RISC-V®

InspireSemi™

Disruptive Next Generation Accelerated Computing Platform

Blistering speed, energy efficiency, versatility, and affordability for HPC/AI and graph analytics applications

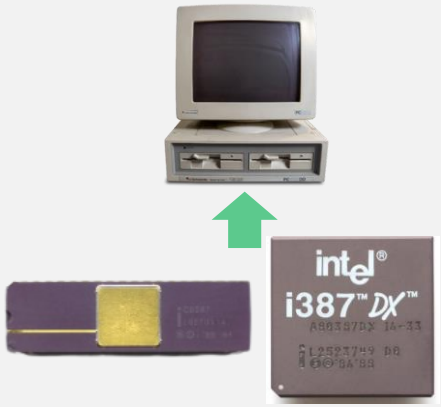
RISC-V Workshop at SC24

November 2024

The Third Wave of Accelerated Computing is Here

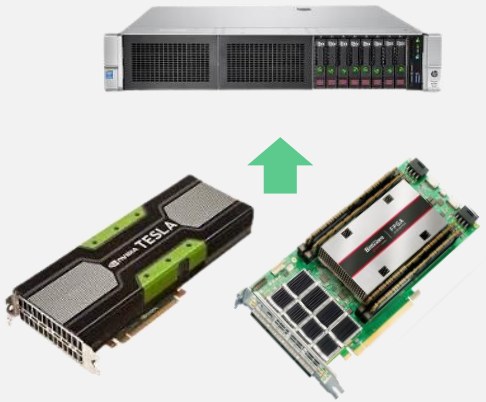
Thunderbird for HPC, AI, Graph Analytics

1980
Math Coprocessor



- Purpose-built widely applicable
- Open software ecosystem
- Plugs into existing computers

2007
GPU, FPGA



- Limited applications benefit
- Proprietary software model
- Plugs into existing servers

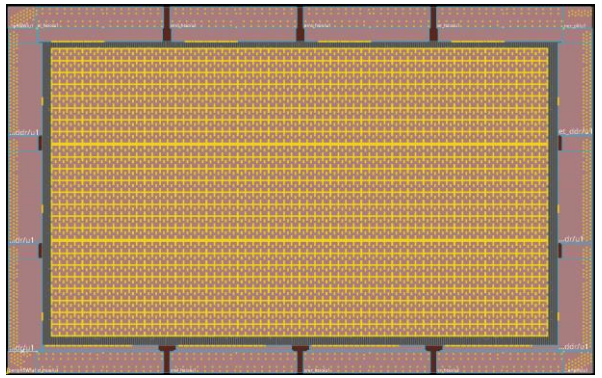
2024+
Thunderbird

- Built for HPC
- Versatile & open software ecosystem
- Plugs into existing servers

Thunderbird Accelerated Computing Solution

- ***Ultra-Efficient, Ultra-compact custom CPU cores***
  - Based on RISC-V instruction set (like ARM but open standard)
  - Modern superscalar, out-of-order, vector-capable cores
  - Ability to add custom instructions
- ***A supercomputer cluster-on-a-chip***
  - 1,536 high-perf 64-bit CPU cores per chip (>6,000 per PCIe card)
  - Comparable to GPU shader count but are independent CPUs
- ***Innovative low latency interconnect fabric***
  - Key to efficient utilization of so many cores
  - Seamlessly spans multiple-chip arrays up to 256 chips!
- ***Best in class energy efficiency***
  - Focus on performance/watt: 75 GLOPS/Watt (FP64)
  - Higher energy efficiency, fits in current datacenters
- ***Supporting existing open RISC-V software ecosystem***
  - Enables customers to easily adapt their software programs
  - Fast – no big investment or training required
- ***Recognized global partners to deliver turnkey solutions***
  - High-volume across multiple markets and geographies

RISC-V®



Lenovo

PENGUIN COMPUTING
A subsidiary of SMART Global Holdings, Inc.

World Wide Technology

POWERED BY DELL Technologies

E4 COMPUTER ENGINEERING

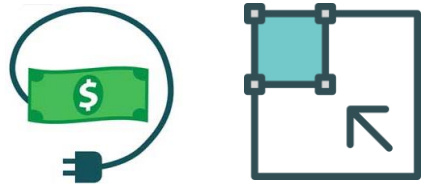
2crsi

Thunderbird Datacenter Impact:

Significant datacenter TCO savings and carbon footprint reduction

- Thunderbird has >20X CPU cores vs. Intel/AMD CPU cores per board
  - Intel and AMD CPUs: Latest have 256 bloated cores per board
  - Thunderbird I: 6,144 custom, low-power, low-area cores per board
- One Thunderbird card has more CPU cores than an ***entire rack of standard Intel/AMD severs!***
  - Less complexity, servers, power, HVAC, interconnects, points of failure, real estate

