Building Computers for AI with RISC-V
Tenstorrent builds **computers for AI**. Our mission is to address the open-source compute demands for software 2.0 through industry-leading AI/ML accelerators, high-performing RISC V-CPUs, and infinitely-configurable ML and CPU Chiplets.
Tenstorrent

- Founded in 2016
- $230M raised from Fidelity, Eclipse, Jim Keller and others
- 280 employees in 5 offices: Toronto, Santa Clara, Austin, Belgrade, Bangalore
- Team background: AMD, ARM, nVidia, Apple, Altera

- Two ML chips - Grayskull and Wormhole - ready for production, working on third
- Mid-way through the design of a high performance RISCV core
Jim Keller is the CEO of Tenstorrent and a veteran hardware engineer.

Prior to joining Tenstorrent, he served two years as Senior Vice President of Intel's Silicon Engineering Group. He has held roles as Tesla's Vice President of Autopilot and Low Voltage Hardware, Corporate Vice President and Chief Cores Architect at AMD, and Vice President of Engineering and Chief Architect at P.A. Semi, which was acquired by Apple.

Jim has led multiple successful silicon designs over the decades, from the DEC Alpha processors, to AMD K7/K8/K12, HyperTransport and the AMD Zen family, the Apple A4/A5 processors, and Tesla's self-driving car chip.
Tenstorrent Executive Team

Jim Keller
CEO: Digital Alpha processor, Apple A series, AMD Zen, Tesla Autonomous Driving system

Matt Matina
VP Machine: Learning ARM

Dan Bailey
Senior Fellow: Tesla, AMD, DEC

David Bennett
Chief CPU Architect: Apple, PA Semi, AMD

Wei-han Lien
Chief CPU Architect: Apple, PA Semi, AMD

Srikanth Arekapudi
RTL/DV Fellow: Cerebras, AMD

Keith Witek
Chief Operating Officer: Google, Tesla, AMD

Jim Montanaro
PD Fellow: Apple, AMD

Yasuo Ishii
Architecture Fellow: Arm, NEC
About Tenstorrent

• Based in North America, with offices in Tokyo, Toronto, Santa Clara, Austin, Belgrade, and Bengaluru.

• Tenstorrent builds the most innovative AI products:
  • Inference and Training, CNNs, LLMs, and NLPs
  • Powerful software stacks for models & bare metal programming

• Tenstorrent created the highest performing RISC-V CPU technology in the world

• Led by industry veteran hardware engineer and CPU architect, CEO Jim Keller.
Hardware Roadmap

2021

Grayskull
ML Processor

- 12nm, 276 TFLOP (FP8)

2022

Wormhole
Networked ML Processor

- 12nm, 328 TFLOP (FP8)
- 200 GB/S Scale-out Ethernet

2023

Black Hole
Standalone ML Computer

- 6nm
- SiFive RISC-V X-280
- Heterogenous compute

2024

Quasar
Low Power, Low Cost ML Chiplet

- ML Chiplet

Grendel
Highly Configurable and Performant ML Chiplet

- CPU + ML chiplets
## AI Solutions Portfolio

### Cards

**Available Now**
- e75 + e150 (Grayskull)
  - 75W-150W TDP
  - Up to 120 x 1GHz Tensix cores
  - 8GB LPDDR4 RAM, 120MB SRAM
  - 246-295 TOPS
  - Edge Deployment Device

**Shipping Q3 2023**
- n150 + n300 (Wormhole)
  - 150W-300W TDP
  - 1.6 TB onboard Ethernet
  - Embedded Workstations, Servers

### Systems

**Shipping Q4 2023**
- **T1000**
  - 1-300W Card
  - Up to 2TB RAM
  - Expansion slot for add-on cards

**Shipping Q4 2023**
- **T3000**
  - 2-300W cards (configurable to 4)
  - Up to 4TB RAM

**Servers**
- DT form factor, rackmount capable
- Entry Data Center Technology with easy installation

### Galaxy

**Shipping Q4 2023**
- **Galaxy 4U**
  - 32x Wormhole modules
  - 7 PetaOPS at BFP8
  - 41.6 Tbps internal connectivity
  - 384GB of globally accessible GDDR6 memory
  - 4GB SRAM
  - 7kW per server
  - Configurable + Scalable implementation

### Cloud

**Available Now**
- **Cloud Compute in Colo**
  - Customers currently utilizing self-serve cloud with ability to deploy API’s
  - Expanding model support and scope in mid-August
  - 20+ Galaxy systems in build to support broad customer base
Tenstorrent Software – Two Distinct Approaches

Buda: Run any model right away
- Great for production customers who want to get models up and running with ease
- Program in Pytorch, Tensorflow, …
- Automatic compilation and optimization

BudaM: Open Access to Tenstorrent Hardware & Software
- Low level hardware access, more like CUDA or Assembly
- Useful for HPC, C++ environments and low level Model development
- Open SW stack for 3rd party compilers and tools
Tenstorrent RISC-V O-o-O Processor Family

Performance

Open & Free

Higher Performance

4-Wide Decode
Sonic Boom with Vector

2-Wide Decode

3-Wide Decode

6-Wide Decode
Server, Laptop, and HPC

4-Wide Decode
Client and Edge

8-Wide Decode
Ascalon

Performance

Open & Free

Higher Performance

4-Wide Decode
Sonic Boom with Vector

2-Wide Decode

3-Wide Decode

6-Wide Decode
Server, Laptop, and HPC

4-Wide Decode
Client and Edge

8-Wide Decode
Ascalon
Scalar Competition Landscape (SPEC2K17INT)

SPEC2K17INT Rate 1

- Amazon Graviton 1: 2.06
- Amazon Graviton 2: 3.70
- Amazon Graviton 3 (projected): 4.63
- AMD Naples: 4.30
- AMD Rome: 4.56
- AMD Milan: 5.91
- AMD Genoa/Bergamo: 6.80
- AMD Zen5 (Projected): 8.84
- NVIDIA Grace (Projected): 7.44
- Intel Xeon: 5.73
- Sapphire Rapids 8480: 7.45

---

Frequency (GHz)

Power Usage (TDP)
Tenstorrent: Open Business Model

- Tenstorrent works with partners to design, create, modify, optimize heterogenous designs

- Key technology providers for wide spectrum of products for our strategy partners
  - AI
  - CPU

HPC/Cloud

Edge Cloud

Edge Devices
## Tenstorrent’s Offerings

<table>
<thead>
<tr>
<th>Product</th>
<th>IP</th>
<th>Chiplet</th>
<th>Chip</th>
<th>Card</th>
<th>Systems</th>
<th>Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>RISC-V</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI Acc</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
Why Tenstorrent is all in on RISC-V: Open and Customizable

Fully customizable / full ownership (you can brand it how you want)

Add datatypes, combined chiplets, or change the ISA (things x86 and ARM can’t)

Optimize performance for your specific workloads

No crazy license restrictions
This collaboration is just a beginning. Tenstorrent’s market leading AI and RISC-V CPU technologies will strengthen SoC competitiveness of LG’s future products while our long-time proven video codec technology will help Tenstorrent take control of data center high-performance processor markets.

LG Corp CTO Byoung-hoon Kim