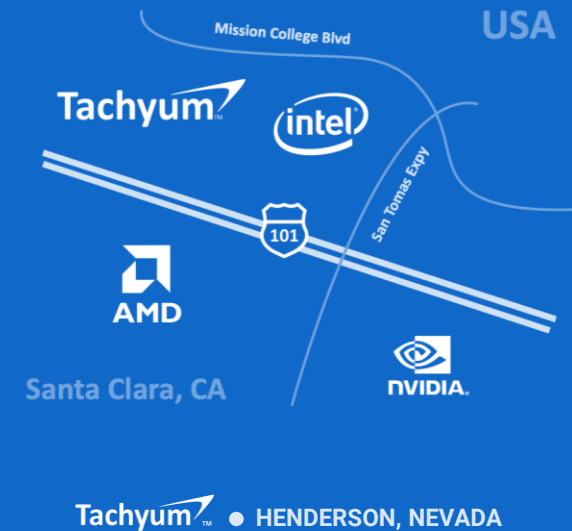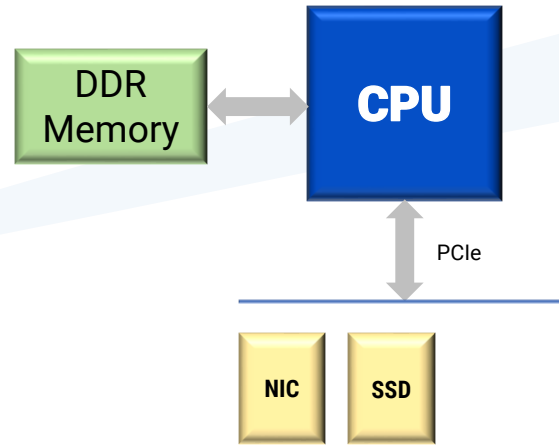# Prodigy Compared with Homogeneous and Heterogeneous Computer Architectures

10/31/2020
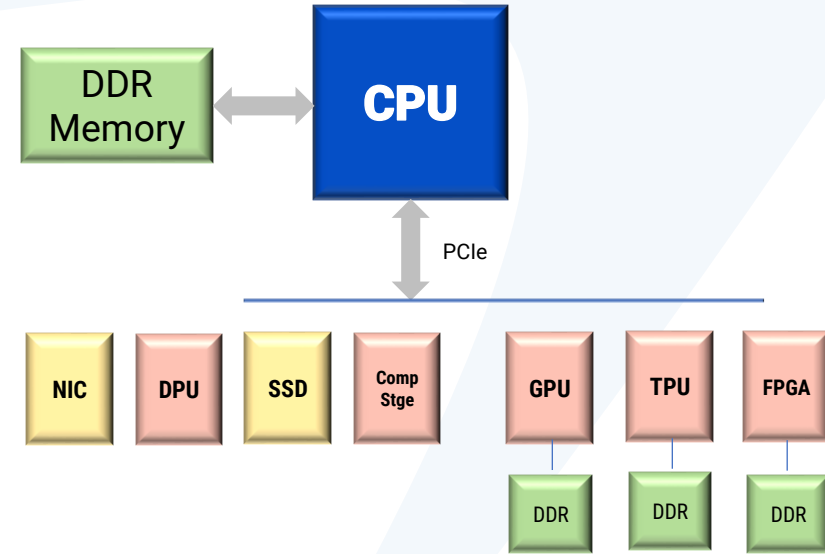
Traditional Homogeneous vs. Heterogeneous Architectures
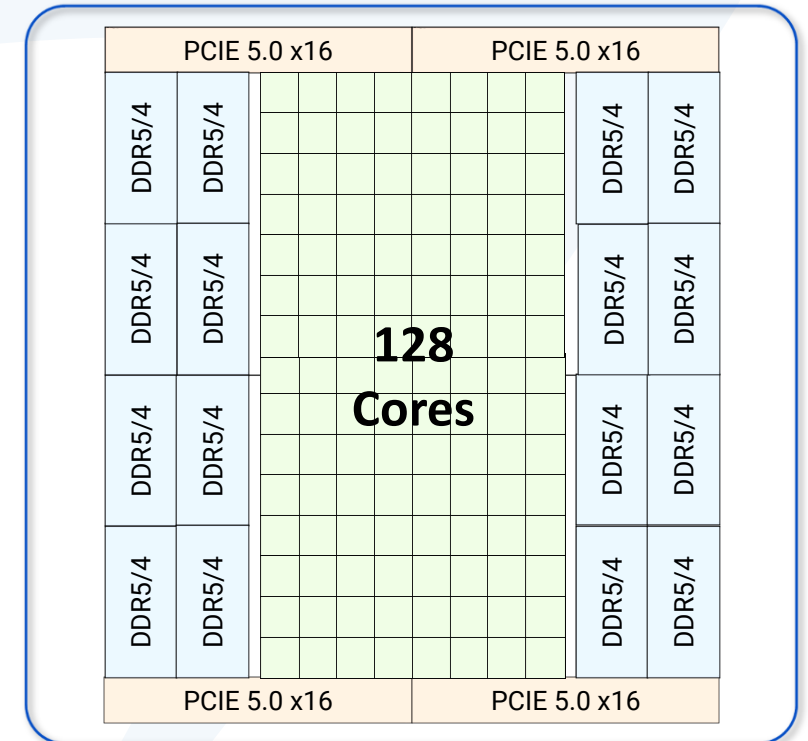
Homogeneous

