TENSTORRENT: A UNIQUE AI STARTUP

KARL FREUND
CAMBRIAN-AI RESEARCH, LLC
APRIL 26, 2023

INTRODUCTION

It is hard to believe the difference a year makes. In 2021, there were over 100 public and venture-backed startups with the same mission, to compete with NVIDIA in producing the fast chips needed to create and run artificial intelligence (AI). Fast forward to 2023, and now many companies are struggling to gain market traction or acquire enough capital to keep going. Part of the problem is undoubtedly the global economy; many AI adopters and investors do not have the resources or courage to give new chips a chance. But the real culprit is NVIDIA; they are proving a lot harder to beat than many companies and their investors ever imagined. So why is Tenstorrent, a Toronto-based startup, any different? Why should we believe that Tenstorrent could succeed where so many are struggling and even failing? This paper will explore what distinguishes Tenstorrent from the scores of other startups from leadership, strategy, and technology perspectives.

THE BACKDROP IS SCARY

Do we really need yet another AI hardware startup? The industry has been flooded with over 100 such firms over the last five years. Some have already closed shop, realizing that NVIDIA technology for data center AI is tough to beat. Consequently, investors have become far more cautious. So, the competitive landscape for Tenstorrent may have gotten a little easier.

Some startups concentrate on areas below NVIDIA's radar, hoping to gain a toehold in the gold rush for Edge AI with low cost and power using techniques such as in-memory, analog, and neuromorphic computing. A few are refining their strategies and technologies to compete with NVIDIA in the data center, such as Groq, Samba Nova, Cerebras, and Graphcore. And five are large, publicly traded semiconductor companies with the deep pockets and the staying power needed to weather the storm: AMD, Amazon, AWS, Google, Intel, and Qualcomm. Realistically, all these companies are trying to vie for a spot as a second source for NVIDIA in AI data center processing for training and inference processing. Could they win? They would probably be thrilled if they could get 10% of the pie combined over the next three years.

Into this storm enters Tenstorrent, the Toronto-based AI Hardware startup with offices in the Bay Area, Austin, and recently Tokyo, Japan. Over the last year, the company has

Copyright ©2023 Cambrian-AI Research

begun to expand from early-phase research and development into becoming a real company, with marketing, sales, support, and functional area execs adding to the engineering talent the company has been recruiting. The company has now grown to over 280 employees. From a leadership perspective, the legendary CPU designer and early angel investor Jim Keller has recently assumed the CEO role, while founder Ljubisa Bajic has left the company. The company has also hired David Bennet as CCO, Keith Witek as COO, Matthew Mattina as VP of Machine Learning, Wei-Han Lien as Chief CPU Architect (from Apple), Mamoru Nakano to head sales in Japan, and Olaf Johansson as VP of Software.

Copyright ©2023 Cambrian-AI Research

THE EMERGING COMPUTE LANDSCAPE

A RICH OPPORTUNITY

The data center is evolving rapidly, with the move to cloud-based resources coinciding with the move to accelerators for HPC and AI. NVIDIA's data center revenue has recently received another shot in the Arm from the explosion of interest in ChatGPT and the new AI battle between Microsoft and Google for the future of search and the applicability of AI to productivity applications. Last quarter, NVIDIA data center sales reached $3.8B, growing at 31%. Markets and Markets projects that the global data center accelerator market size (measuring system revenue, not semiconductor revenue) could reach $64.0 billion by 2027, growing at a CAGR of 24.7% during the forecast period.

Meanwhile, the market for CPUs is starting to fracture considerably, with AMD growing from 10.7% to 17.6% in the data center in the latest quarter at Intel's expense. Meanwhile, Arm continues to grow, and Gartner projects to achieve some 19% of server shipments in 2026. Now, along comes open-source RISC-V, which is penetrating the microcontroller market and has ambitions in the data center as well. Semico Research projects RISC-V will account for 25 billion AI SoCs By 2027. That's a lot of silicon.

