

NEWS RELEASE

## Tachyum Books Purchase Order to Build System with 25,000x ChatGPT4 Capacity and 25x Faster than Current Supercomputers

10/3/2023

LAS VEGAS--(BUSINESS WIRE)-- **Tachyum** ® announced that it has accepted a major purchase order from a US company to build a large-scale system, based on its 5nm Prodigy® Universal Processor chip, which delivers more than 50 exaflops performance that will exponentially exceed the computational capabilities of the fastest inference or generative AI supercomputers available anywhere in the world today.

Prodigy, the world's first Universal Processor, is engineered to transform the capacity, efficiency and economics of datacenters through its industry-leading performance for hyperscale, high-performance computing and AI workloads. When complete, the Prodigy-powered system will deliver a 25x multiplier vs. the world's fastest conventional supercomputer built just this year, and will achieve AI capabilities 25,000x larger than models for ChatGPT4.

Exascale supercomputers can quickly analyze and process massive amounts of data to solve complex problems that had previously been intractable. Prodigy overdelivers on traditional exascale capabilities by providing industry-leading performance, significantly reduced energy consumption, improved server utilization and space efficiency. Prodigy's exponential increases in memory, storage and compute to enable breakthroughs across datacenter, AI and HPC workloads in government, research and academia, business, manufacturing and other industries.

The human brain consists of around 100 billion neurons and approximately 200 trillion synaptic connections. Assuming a few bytes per synaptic connection, it would require 100's of terabytes of memory. Therefore, the Prodigy system with hundreds of petabytes of DRAM will have 100x more memory than needed.

Installation of this Prodigy-enabled solution will begin in 2024 and reach full capacity in 2025. Among the deliverables are:

- 8 Zettaflops AI training for big language models
- 16 Zettaflops of image and video processing
- Ability to fit more than 100,000x PALM2 530B parameter models OR 25,000x ChatGPT4 1.7T parameter models with base memory and 100,000x ChatGPT4 with 4x of base DRAM
- Upgradable memory of the base model system
- Hundreds of petabytes of DRAM and exabytes of flash-based primary storage
- 4-socket, liquid-cooled nodes connected to 400G RoCE ethernet, with the capability to double to an 800G all non-blocking and non-overprovisioned switching fabric

Tachyum's proprietary TPU® AI Inference IP supports Tachyum AI (TAI) data type and provides even more; breakthrough efficiency for video and large language model data formats that otherwise would require excessive power and expensive multipliers in matrix multiplication to achieve.

"The unprecedented scale and computational power required as part of this installation simply could not be provided by any chip manufacturer on the market today," said Dr. Radoslav Danilak, founder and CEO of Tachyum. "While there are startups receiving billions of dollars, based on their promise of achieving similar capabilities sometime in the future, only Tachyum is positioned to deliver the capability to economically build order-of-magnitude bigger machines that potentially enable the transition to cognitive AI, beginning later this year. This purchase order is a testament to our first-to-market position and our ability to provide a positive impact to worldwide AI markets."

As a Universal Processor offering utility for all workloads, Prodigy-powered data center servers can seamlessly and dynamically switch between computational domains (such as AI/ML, HPC, and cloud) on a single architecture. By eliminating the need for expensive dedicated AI hardware and dramatically increasing server utilization, Prodigy reduces CAPEX and OPEX significantly while delivering unprecedented data center performance, power, and economics. Prodigy integrates 192 high-performance custom-designed 64-bit compute cores, to deliver up to 4.5x the performance of the highest-performing x86 processors for cloud workloads, up to 3x that of the highest performing GPU for HPC, and 6x for AI applications.

## Follow Tachyum

<https://twitter.com/tachyum>

<https://www.linkedin.com/company/tachyum>

<https://www.facebook.com/Tachyum/>

## About Tachyum

Tachyum is transforming the economics of AI, HPC, public and private cloud workloads with Prodigy, the world's first Universal Processor. Prodigy unifies the functionality of a CPU, a GPGPU, and a TPU in a single processor that delivers industry-leading performance, cost, and power efficiency for both specialty and general-purpose computing. When hyperscale data centers are provisioned with Prodigy, all AI, HPC, and general-purpose applications can run on the same infrastructure, saving companies billions of dollars in hardware, footprint, and operational expenses. As global data center emissions contribute to a changing climate, and consume more than four percent of the world's electricity—projected to be 10 percent by 2030—the ultra-low power Prodigy Universal Processor is a potential breakthrough for satisfying the world's appetite for computing at a lower environmental cost. Prodigy, now in its final stages of testing and integration before volume manufacturing, is being adopted in prototype form by a rapidly growing customer base, and robust purchase orders signal a likely IPO in late 2024. Tachyum has offices in the United States and Slovakia. For more information, visit <https://www.tachyum.com/>.

Mark Smith  
JPR Communications  
818-398-1424  
[marks@jprcom.com](mailto:marks@jprcom.com)

Source: Tachyum