Tachyum Prodigy

The World's First Universal Processor

- Tachyum is developing the industry's first Universal Processor, AI, and supercomputing chip – Prodigy
- Prodigy has up to 18x higher performance and up to 6x better performance per watt than its competition
- Prodigy solves key issues plaguing today's data centers, including high power consumption, low server utilization, and the processor performance plateau that is limiting performance

Company and Product Overview

Tachyum is a semiconductor company developing the world's first Universal Processor, Prodigy, which unifies the functionality of CPU, GPGPU, and TPU into a single monolithic device, delivering unprecedented performance, power efficiency, and TCO reduction for a wide range of applications and workloads, including cloud AI, and HPC.

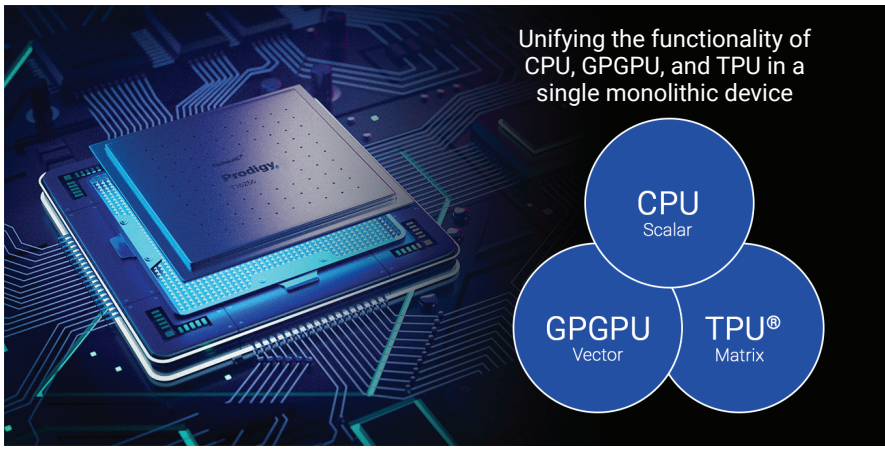
Prodigy's revolutionary architecture solves key issues that are plaguing today's data centers, including high power consumption and low server utilization. With Prodigy, hyperscale data centers can run cloud workloads during peak hours and AI workloads during off hours, keeping servers running 24/7.

Prodigy eliminates the need for costly and power-hungry accelerators, enabling high performance data centers to be deployed with a homogeneous architecture, enabling a simple software model and easy, straightforward maintenance.

In addition to running its native instruction set architecture, Prodigy also runs the binaries for x86, Arm, and RISC-V, providing fast, easy, out-of-the-box testing and evaluation.

Tachyum has approximately 100 employees with engineering teams in the Silicon Valley, Slovakia, and the Czech Republic, and the corporate office in Las Vegas, Nevada.

Tachyum Prodigy



Target Markets and SKUs

The Prodigy family of processors encompasses multiple product SKUs ranging from 256 to 64 cores with a wide range of performance, power, and features to address a wide array of important markets. Both the markets and SKUs are shown below, highlighting the Prodigy Series flexibility and ability to excel in a broad array of applications and workloads.

- HPC, Big AI
- Exascale Supercomputers
- Cloud / Hyperscale
- Edge / Telco
- Big Data, Analytics, Databases, Storage
- Crypto / Digital Currency

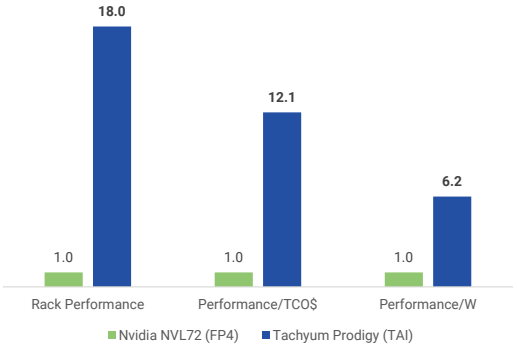
Prodigy SKU Summary

SKU	Cores	Frequency (GHz)	Scalability	Memory Controllers	PCIe Lanes	TDP (W)	Applications
T16256-AIX	256	5.7	4S	16	96	1030	Top-End, HPC, Big AI
T16240-AIM	240	5.7	4S	16	96	965	Top/Mid HPC, Big AI
T16240-AIE	240	4.5	4S	16	96	760	Entry HPC, Big AI, Crypto/Digital Currency
T16192-HT	192	4.5	4S	16	96	605	Analytics, Big Data, In-Memory Databases
T16120-HS	120	5.7	2S	16	96	480	Cloud, Databases, Edge/Telco
T8120-EN	120	4.5	1S	8	48	380	Entry Cloud, Databases, Storage
T864-LP	64	3.2	1S	8	48	140	Low Power, Hosting, Storage