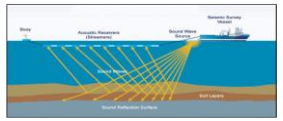
4U Server with 8 Thunderbird cards



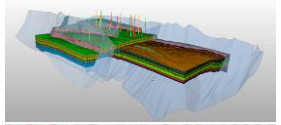
Addressing the Need to Accelerate All HPC & AI Software

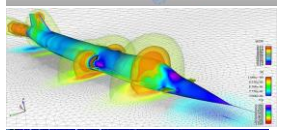
What customers always wanted…not "yet another GPU"

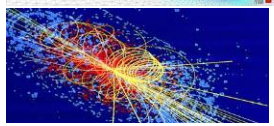

Financial simulations

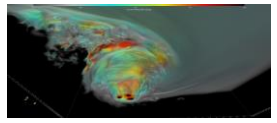

Geology: Seismic


Financial Trading & Graph Analytics

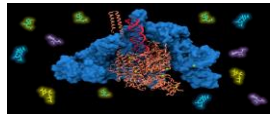

Energy: Reservoir Modeling & Sim


CAE/Computational Fluid Dynamics

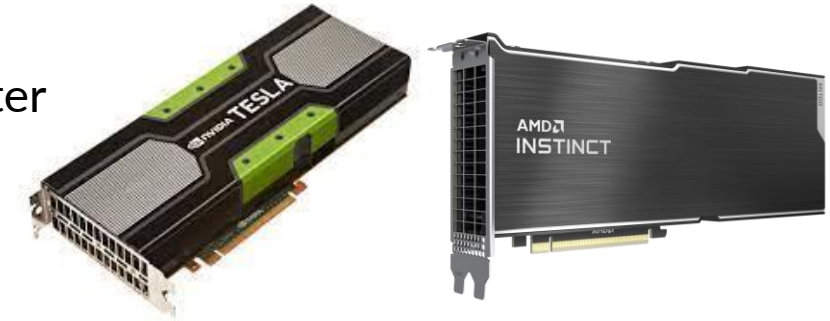

Government Lab Simulations


Climate & Weather Modeling


Cybersecurity & Cryptography


Genomics, Pharma, Life Sciences

Datacenter GPUs



Datacenter GPUs do it all, right?

InspireSemi Thunderbird

Highly differentiated "supercomputer-cluster-on-a-chip"

- Versatile platform delivers unprecedented capability
- 4 chip PCIe card delivers >6,000 64-bit CPU cores (FP64)
- Innovative high bandwidth, low-latency on-chip network
- Best-in-class for both Performance/$ and Watt
- Large scale computing power with much lower TCO - can replace many racks of servers and expensive high-speed networking

Open Software Ecosystem Solves Customer Porting Challenges

- Supports standard operating systems (unlike competing compute accelerators)
  - **Linux!**
  - Real-Time Operating Systems (RTOS)
  - Kitten lightweight kernel (LWK for DOE labs)
  - Eliminates need to learn and use proprietary software stacks
  - Eliminates vendor lock-in
- Uses standard CPU-style programming models
  - No need for CUDA, ROCM, etc. that GPUs require
  - No need for disruptive software algorithm rewrites
  - Standard compiler, OpenMP, MPI, etc. approaches
- Leverages key RISC-V software ecosystem
  - Address-accurate QEMU model
  - Standard compilers, e.g.- GCC, Gfortran, GDB toolchains
  - Standard HPC libraries, e.g. – BLAS, LAPACK, FFTW

Thunderbird Performance Specifications

6 – 8 TFLOPS
per chip (FP64)
24 – 32 TFLOPS
per card (FP64)

75 GFLOPS/ Watt
(FP64)

20x
reduction in latency

30 – 60%
lower power consumptions

Early customer feedback is unlike GPUs, Thunderbird is likely to hit its peak performance numbers due to its architecture and interconnect technology

InspireSemi Thunderbird in the News

techradar.

Supercomputer-on-a-chip goes live: single PCIe card packs more than 6,000 RISC-V cores, with the ability to scale to more than 360,000 cores

- InspireSemi has announced the successful tapeout of the Thunderbird I Accelerated Computing chip for fabrication at TSMC

tom's HARDWARE

Thunderbird packs up to 6,144 CPU cores into a single AI accelerator and scales up to 360,000 cores — InspireSemi's RISC-V 'supercomputer-cluster-on-a-chip' touts higher performance than Nvidia GPUs

- The Holy Grail of supercomputing chip design is an architecture that combines the versatility and programmability of CPUs with the explicit parallelism of GPUs, and InspireSemi strives to achieve just that

TECHSPOT

Move over GPUs, with 1,536 cores the Thunderbird RISC-V CPU is ready to eat your lunch

