de GPU à GPU, de Power à GPU)
- Contourner la limite de liaisons : Switch PCIe
(augmentation de la latence)

Switching PCIe

Pro	Cons
+ Lanes PCIe	Prix Encombrement
DMA entre devices	Latence



Impact du Switching PCIe dans les DTN ?

La latence impacte les petits échanges (accès aléatoires).

Objectif : Voir l'impact des switchs PCIe dans les gros transferts séquentiels (Data Transfer Nodes)

Matériel : Serveur 2 x Xeon E5, 24 x NVMe Intel PCIe Gen3 x4

- 2 Switchs PCIe Gen3 x16 vers x48 (12 NVMe, ratio 3:1)
- Un switch par CPU : x16 dispo pour cartes 100Gb/s

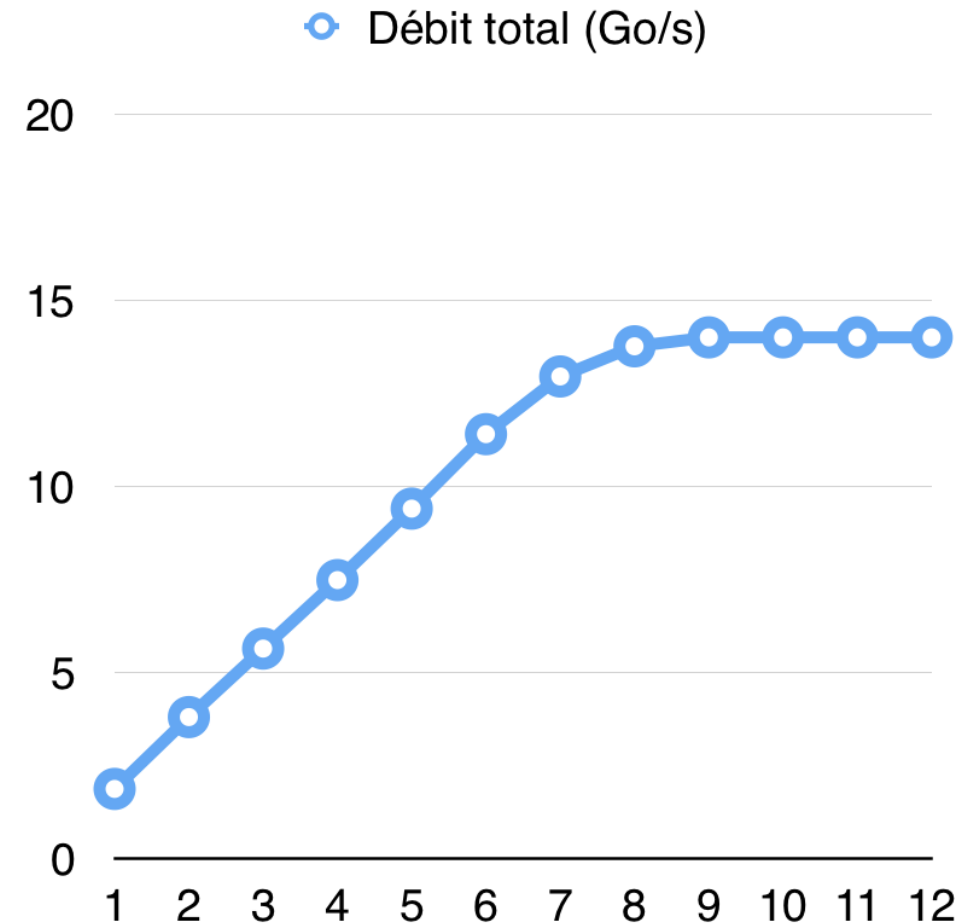
Tests : LCaltech + démo au SC17 sur SCinet



Caltech

Résultats en local

- Débit : parfois 2 Go/s sur PCIe x4 (= 4 Go/s)
- Les disques ont jusqu'à 50% de latence
- Sequential R/W avec Queue
 - 1 NVMe : Très peu d'impact
 - 4 NVMe : Parfaitement proportionnel
 - 7 NVMe : Toujours proportionnel !
 - > 7 NVMe : Saturation PCIe x16
- Les switchs PCIe sont bénéfiques
 - (pour les gros transferts)
 - (avec devices ayant de la latence)



