The ANTON 2 Chip

A Second-Generation ASIC for Molecular Dynamics

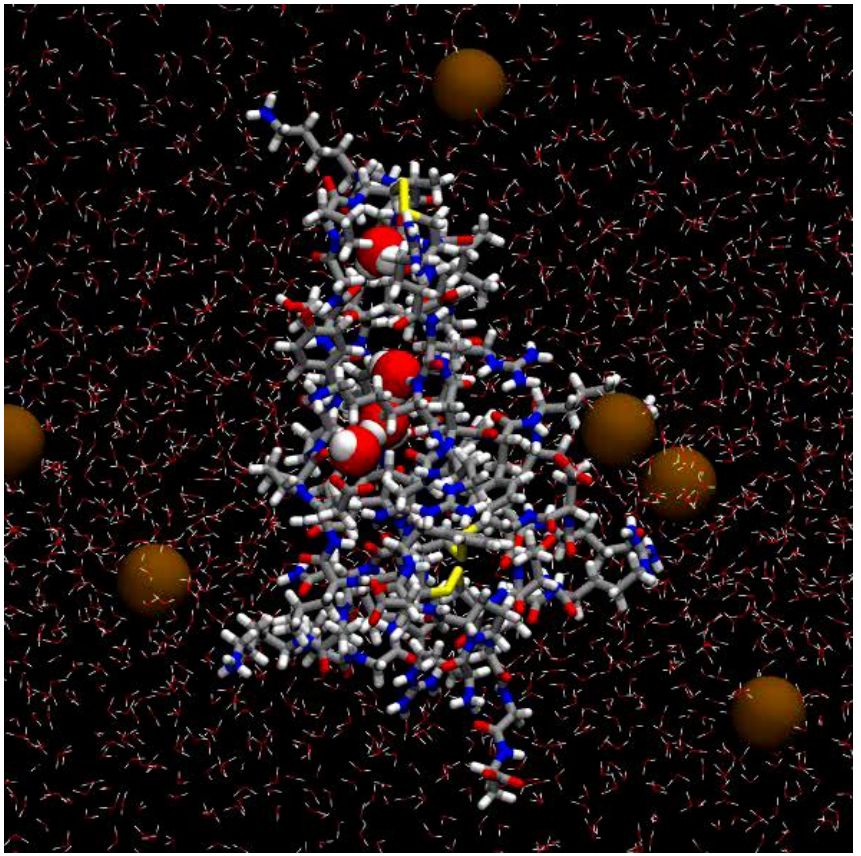
The Hardware Team: All 2^5 of Us

J. Adam Butts, Brannon Batson, Jack C. Chao*, Martin M. Deneroff*, Ron O. Dror*, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, J.P. Grossman, C. Richard Ho*, Jeffrey S. Kuskin, Richard H. Larson*, Timothy Layman*, Li-Siang Lee*, Chester Li*, Shark Yeuk-Hai Mok*, Mark A. Moraes, Rolf Mueller*, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot*, Naseer Siddique, Jochen Spengler, Ping Tak Peter Tang*, Michael Theobald, Horia Toma*, Brian Towles, Stanley C. Wang, and David E. Shaw

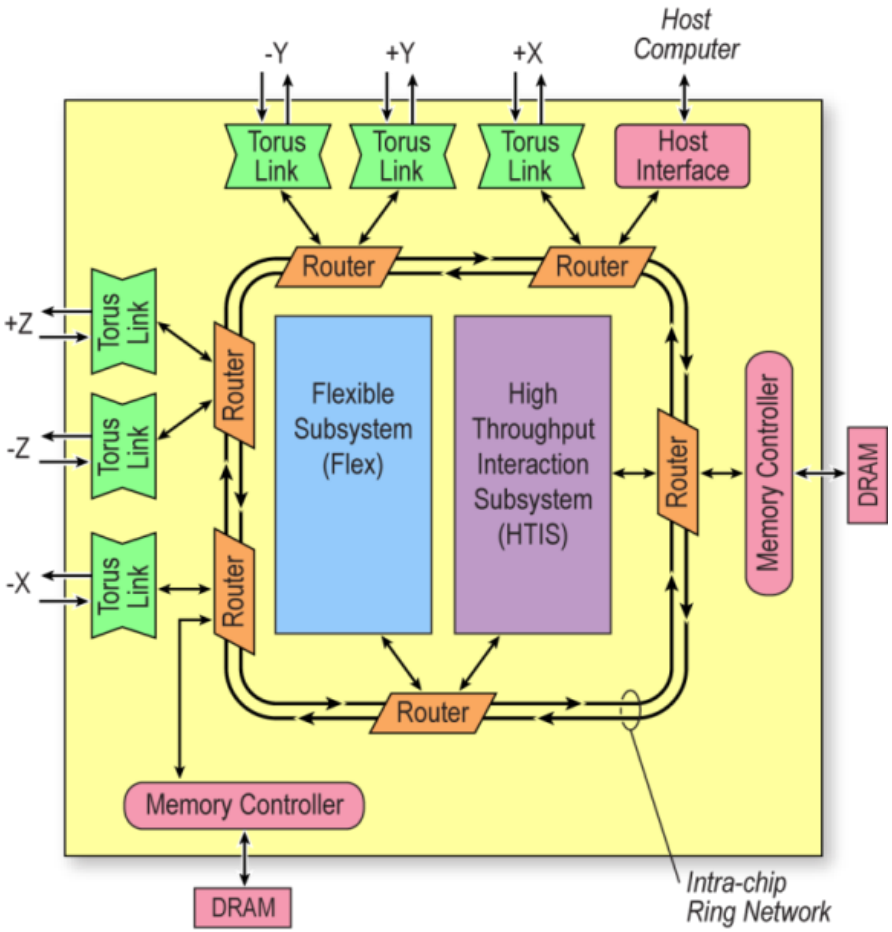
* Work conducted while at D. E. Shaw Research; author's affiliation has subsequently changed.