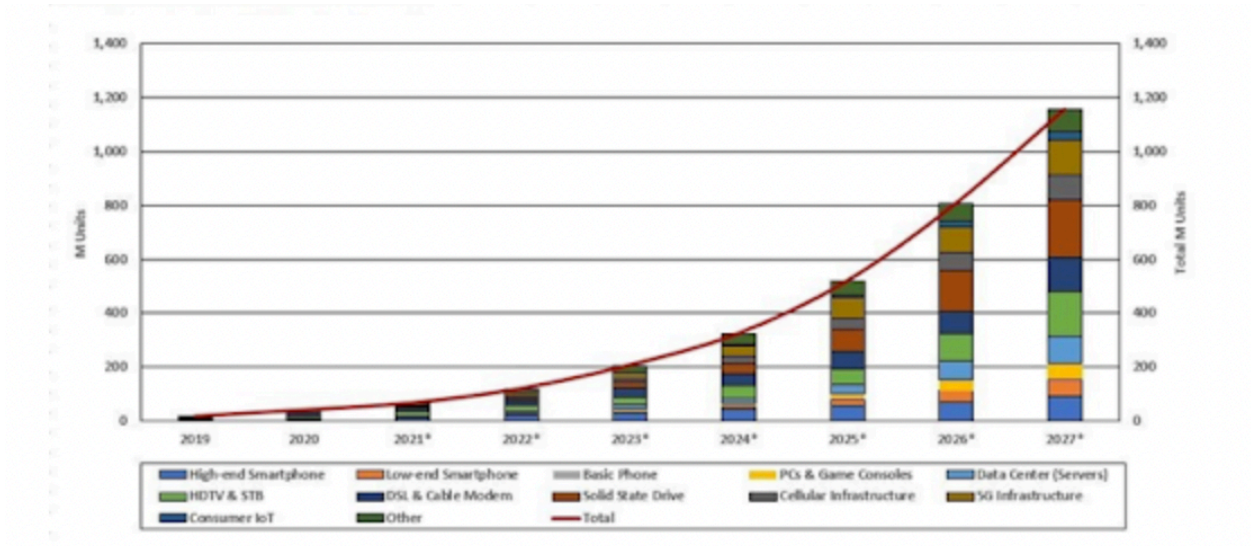

Figure 1: Total market projections for RISC-V CPUs, courtesy of Semico Research.

Since, as we shall see, the AI Accelerator SoCs will increasingly become integrated with CPU cores to perform scalar and management operations, Tenstorrent believes that it should develop its own RISC-V cores instead of relying on a third party. If Tenstorrent can deliver superior RISC-V cores, it could create a second revenue stream for IP or

Copyright ©2023 Cambrian-AI Research

silicon. So, when one combines the $64B in AI accelerators with the 25 billion RISC-V SoCs, the Tenstorrent opportunity becomes more evident.

THE COMPUTE CHALLENGE

Modern AI requires a massive amount of computing, especially when training the large language or foundation models that are quickly becoming the driving force behind AI adoption. Today, the required billions of trillions of calculations are performed primarily by GPUs, whose parallel processing is well-suited for the task. However, many startups believe their innovations will surpass NVIDIA GPUs since they are designed from the ground up for one workload: training and processing deep neural networks.

While that hypothesis sounds "sound," the endeavor has proven much more challenging than investors had anticipated. With the explosion in foundation models thanks to OpenAI ChatGPT, many startups now have to rethink their designs to handle the massive number of parameters needed in fast memory. A few companies have already gone out of business or have reverted to a second stealth mode while they rework their technology. Consequently, the AI HW Landscape has darkened for many startup competitors, while the more significant players (NVIDIA, AMD, Google, Intel, and now AWS) have solidified their offerings and roadmaps.

Given the challenge, some are abandoning the data center altogether, focusing on Edge AI for embedded accelerators and ASICs, a market many believe to lie under NVIDIA's radar. Others are changing their business models to become service providers. But nearly all are struggling to find the venture capital for their next round as the global economy begins to show weakness and new AI chips fail to gain adequate market traction.

Now, with OpenAI ChatGPT and Google's Lamba-based Bard, the computing challenge and opportunity have increased significantly. Tenstorrent stands to benefit from these, but only if they can hire the right engineers and execute the company's aggressive multi-prong strategy.