2CRSI SERVER WITH INTEL NVMe DRIVES INTERNAL I/O LOCAL PERFORMANCE

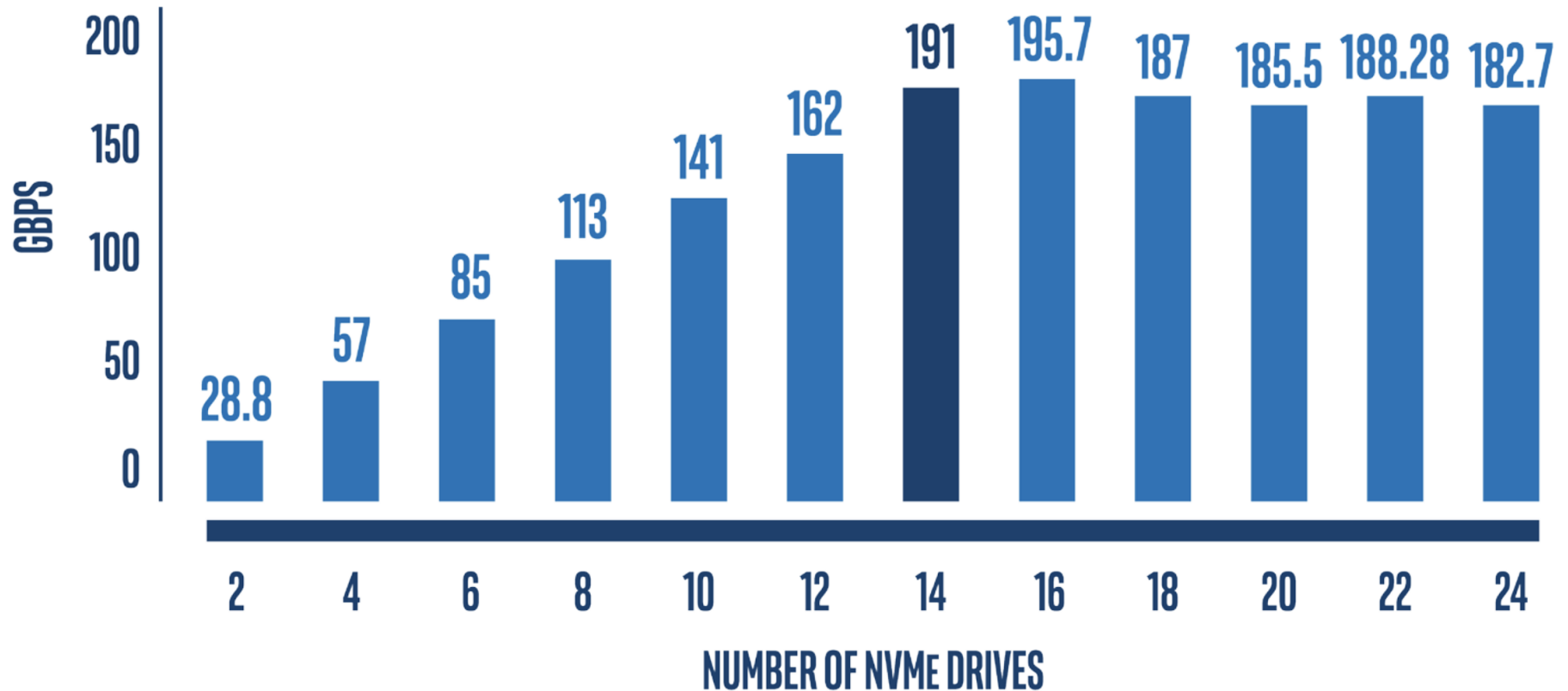


Figure 6: Max throughput reached at 14 drives (7 drives per processor). A limitation due to combination of single PCIe x 16 bus (128Gbps), processor utilization and application overheads.⁴ Source - Caltech.

Q/A