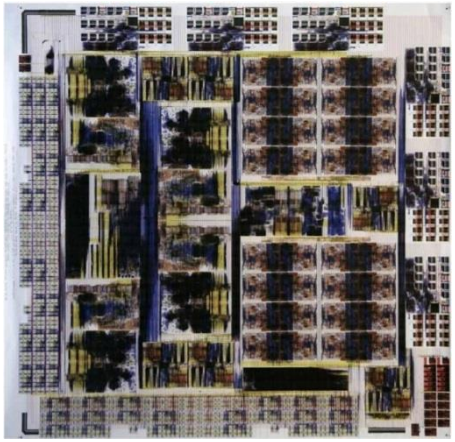
Molecular Dynamics (MD) and Why It's Hard

- Simulation of the motions of all atoms in a *biochemical system*
  - Calculate all forces on each atom of the system in each discrete time step
  - Each simulation time step is ~2 fs
- Efficient parallelization of force calculation is hard
  - Time steps are sequentially dependent
  - All atoms interact with all other atoms
- Massive computational task!
  - Many interesting biochemical systems have 10^5 – 10^6 atoms
  - Many biological processes take milliseconds (i.e., 10^{12} time steps)
  - ~10^4 FLOP/atom/step, even with range-limited approximations



ANTON: The Original Series*





* D. E. Shaw *et al*., "Millisecond-Scale Molecular Dynamics Simulations on Anton," in *Proc. Supercomputing (SC-09)*, Nov. 2009.