Pros	Cons
• General purpose, flexible • Easy deployment/ maintenance	• Not designed for HPC or AI • Low parallel performance for modern workloads

Heterogeneous

Pros	Cons
• Accelerates specific workloads, including HPC and AI • Scalable	• Requires special programming/config • Expensive, power-hungry • Under-utilized – contrary to software defined data center

Tachyum™

Tachyum Prodigy – Advantages of Homogeneous and Heterogeneous Architectures without the Disadvantages

- **High Integer Performance for General Purpose Workloads**
  - Up to 128 general purpose cores

- **High Floating Point Performance for Parallel Workloads**
  - Dual 512b vector units provide high performance HPC, AI/ML

- **Scalable**
  - Family of 16 – 128 core devices with support 2P and 4P platforms

- **Common Software – Easy Deployment/ Maintenance**
  - All cores part of same ISA

- **High Memory Bandwidth**
  - 16 DDR5 controllers provides best in industry bandwidth

Samples Q4 '21

	PCIE 5.0 x16	PCIE 5.0 x16	
DDR5/4 DDR5/4			DDR5/4 DDR5/4
DDR5/4 DDR5/4	**128 Cores**		DDR5/4 DDR5/4
DDR5/4 DDR5/4			DDR5/4 DDR5/4
DDR5/4 DDR5/4			DDR5/4 DDR5/4
	PCIE 5.0 x16	PCIE 5.0 x16	