THE TENSTORRENT STRATEGY

Jim Kelly has worked closely with infrastructure buyers for years; he knows what the customer wants. Simply put, they want an open platform for AI compute that is easy to deploy at scale and provides a lower TCO than current alternatives. Tenstorrent has an AI chip today, a roadmap for the future that promises those benefits, and the credibility and track record of Mr. Keller to pull it off.

Copyright ©2023 Cambrian-AI Research

Tenstorrent is different from the rest of the field and may have a higher probability of success. First, the company has a software strategy that encourages open-source community innovation. Second, Tenstorrent is the only startup with an AI accelerator and RISC-V CPU designs and ambitions. Finally, Tenstorrent has attracted a world-class engineering team, and the company is now led by perhaps the industry's best-known CPU designer, Jim Keller.

Open Software

Writing AI software and porting AI models to a new chip can be labor intensive, especially when optimizing these AI tools to match a new hardware design. With NVIDIA, internal engineers (a.k.a. "ninjas") write new CUDA code, the company's closed-source library of math routines. With Tenstorrent, most optimizations, such as tensor fusion and linear algebra, can be written in high-level kernels in C++, which then call on a smaller set of closed low-level hardware-aware kernels.

AI and RISC-V Technologies

Second, as Jim Keller began to explore his options for the CPU cores in the company's AI SoC, he realized he could create better CPU cores with RISC-V than were available on the market. These fast cores could be built into larger CPU complexes to compete directly with X86 and Arm CPUs, should Tenstorrent decide to productize these designs or license them to large users. This strategy could provide Tenstorrent additional revenue streams unavailable to its competitors.

The Tenstorrent Team and Leadership

The company was founded in 2016 by Ljubisa Bajic, and his friend and colleague from AMD, Jim Keller, was an early advisor and angel investor. After stints at Tesla and Intel, Jim realized that his passion and drive to innovate could be satisfied by joining Tenstorrent full-time, and he became CEO in 2022. Jim's vision for the company, and his network of colleagues from the past, helped him build a strong engineering and leadership team, and the company continued to hire and grow while its competitors were laying people off. Jim's reputation will help the company continue to attract the best engineers.

One of the benefits of having Jim Keller on board is his reputation and network. Jim can open the doors to design engagements with the Super 7 data centers, enabling Tenstorrent to benefit from co-design with the largest buyers of IT tech and the leading researchers in AI. The team doesn't have to guess what, say, Meta or Google or Microsoft might say about a CPU or accelerator roadmap.

Copyright ©2023 Cambrian-AI Research

CONCLUSIONS

While we wish we had better insights into the eventual product mix and revenues Tenstorrent might realize, we believe that the RISC-V and AI Accelerators technology has performance and TCO advantages over the company's many competitors. AMD, Intel, and NVIDIA have already embraced the idea of combining GPU and CPU; all three point to the memory and throughput advantages a combination can offer. These advantages are significant to meeting large language models' needs for training and inference processing. Tenstorrent is the only startup we know of that can engage at this level of technology integration.

And having what could be the best RISC-V cores available to large buyers as IP or chiplets could create another revenue stream or an attractive exit strategy. Consequently, when anyone asks us who looks good to be able to compete in this new world, we invariably point to Tenstorrent as a likely winner.

Copyright ©2023 Cambrian-AI Research

Important Information About This Paper

Contributors and Publisher

Karl Freund, Founder and Principal Analyst, Cambrian-AI Research LLC

Inquiries

Contact us if you would like to discuss this report, and Cambrian-AI Research will respond promptly.

Citations

This paper can be cited by accredited press and analysts but must be mentioned in context, displaying author's name, author's title, and "Cambrian-AI Research". Non-press and non-analysts must receive prior written permission by Cambrian-AI Research for any citations.

Licensing

This document, including any supporting materials, is owned by Cambrian-AI Research. This publication may not be reproduced, distributed, or shared in any form without Cambrian-AI Research's prior written permission.

Disclosures

This document was developed with Tenstorrent Inc. funding and support. Although the document may utilize publicly available material from various vendors, it does not necessarily reflect the positions of such vendors on the issues addressed in this document.

Disclaimer

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Cambrian-AI Research disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Cambrian-AI Research and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

©2023 Cambrian-AI Research. Company and product names are used for informational purposes only and may be trademarks of their respective owners.