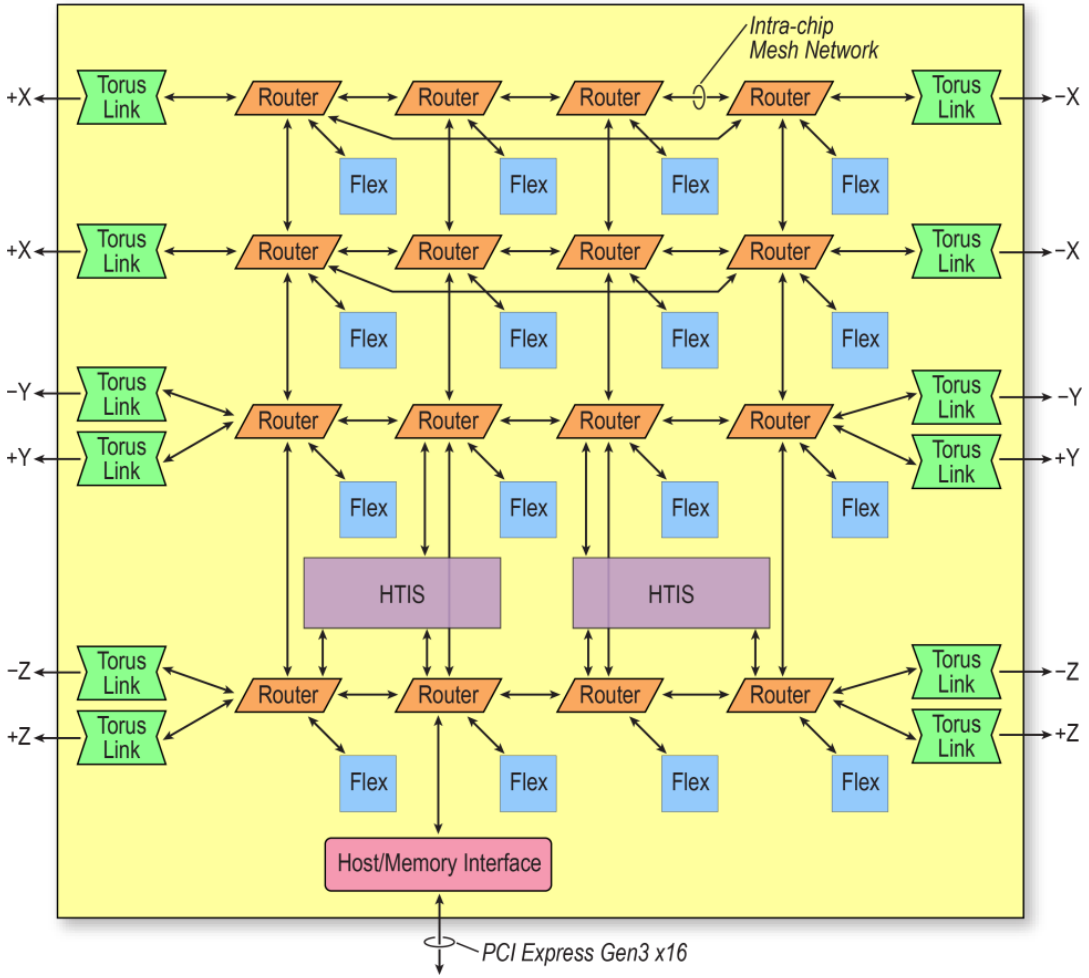
ANTON 2: The Next Generation

- Increased speed
  - More cores, pipelines
  - Higher clock frequency
  - In-house physical design
- Enhanced capability
  - More flexible fixed-function pipelines
  - Higher capacity (atoms/ASIC)
- Simplified software
  - Single ISA for all CPU cores
  - Optimized compiler port (gcc)

	ANTON	ANTON 2
CPU cores	13	66
CPU clock	485 MHz	1,650 MHz
Interaction pipelines	66	152
Pipeline clock	970 MHz	1,650 MHz
Peak throughput*	2.73 TFXOPS	12.7 TFXOPS
Main SRAM	128 KB	4,096 KB
Atoms/ASIC	460	8,000
Channel B/W	607 Gb/s	2.7 Tb/s

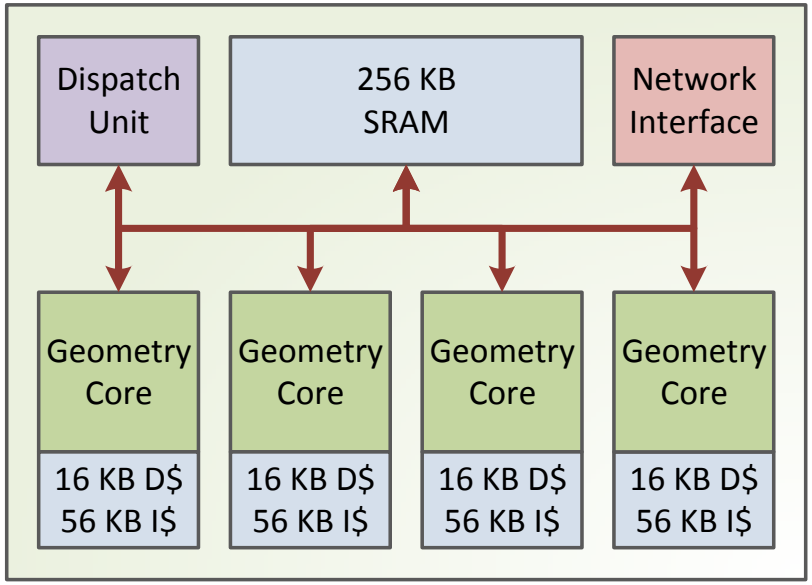
*TFXOPS = 10^{12} 32-bit fixed-point operations per second.

ANTON 2 ASIC Architecture



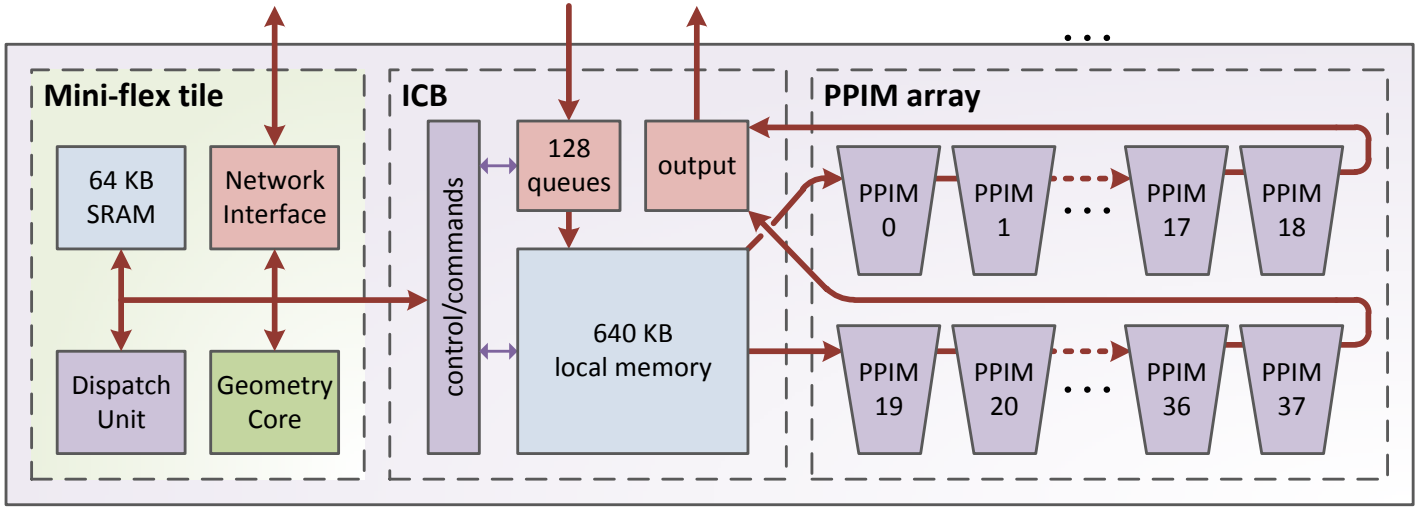
Flexible Subsystem Tile (Flex)