Performance

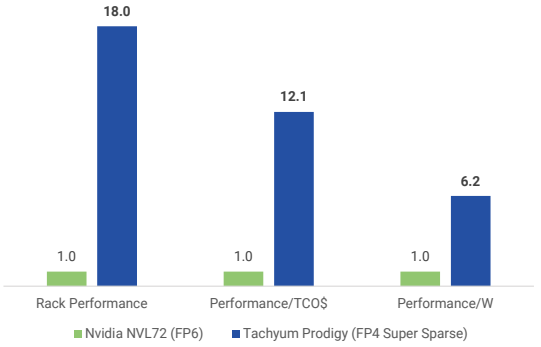
AI Processing Rack Performance vs. Nvidia

Prodigy TAI vs. NVL72 FP4

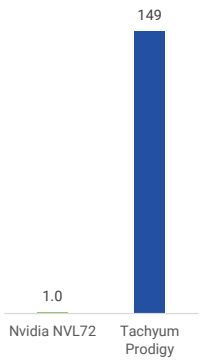


AI LLM Rack Performance and Memory Footprint vs. Nvidia

Prodigy FP4 Super Sparse vs. NVL72 FP6



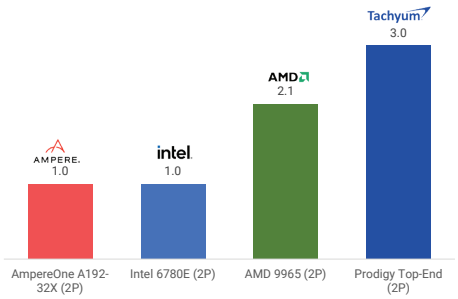
Maximum Rack Memory



Prodigy Racks Deliver **18x Higher AI Performance** and **150x More Memory** than Nvidia Blackwell NVL72

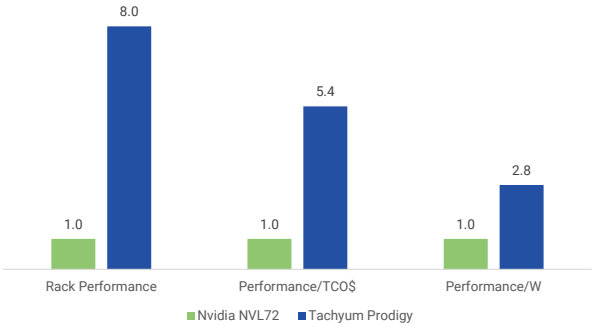
Cloud Performance vs. x86 and Arm

SPECrate 2017 Integer



HPC Rack Performance vs. Nvidia

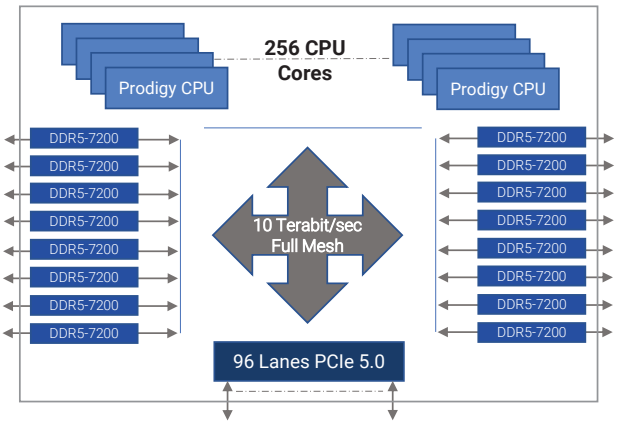
Prodigy vs. NVL72 64-bit Floating-Point Performance



Prodigy Racks Easily Switch from **3x Higher Cloud Performance** than x86 to **8x Higher HPC Performance** than Nvidia Blackwell NVL72

Device Architecture

Prodigy's groundbreaking universal processor architecture includes 256 high-performance CPU cores, 16 DDR5-7200 memory controllers, and 96 lanes of PCIe 5.0 that are connected with a 10 terabit/sec full mesh network, delivering processing power and high memory and I/O bandwidth that is designed to be balanced, optimizing system performance and avoiding bottlenecks. Prodigy will be manufactured in 5nm process technology. Key features and corresponding benefits are highlighted below.



