4 Exaflops AI Training 32 Racks Tachyum System

Dr. Radoslav Danilak, CEO of Tachyum
Tachyum 4 Exaflops AI Training in 32 Racks

Tachyum Prodigy
625 AI Teraflops

Tachyum 4 sockets
2.5 AI Petaflops

52U Rack
125 AI Petaflops

32 racks
4 AI Exaflops

Tachyum

• **Fabless Semiconductor Company**
  • Designing and manufacturing universal processor chips for servers, HPC and AI
  • Targeting hyperscale, telecommunications, service providers, HPC, AI and government customers
  • Tachyum will provide chips for its partners motherboards, servers, systems and solutions

• **Value Proposition**
  • 3x performance / price on CAPEX, and 4x lower TCO for the same performance
  • Universal processor chip unifies CPU, GPGPU and AI accelerators into a homogenous system
  • Up to 10x lower power consumption

• **Headquarters and Offices in Europe and USA**
  • Tachyum s.r.o., Karadžičova 14, 821 08 Bratislava, Slovak Republic
  • Tachyum Inc, 2520 Mission College Blvd, Suite 201, Santa Clara, CA 95054, USA
  • Additional offices planned in Asia for opening 1H 2021
Tachyum Architecture

• **Processor Performance Plateau**
  • 2005 3.8GHz Pentium-4
  • Transistors 30% faster every 2 years
  • Should be over 20GHz now

• **Tachyum Removed Wires From Critical Paths**
  • New architecture takes into account the physical distances
  • Avoids moving data to increase speed and save power
  • Scales performance with transistor speed

• **Universal Processor**
  • Unifies server processor, HPC and AI into single architecture
  • Moving from heterogenous to homogenous architectures
Tachyum Microarchitecture

• **Maximum issue rate per clock**
  • 2 x 512-bit multiply-add
  • 2 load + 1 store

• **Floating-Point/Integer execution units**
  • IEEE double, single and BF16 FPU
  • AI 8-bit floating-point data type
  • 2 x 512-bit multiply-add vector/matrix units

• **Vector and Matrix operations**
  • Matrix operations: 4x less power
  • 16b Int/FP 8x8, FP64, FP32 4x4
  • 8x8 matrix multiply-add = 1,024 Flops uses 6 source and 2 destination registers
  • Can increase performance 2x in the future
Tachyum Prodigy Is Universal Processor

- **Universal Processor For**
  - Hyperscale servers
  - Supercomputers (HPC)
  - AI Training and Inference

- **1st human brain sized AI**

- **Prodigy Processor**
  - Faster than GPU/TPU
  - 10x less power of Xeon
  - 1/3 cost
Tachyum Prodigy Universal Processor Chip

- **Faster than Xeon, and smaller than ARM**
  - Shorter wires with compiler that helps reduce delays
  - Legacy binaries run through our binary translations

- **64 cores, each faster than Xeon core**
  - 4GHz 7nm, 8 DDR5, 64 PCIE, 2 400G Ethernet
  - 32 and 16 core SKUs, single package with 2 x 64 cores

- **Performance**
  - 8 Teraflop of IEEE Double-Precision Floating-Point
  - 1 Petaflop on compressed training
  - 4 Peta-op on compressed inference
Reference Design

• **4 Socket standard size motherboard**
  - 4 Prodigy processors with 128 / 64 / 32 cores socket packages
  - 64 DDR5 RDIMM x 16-512GB for 32TB in a single node
  - Fits into a standard 19” rack as well as an Open Compute V3 rack
  - 2 x 400GE, 2 x 1Gb management ports, TPM, BMC
  - Air-cooled 2U or 1U chassis with 48V power supply

• **Software provided with motherboard**
  - Tachyum provides UEFI BIOS replacement
  - OpenBMC management software
  - Linux 5.8, GCC 9.4, DBG, KDB, LAMP & system applications
  - User application will be provided through the web separately
Chassis and Rack

• **2U and 1U air cooled full depth chassis**
  • Unified compute, storage or management node
  • 1-36 x 1-32TB NF1 SSD, 1-32TB DDR5 DRAM
  • For a standard 19” wide rack, with an adaptor for a OCP V3 rack
  • USB for KVM, VGA, and OpenBMC management
  • 2 x 400G Ethernet

• **19” Rack with 48” depth or OCP V3 rack**
  • 42” – 52” rack supported
  • 32-50 1U nodes or 16-25 2U nodes
  • 1U 128x100G or 256x100G switch in middle of the rack
  • Copper from servers to switch and fiber to spine switch
Networking and Storage

- **Networking**
  - 1U 128x100G or 256x100G switch in middle of the rack
  - 4-16 rack can be connected peer-to-peer using middle of the rack switches
  - 64 racks can be connected by 64 x QSFP-DD 400G (800G in 2022) 2U switch
  - 512 racks can be connected by 400G or 800G 21U CLOS switch
  - Tachyum will validate SONiC software for supported configurations

- **Storage**
  - Each chassis can host up to 1PB of flash with 36 NF1 drives x 32TB each
  - Storage can be shared as VNME or file storage using open source stack
Software

- **Tachyum provided ported software**
  - UEFI, Core-Boot, BMC, Linux, Device Drivers, GCC, debuggers, ...
  - Java, Python, TensorFlow, PyTorch, ..., LLVM and FreeBSD in 1Q2021
  - Open source applications like LAMP, Hadoop, Sparc, databases
  - Tachyum to support entire Linux distribution with applications over time

- **Open and customer owned source software**
  - Recompile to run at full speed or run binaries at 60-75% speed
  - Tachyum will help in porting customers based on support contracts
  - Tachyum partners, ISV, IHW and ecosystem to help customers

- **Legacy binaries**
  - X86, ARM V8, RISC-V binaries run through our software emulation
  - Allow mixing of native with emulated binaries
Availability

• **Reference Systems Availability 2H 2021**
  • Tachyum 4 socket reference design including rack, storage and networking integration
  • Engineering samples from Tachyum OEM/ODM and integrators
  • Production systems delivery and power-up in 4Q 2021

• **Visit Tachyum to See Reference Designs in Europe and USA**
  • Tachyum s.r.o., Karadžičova 14, 821 08 Bratislava, Slovak Republic
  • Tachyum Inc, 2520 Mission College Blvd, Suite 201, Santa Clara, CA 95054, USA
  • 4Q 2021 Slovak Academy of Sciences, Bratislava, Slovak Republic

• **Early Adopters and Strategic Partners**
  • Tachyum software emulation binary translator to X86 with Linux and development tools August 2020
  • Tachyum Prodigy FPGA emulation system available starting in October 2020
  • Tachyum prodigy sample chips and reference designs available in 2Q 2021
Tachyum 4 Exaflops AI Training in 32 Racks

Tachyum Prodigy
625 AI Teraflops

Tachyum 4 sockets
2.5 AI Petaflops

52U Rack
125 AI Petaflops

32 racks
4 AI Exaflops

Prodigy
Prodigy
Prodigy
Prodigy
Mgmt port + USB

36 NF1 SSD

4x400G

Fans

PS # 1
PS # 2

~ 34" Space for OCP
NIC + Standard PCIe Card
19"