- 4 geometry cores (GCs)
  - Fixed-point SIMD processors
  - Compiler-friendly ISA optimized for MD
  - Large instruction cache
- Dispatch unit*
  - Low-latency hardware event synchronization
  - Enables exploiting fine-grained parallelism
- Globally shared SRAM
  - Stores particle data, additional code
  - Built-in atomic operations
- In-tile network with gateway to on-chip mesh



* J.P. Grossman *et al*., "Hardware support for Fine-Grained Event-Driven Computation in Anton 2," in *Proc. ASPLOS-18*, Mar. 2013.

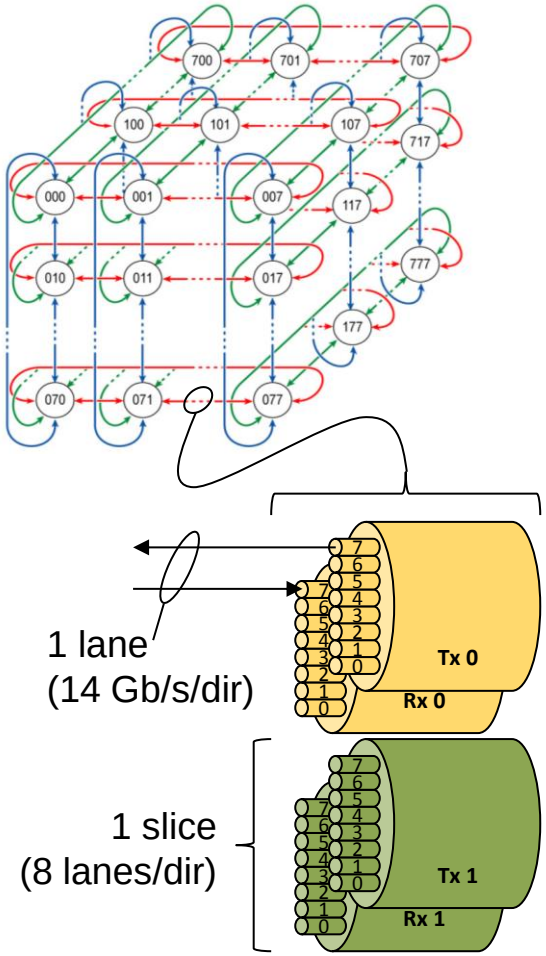
High-Throughput Interaction Subsystem (HTIS)



- Pairwise Point Interaction Modules (PPIMs)
  - Two specialized compute pipelines each
  - Over 250 billion interactions/s/ASIC
  - Significant flexibility
    - Configurable pair selection
    - Flexible interaction functions
    - Particle/pair parameter overrides
- Single-GC “mini-flex” tile for control
- Interaction Control Block (ICB)
  - Queues for receiving different incoming particle streams
  - Streaming memory supplies high-bandwidth to PPIM array
- Output block merges results from two PPIM rows

Network Architecture*

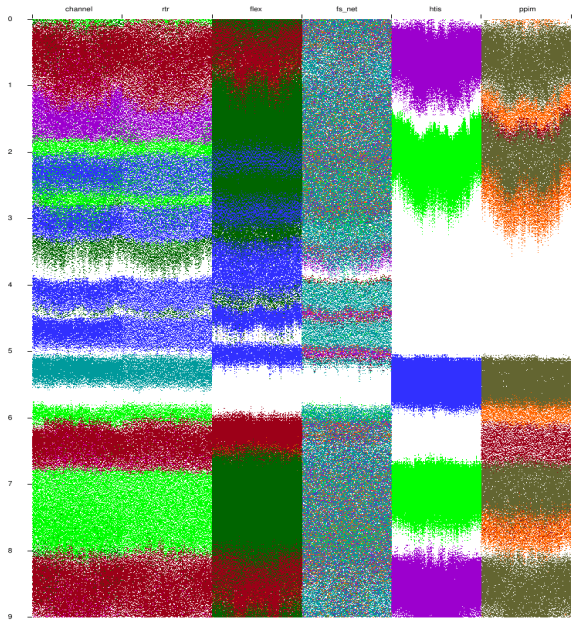
- On-chip mesh network
  - 2.5 Tb/s bisection bandwidth
  - Also carries inter-chip through traffic
- Inter-chip 3D torus network
  - Mirrors spatial geometry of MD
  - Network sliced for load-balancing
  - 96 bi-directional lanes @ 14 Gb/s
    - 2.7 Tb/s total bandwidth / ASIC
    - 512-node bisection bandwidth of 57 Tb/s
  - Custom channel protocol
- Optimized for exploiting parallelism in MD
  - Arbitrary dimension-order routing
  - Table-based multicast
  - In-network reductions