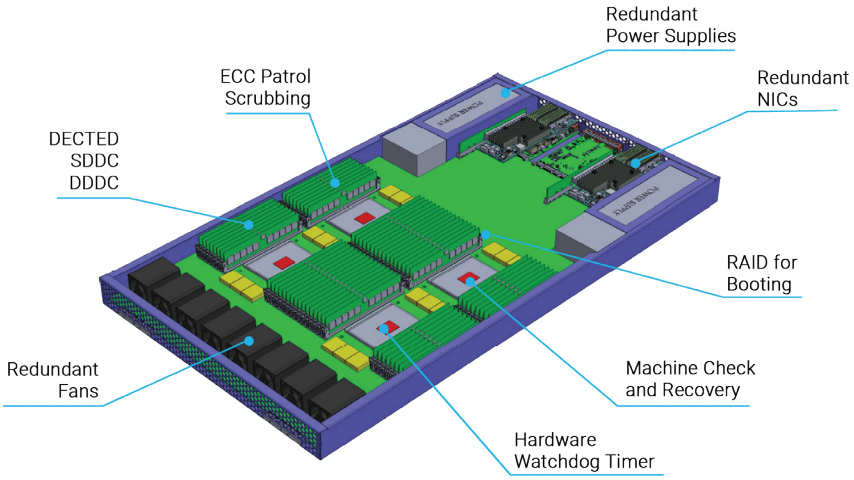
Prodigy Block Diagram

Key Features

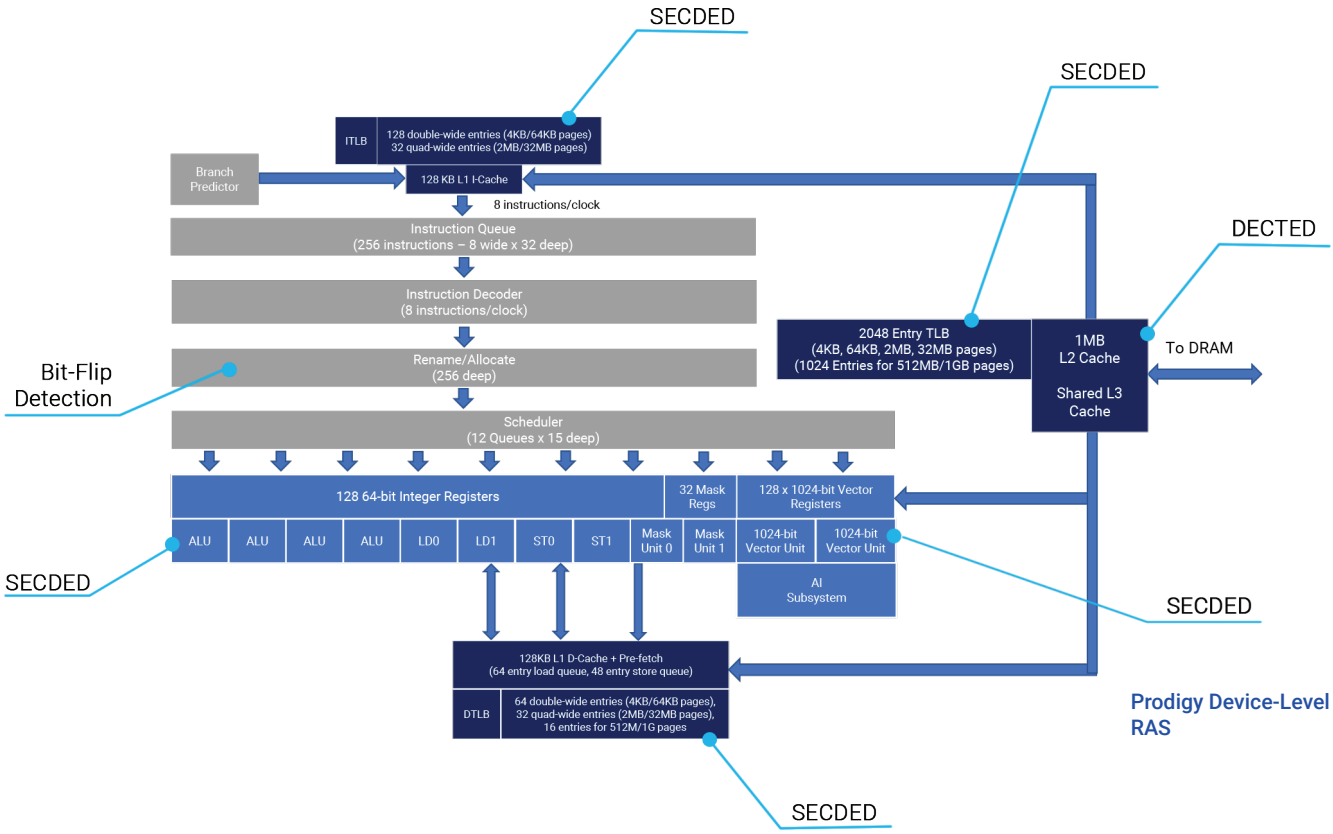
Features	Benefits
256 64-bit cores running at 5.7+ GHz	Highest performance for compute workloads
HW coherency supports 2 and 4-socket systems	High scalability for powerful compute nodes
16 DDR5-7200+ memory controllers	High memory capacity and bandwidth for LLMs
1TB / 2TB* of memory bandwidth	*Bandwidth amplification doubles bandwidth
96 lanes of PCIe 5.0 with 48 controllers	High performance NICs, large NVMe storage arrays
10 Terabit/sec Full Mesh	High bandwidth, low latency between subsystems
Runs native and x86, Arm, and RISC-V binaries	Fast, easy, out-of-the-box testing and evaluation
2 x 1024-bit vector units per core	High-performance HPC
2 x 1024-bit matrix processor per core	High-performance AI
FP64, FP32, TF32, BF16, Int8, FP8, TAI data types	Converged, homogeneous data centers
Sparsity and Super-Sparsity	Maximum AI performance and memory efficiency

