Millisecond-Scale Molecular Dynamics Simulations on Anton

David E. Shaw*, Ron O. Dror, John K. Salmon, J.P. Grossman, Kenneth M. Mackenzie, Joseph A. Bank, Cliff Young, Martin M. Deneroff, Brannon Batson, Kevin J. Bowers, Edmond Chow, Michael P. Eastwood, Douglas J. Ierardi, John L. Klepeis, Jeffrey S. Kuskin, Richard H. Larson, Kresten Lindorff-Larsen, Paul Maragakis, Mark A. Moraes, Stefano Piana, Yibing Shan, and Brian Towles

D. E. Shaw Research, New York, NY 10036, USA

- Overall:
- This paper presents initial performance results from Anton and describes the relationships between algorithms and hardware that lead to its high performance.

- Contributions:

- co-designed application-specific software hardware for Anton to improve MD simulation.

- report Anton's performance when executing actual MD simulations.

- The first time achieve the simulation of biological molecules at an atomic level of detail for periods on the order of a millisecond.

Molecular Dynamics (MD):

MD simulations of biological molecules
-> trace atomic motions and give deep insights into molecular mechanisms
that experimental approaches could not have achieved alone

functionally important biological events that typically occur on timescales of many microseconds or milliseconds

-the "folding" of proteins into their native three-dimensional structures
-the structural changes that underlie protein function
-the interactions between two proteins or between a protein and a candidate drug molecule

Long timescale simulations are more difficult than simulations of larger chemical systems.
increasing chemical system size can be addressed by adding more parallelism
extending a simulation in time can not be easily addressed.

Commonly used biomolecular force fields express the total force on an atom as a sum of three components:
1) bonded forces, interactions between small groups of atoms connected by one or more covalent bonds
2) van der Waals forces: interactions between all pairs of atoms, fall off quickly with distance, only evaluated for nearby pairs of atoms
3) electrostatic forces: interactions between all pairs of atoms, fall off slowly with distance.

Anton:

Anton is a massively-parallel, specialized supercomputer designed for MD simulations of biomolecular systems.

The machine's specialized hardware dramatically increases the speed of MD calculations

The algorithms and numerical techniques employed by Anton were co-designed with its hardware architecture.

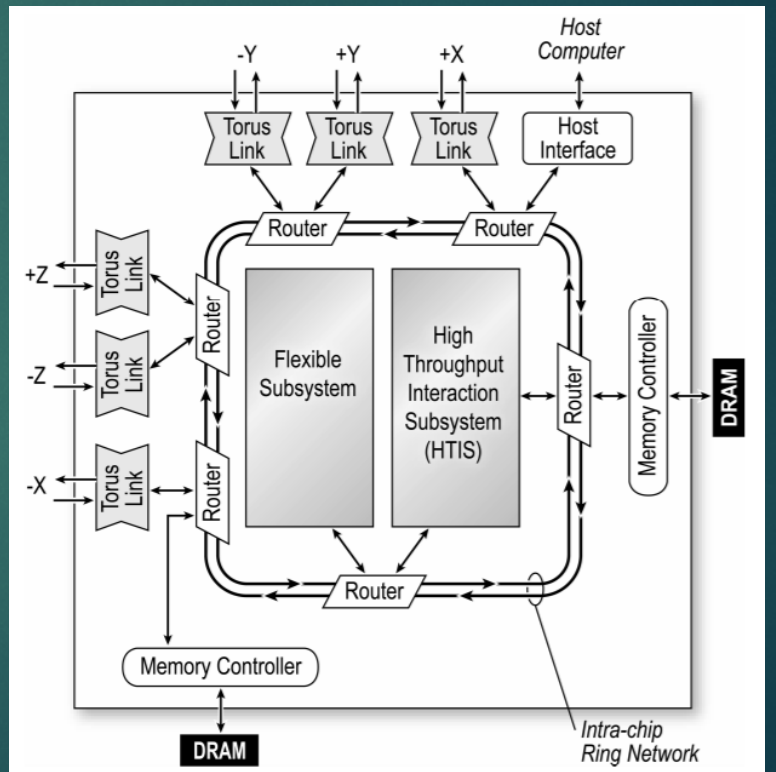

Figure 2: Anton ASIC block diagram.