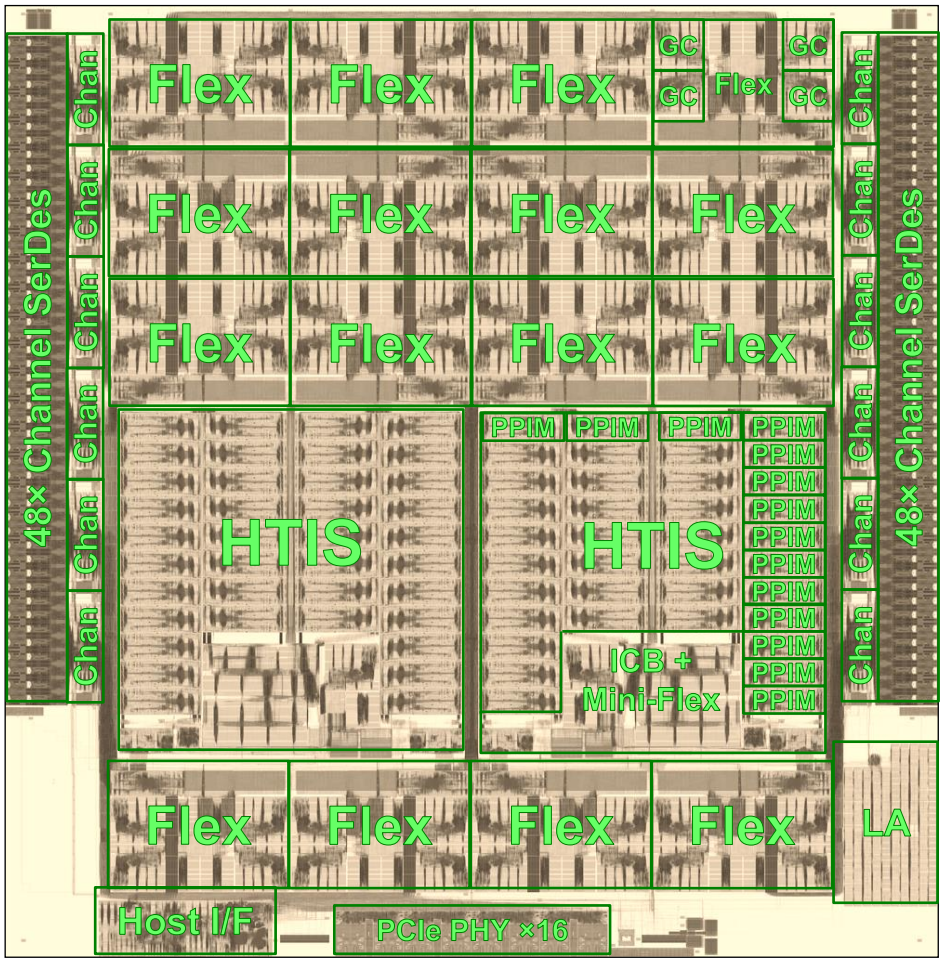
* B. Towles *et al.*, "Unifying on-chip and inter-node switching within the Anton 2 network," in *Proc. ISCA-41,* Jun. 2014.

Host and Debug Interfaces

- Standard Host Interface
  - High-bandwidth PCIe Gen3 ×16
  - Direct handling of GC cache misses to host-mapped memory
  - Supports DMA to/from on-chip Flex tile memory
  - Full-swing CMOS interface for configuration and debug
- On-chip Logic Analyzer (LA)
  - Flexible triggering and monitoring of on-chip state
  - Trace capacity of over 650,000 cycles
  - Extensively used for performance tuning