- Open source enables small industries to participate in the accelerator boom

ALL ABOUT CIRCUITS

InspireSemi Announces Tapeout of Thunderbird Accelerated Computing Chip

- InspireSemi's new compute chip couples the parallel processing of GPUs with the versatility of CPUs

JPR Jon Peddie Research

InspireSemi announces tape-out of RISC-V HPC chip

- InspireSemi's Thunderbird I RISC-V chip offers high-performance computing for underserved applications, with an emphasis on energy efficiency and competitive pricing

Thunderbird Addresses ALL HPC & AI Customer Needs

	InspireSemi Thunderbird	CPU	GPU	FPGA	AI Accelerators
Architecture	Many programs, many data streams	Few programs, few data streams	Few programs, many data streams	Programmable logic elements	Single program, many data streams
Performance	High for broad range of HPC apps	Slow, need h/w accelerators	High for AI and some HPC apps	Medium	High for AI only
Cost	Low $6,500 for 2 chip PCIe card	High ~$1K-8K (+ more servers)	High ~$7K-48K	High $8K-$10K	High ~$10K - $2.2M
Energy consumption	Low ~50W/chip (150W max)	Med 290W+/chip (+ more servers)	High 700-1,000W	High ~300W	High ~300W – 20kW
Scalability	256 chips	1-4 chips	2-8 chips	1 chip	1-2 chips
Programming model	Standard CPU-like, Any language, Full instruction set	Standard CPU, Any language, Full instruction set	Specialized C variant (CUDA, ROCM, SYCL)	Hardware description language	Proprietary, obscure
Software ecosystem	Open-source, Linux, compilers, libraries, AI frameworks, existing applications	Robust	Limited, proprietary	None	AI frameworks and proprietary software stacks

Disclaimer

This presentation contains statements which constitute “forward-looking information” within the meaning of applicable securities laws, including statements regarding the plans, intentions, beliefs and current expectations of InspireSemi with respect to future business activities and operating performance.

Often, but not always, forward-looking information can be identified by the use of words such as “plans”, “expects”, “is expected”, “budget”, “scheduled”, “estimates”, “forecasts”, “intends”, “anticipates”, or “believes” or variations (including negative variations) of such words and phrases, or statements formed in the future tense or indicating that certain actions, events or results “may”, “could”, “would”, “might” or “will” (or other variations of the forgoing) be taken, occur, be achieved, or come to pass. Forward-looking information includes, but is not limited to, information regarding: (i) the business plans and expectations of the Company including expectations with respect to production and development; and (ii) expectations for other economic, business, and/or competitive factors. Forward-looking information is based on currently available competitive, financial and economic data and operating plans, strategies or beliefs as of the date of this presentation, but involve known and unknown risks, uncertainties, assumptions and other factors that may cause the actual results, performance or achievements of InspireSemi, to be materially different from any future results, performance or achievements expressed or implied by the forward-looking information. Such factors may be based on information currently available to InspireSemi, including information obtained from third-party industry analysts and other third-party sources, and are based on management’s current expectations or beliefs. Any and all forward-looking information contained in this presentation is expressly qualified by this cautionary statement.

Investors are cautioned that forward-looking information is not based on historical facts but instead reflect InspireSemi’s management’s expectations, estimates or projections concerning future results or events based on the opinions, assumptions and estimates of management considered reasonable at the date the statements are made. Forward-looking information reflects InspireSemi’s current beliefs and is based on information currently available to it and on assumptions it believes to be not unreasonable in light of all of the circumstances. In some instances, material factors or assumptions are discussed in this presentation in connection with statements containing forward-looking information. Such material factors and assumptions include, but are not limited to: the impact of the COVID-19 pandemic on the Transaction or the company; the ongoing conflict between Russia and Ukraine and any actions taken by other countries in response thereto, such as sanctions or export controls; and anticipated and unanticipated costs and other factors referenced in this presentation and the Filing Statement, including, but not limited to, those set forth in the Filing Statement under the caption “Risk Factors”. Although the Company has attempted to identify important factors that could cause actual actions, events or results to differ materially from those described in forward-looking information, there may be other factors that cause actions, events or results to differ from those anticipated, estimated or intended. Forward-looking information contained herein is made as of the date of this presentation and, other than as required by law, the Company disclaims any obligation to update any forward-looking information, whether as a result of new information, future events or results or otherwise. There can be no assurance that forward-looking information will prove to be accurate, as actual results and future events could differ materially from those anticipated in such statements. Accordingly, readers should not place undue reliance on forward-looking information. Should one or more of these risks or uncertainties materialize, or should assumptions underlying the forward-looking information prove incorrect, actual results may vary materially from those described herein as intended, planned, anticipated, believed, estimated or expected.