Anton Architecture

Anton comprises a set of nodes connected in a toroidal topology
the 512-node machines have an 8×8×8 toroidal topology, corresponding to an 8×8×8 partitioning of a chemical system with periodic boundary conditions.

Each node includes an ASIC with two major computational subsystems: high-throughput interaction subsystem (HTIS), the flexible subsystem.

HTIS is designed for computing massive numbers of range-limited pairwise interactions of various forms using an array of 32 hardwired pairwise point interaction pipelines (PPIPs). PPIPs have different computation form determined by various modes, user-specified lookup tables, parameter values.

Flexible subsystem is composed of programmable cores used for the remaining, less structured part of the MD calculation.
It contains eight geometry cores (GCs) for fast numerical computations
Four Tensilica LX processors that control the overall data flow in the Anton system
Four data transfer engines that allow communication to be hidden behind computation

ASIC also contains a pair of DDR2-800 DRAM controllers, six high-speed channels that provide communication to neighboring ASICs, and a host interface that communicates with an external host computer for input, output, and general control.

Units above are connected by a bidirectional on-chip communication ring.

The ASICs are implemented in 90-nm technology and clocked at 485 MHz
(the PPIP is clocked at 970 MHz)

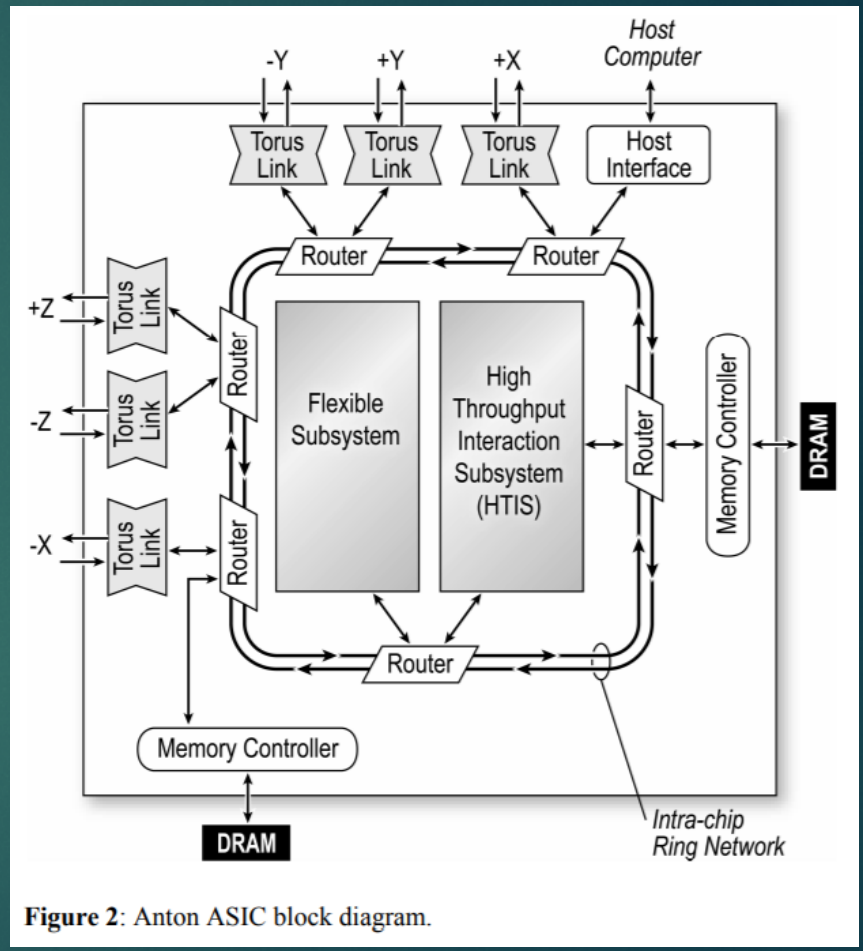

Figure 2: Anton ASIC block diagram.

Anton Advantage

on the x86, electrostatic and van der Waals interactions between pairs of atoms separated by less than 9 Å cost 64% of time. The Fourier computation cost 14% of time, which is the second largest.

Anton's performance is optimized by choosing a larger cutoff and a coarser mesh than are typically used on commodity hardware.