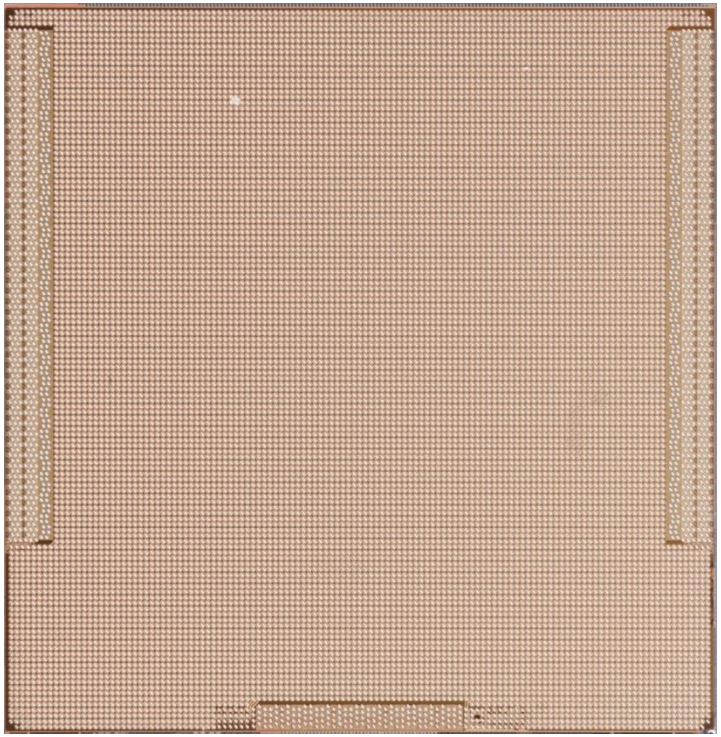


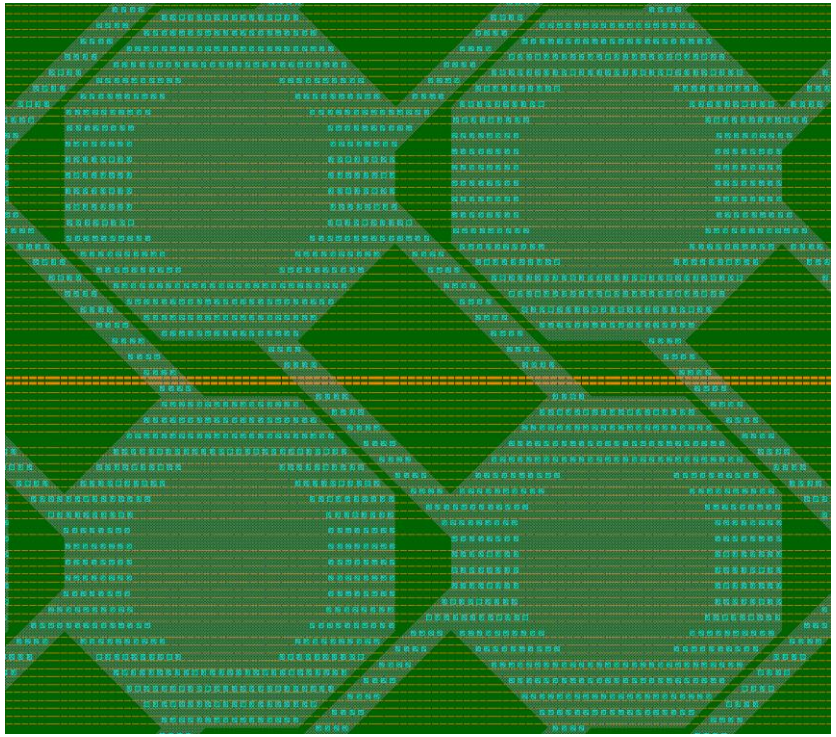
ANTON 2 ASIC



CMOS process	40 nm
Die size	408 mm^2
Total transistors	2.0 G
Gates	190 M
SRAM bits	120 M
Placed instances	51 M
Power dissipation	190 W
TDP	250 W