Q2 '21:   Emulation for early adopters

Q3 '21:   Emulation system general access

Q4 '21:   Device samples

Tachyum

Prodigy Delivers Key Requirements for Target Markets
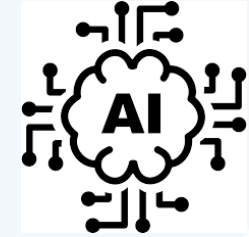
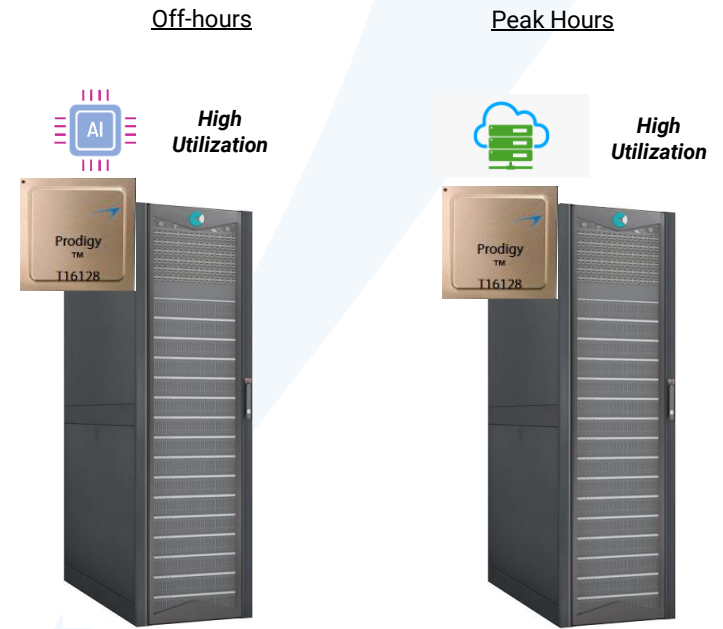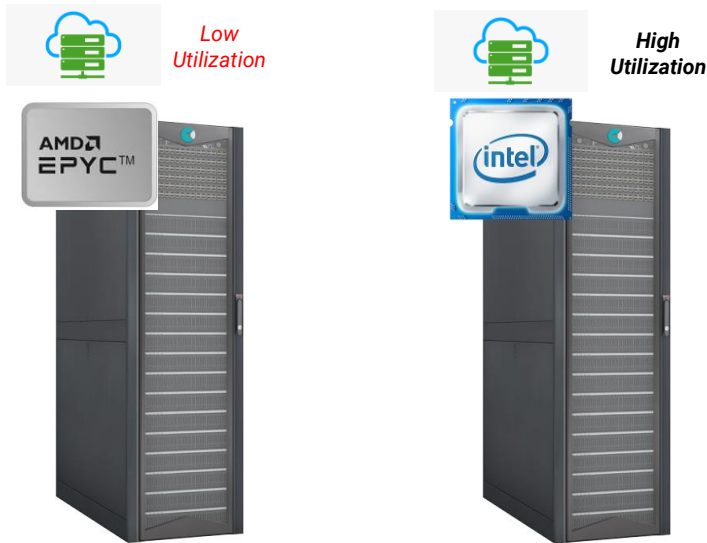|	Hyperscale/ Cloud	HPC	AI/ML
High Integer Performance	✔		
High Single-thread Performance	✔		
High Performance Parallel Processing	✔	✔	✔
High Memory Bandwidth	✔	✔	✔
Scalable, including large memory footprint	✔	✔	✔
Easy Deployment and Maintenance	✔	✔	✔
Cost and Power Efficient	✔	✔	✔
Special and compressed data types			✔
4-bit data pipes for inference – Int4			✔

Case Study: Repurposing Idle Servers in Hyperscale Data Center

Facebook web servers sit idle during off hours

Facebook Idle Servers In Day

Idle Servers

9pm   3am   9am   3pm

Off-hours   Peak Hours

Low Utilization

High Utilization
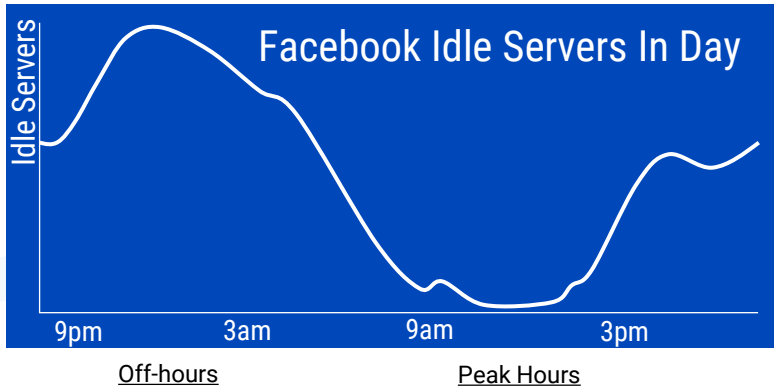
Prodigy keeps servers fully utilized 24/7

- AI/ML workloads during off hours
- Web servers during peak hours
- High efficiency, Low TCO

Off-hours   Peak Hours

High Utilization   High Utilization

Thank You!

visit

www.Tachyum.com