The time required for Fourier computation is determined by tunable parameters in the Ewald decomposition:

electrostatic potential due to a point charge as a sum of two components:
- one that falls rapidly to zero as distance from the charge increases
- one that is smooth throughout space and can thus be computed on a mesh.

Increasing the smoothness of the second component allows the use of a coarser mesh, but also makes the first component fall to zero less rapidly
-> an increased cutoff is required for the range-limited interactions.

On the x86, this parameter change leads to an overall slowdown of nearly twofold, whereas on Anton, it results in a speedup of more than twofold.

	x86 core		Anton	
	small cutoff (9 Å) fine mesh (64^3)	**large cutoff (13 Å) coarse mesh (32^3)**	**small cutoff (9 Å) fine mesh (64^3)**	**large cutoff (13 Å) coarse mesh (32^3)**
Nonbonded forces				
Range-limited forces	56.6 ms (64%)	164.4 ms (89%)	1.4 μs (4%)	1.9 μs (12%)
FFT & inverse FFT	12.3 ms (14%)	1.4 ms (1%)	24.7 μs (63%)	8.9 μs (58%)
Mesh interpolation	9.6 ms (11%)	8.8 ms (5%)	9.5 μs (24%)	2.0 μs (13%)
Correction forces	4.0 ms (5%)	3.8 ms (2%)	2.5 μs (6%)	2.5 μs (16%)
Bonded forces	2.7 ms (3%)	2.7 ms (1%)	3.5 μs (9%)	4.1 μs (27%)
Integration	3.4 ms (4%)	3.4 ms (2%)	1.6 μs (4%)	1.6 μs (10%)
Total	88.5 ms (100%)	184.5 ms (100%)	39.2 μs (100%)	15.4 μs (100%)

Anton's Algorithm

Anton's performance depends on its ability to keep the many arithmetic units on each node busy with useful computation.

To achieve this, the algorithms should be chosen wisely.

1. requires algorithmic choices that result in a predictable pattern of data movement.

->Each computational subunit on an Anton ASIC has its own low-latency dedicated memory. Intra-node data transfers between these subunits are carefully choreographed to minimize data movement and to deliver data just when it is needed

2. requires algorithms that use many short messages.

-> On Anton, inter-node latency is tens of nanoseconds, and messages with as little as four bytes of data can be sent efficiently.

Anton Numerical Techniques

Anton achieves both a performance advantage and desirable numerical properties by using customized numerical formats (fixed-point arithmetic).

The use of fixed-point arithmetic has two substantial advantages over floating point

1. fixed-point hardware adders have significantly lower area and latency than their floating-point equivalents:

2. unlike floating-point addition, fixed-point addition is associative. In other words, the order of summation of fixedpoint numbers will not affect numerical results

Which leads to good properties of Anton:

Determinism: simulations on the same hardware with the same inputs produce same result

Parallel invariance: a given simulation will evolve in exactly the same way on any single- or multi-node Anton configuration

Exact reversibility: run 400 million time steps then negated the instantaneous velocities of all the atoms, recover the initial conditions bit-for-bit after another 400 million time steps (without constraints, temperature control or pressure control)