High Power: On-Die

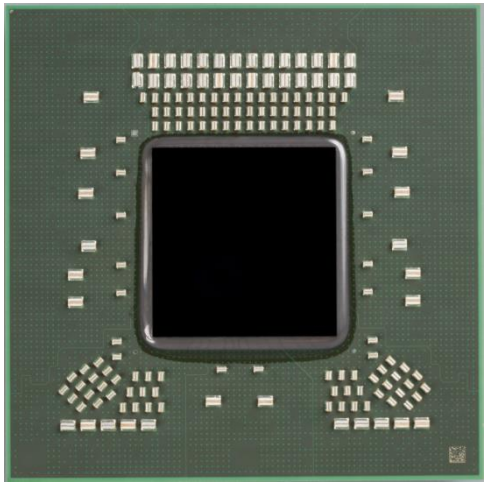


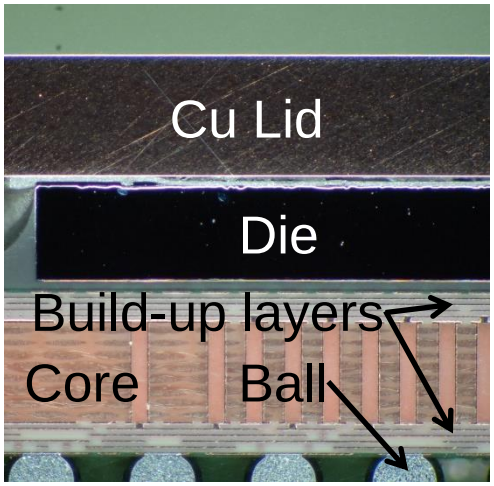


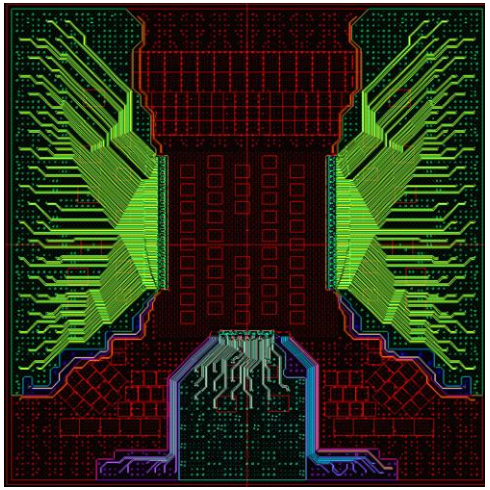
- > 14,500 core supply bumps
- Custom decoupling capacitors
- Careful attention to cell density, power grid continuity

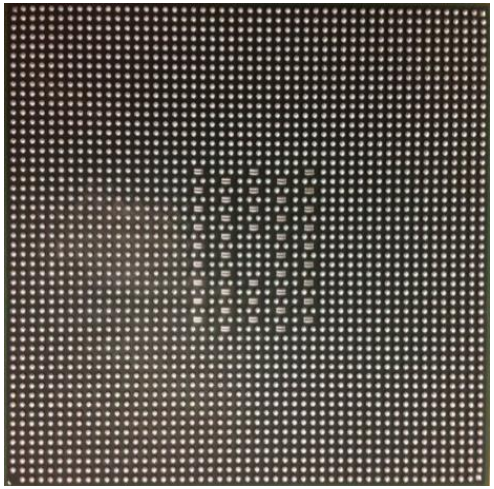
- > 25% of routable metal tracks for core power delivery
- Distribution layers use 98.5% of allowable metal for power delivery