Reliability, Availability, and Serviceability

Prodigy's revolutionary new architecture employs state-of-the-art technology which requires that the design be complemented by high reliability, availability, and serviceability, or RAS, to ensure that customer platforms are not only high performance but reliable and easy to service in the field. Prodigy's RAS strategy is very comprehensive, encompassing multiple facets at the silicon, platform, and system level to ensure Prodigy deployments provide high performance along with high reliability and availability at all levels. Prodigy's device RAS features are designed to detect and, if possible, seamlessly correct errors in the CPU's internal memory blocks and attached DDR modules caused by external events. Prodigy's RAS features are summarized in the associated diagrams.



Prodigy Platform and System-Level RAS



Prodigy Device-Level RAS

Post-Quantum Cryptography Support

In 2016 the National Institute of Standards and Technology (NIST) launched the post- quantum cryptography program (PQC) to address the threat with quantum computers and ensure that no QC can be used to break crypto codes. By 2022 four algorithms were selected and in 2023 three of the four had draft versions released and the fourth was expected by late 2024, so it should be released in 2025. In the information box there are the new PQC quantum-resistant algorithms that were selected in 2022.

Tachyum's world-class software engineering team has ported and verified all the new quantum-resistant asymmetric algorithms on Prodigy. The algorithms are running as part of Prodigy's standard software distribution for all customers and partners, and Tachyum continues to optimize them to ensure it has the fastest possible solution to deploy. Prodigy also supports the quantum-safe AES-256 which has already been optimized.

Key Establishment

- FIPS 203: ML-KEM or CRYSTALS-Kyber

Digital Signature

- FIPS 204: ML-DSA or CRYSTALS-Dilithium
- FIPS 205: SLH-DSA or Sphincs+
- FIP 206: Falcon

Prodigy's swift support for these vital new PQC algorithms underscores Tachyum's commitment to data security, ensuring that Prodigy-based systems will be future-proofed as new threats emerge.

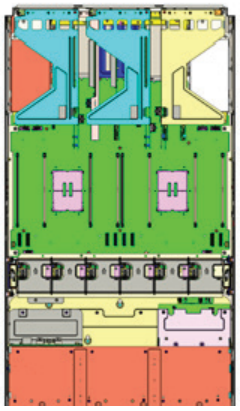
Platform Strategy

Prodigy's platform strategy includes two types of evaluation platforms. A standard, air-cooled 2-socket platform will be used for cloud and AI workloads, and for lead customers who require the highest performance, there is a liquid-cooled 4-socket platform that will be targeted for maximum AI and HPC performance. PCIe slots on both platforms support standard and OCP form factors.

The platforms support simple, out-of-the-box evaluation with an SDK that includes Linux, gcc compiler, software libraries, and a large ecosystem of recompiled native applications, streamlining software development.

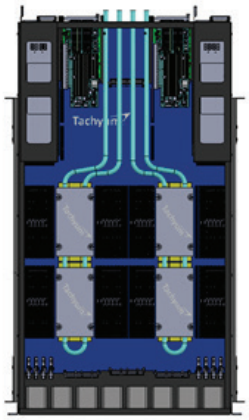
The evaluation platforms provide early customer access, and enable ODM/OEM partners to leverage the platforms to develop production designs and platforms.