Results

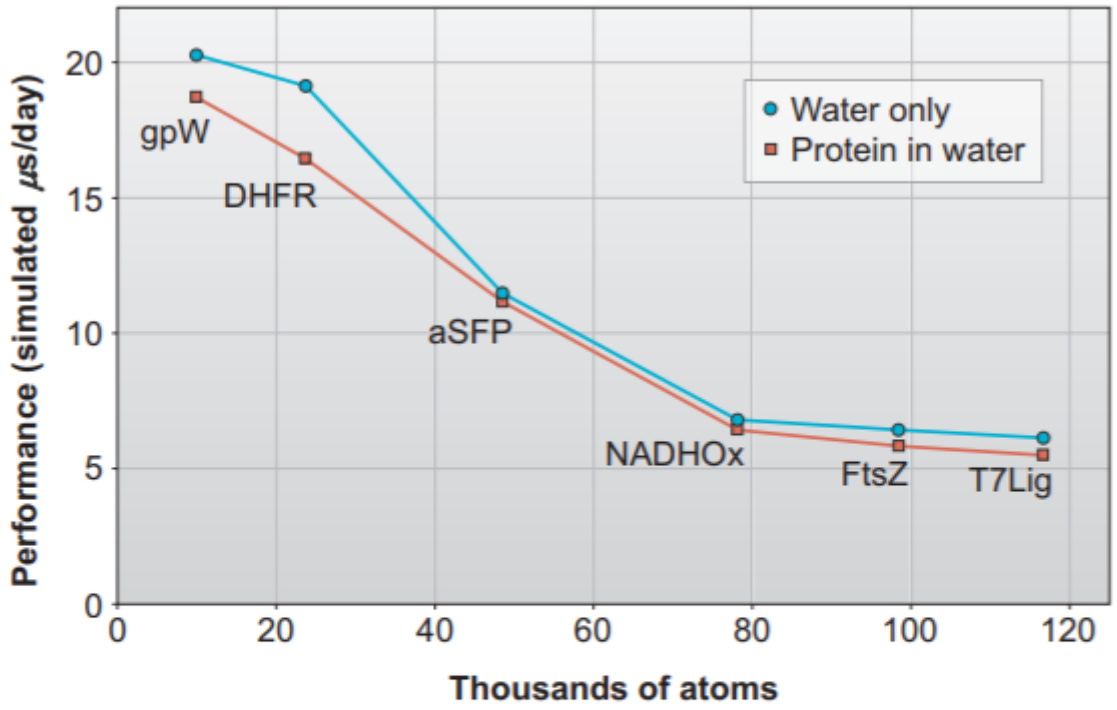

Figure 5: Performance of a 512-node Anton machine for chemical systems of different sizes. Simulation parameters are specified in Table 4 (water systems use the same parameters as similarly sized protein systems). Performance was measured over ten million time steps for each system.

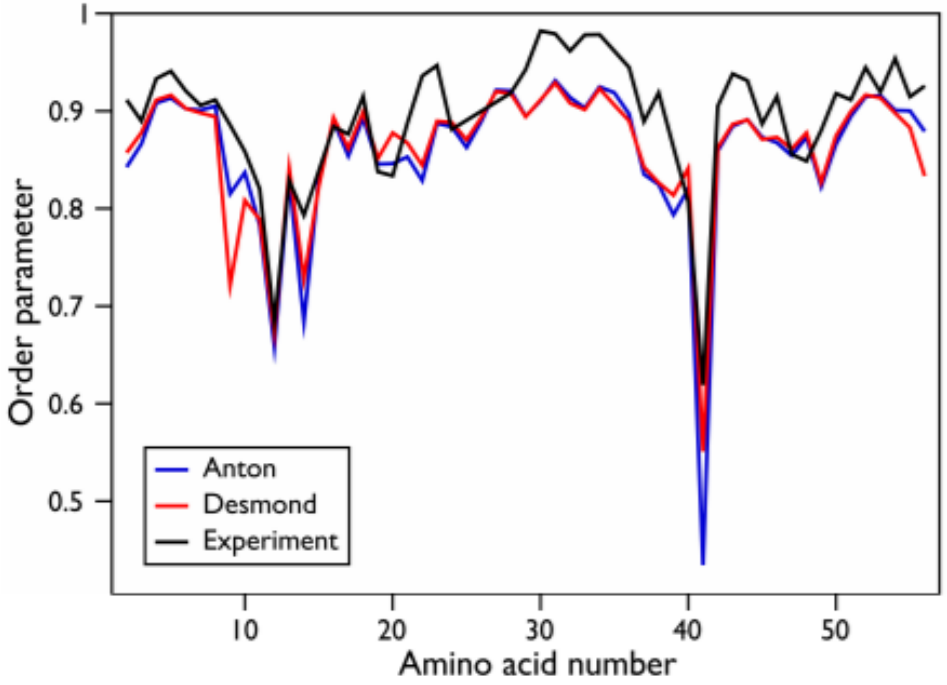

Figure 6: Amount of motion of each amino acid in the protein GB3, as estimated from 1-μs simulations on Anton (blue) and Desmond (red), and as measured experimentally by NMR (black). Smaller values on the vertical axis indicate more motion. Simulations used the AMBER99SB force field [17].

Thanks

Questions?