High Power: Package





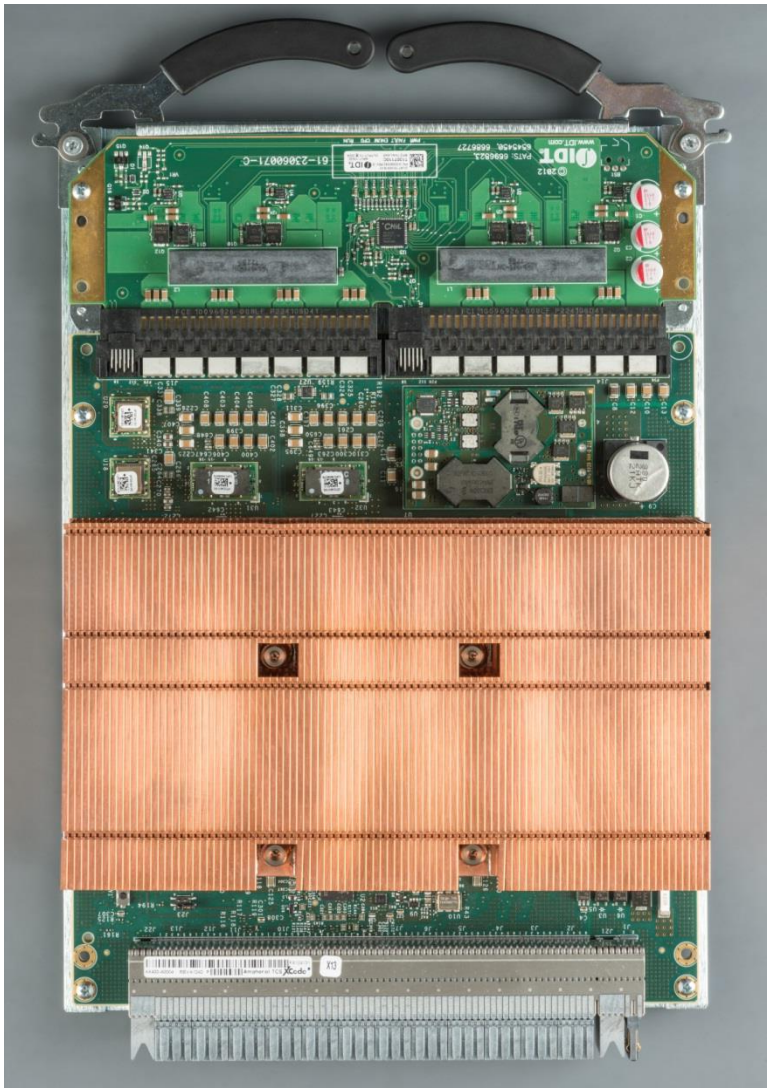




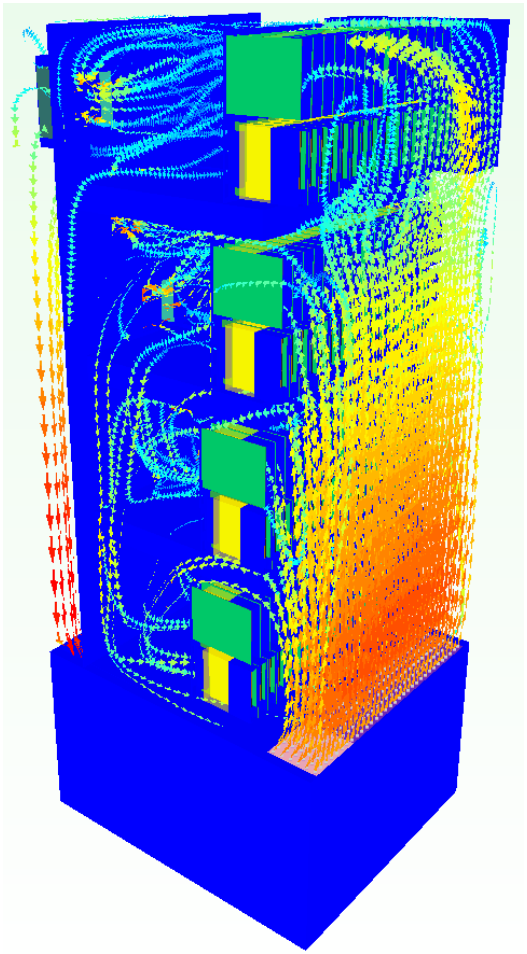
- Organic flip-chip ball grid array
  - 12 routing layers
  - 2511 pins (>40% supply)
  - Copper heat spreader lid Θ_{JC} ~ 0.08 °C/W
- Power integrity
  - > 200 package capacitors
  - 183 core supply capacitors
  - < 1 mΩ worst-case impedance
- Signal integrity
  - < 1.5 dB insertion loss
  - < −50 dB near- or far-end crosstalk
  - Non-overlapping core and analog supply planes

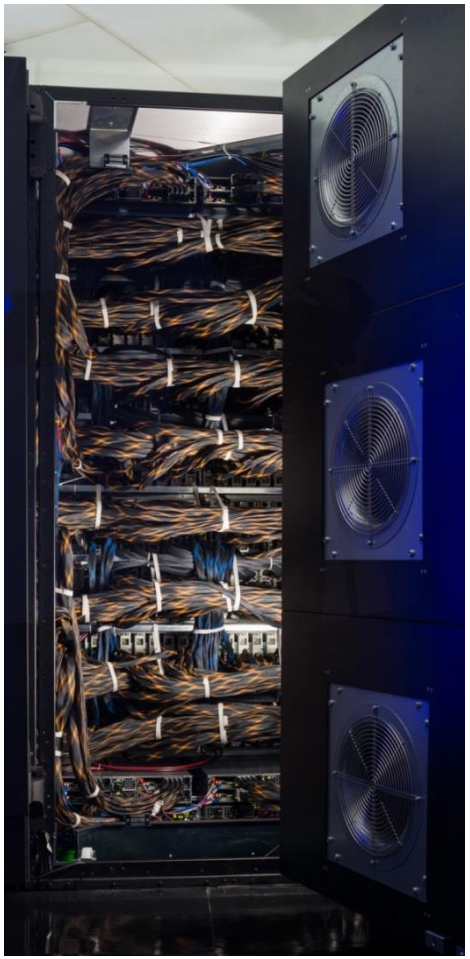
High Power: Node Board

- 24-layer circuit, single-node board
  - Single-node board enhances serviceability
  - Critical signals routed with thin dielectric layers
  - Blind and buried vias
- Core voltage regulator module (VRM)
  - Custom, eight-phase design
  - Bench tested to 300 W
  - Field-replaceable daughter card
  - Supplied via 48-to-12V VRM on node board
- Large Cu + Al heat sink



High Power: Cooling System





- Extensive modeling
- 3 blowers in rack door
  - 1.7 kW each
  - 7,600 ft^3/min airflow
  - Door weighs ~325 lbs.
- 56 kW chilled-water heat exchanger under floor
- Air temp. 16→30 °C
- Baffles balance airflow through node boards
- 42 kW / rack
  - ~22 kW due to ASIC
- Average die temp. 50 °C

Not All of the Cards are Stacked Against Us

- Well-understood workload
- Output trajectory supports frequent checkpointing
- Flexible frequency targets for core and channel
- Driven by performance, not yield
- High compute to I/O ratio
- No need to minimize standby power dissipation
- Moderate, well-controlled operating temperature

The Birth of a New Machine

- Timeline
  - Tapeout December 2012
  - 1st ASIC powered on April 2013
  - 1st 512-node segment December 2013
- Multiple 512-node segments in operation
- Current activities
  - Build-out continuing
  - Performance tuning
  - Biochemistry research!


ANTON2