Standard Platform for Cloud/AI



Air-Cooled 2-Socket Evaluation Platform

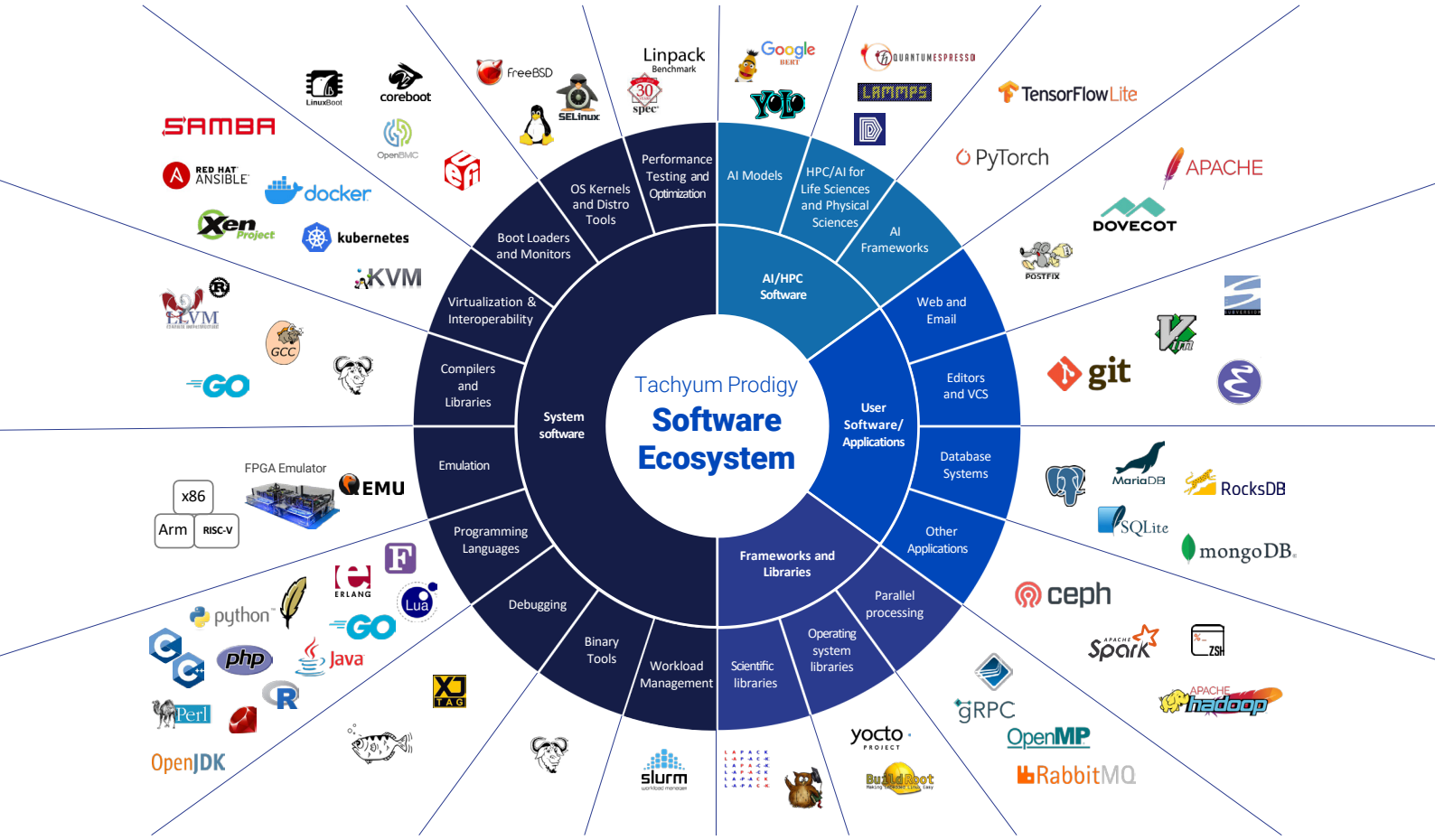
Lead Customer Platform for Maximum AI/HPC Performance



Liquid-Cooled 4-Socket Evaluation Platform

Prodigy Software Ecosystem

Prodigy has a rich ecosystem of development tools, operating systems, application software, and software libraries to enable fast, easy development and quick time to market.



Complete Software Ecosystem at www.tachyum.com/sw

Tachyum

www.tachyum.com

Tachyum Inc., 8275 South Eastern Ave, Ste 233, Las Vegas, NV 89123, U.S.A.
Tachyum s.r.o., Karadžičova 14, CBC IV., 3rd floor, 821 08 Bratislava, Slovakia

Scan the QR Code to discover our products

Brochure v2_250306