Toward Millisecond-Scale Molecular Dynamics Simulations of Proteins

David E. Shaw

D. E. Shaw Research and

Center for Computational Biology and Bioinformatics
Columbia University

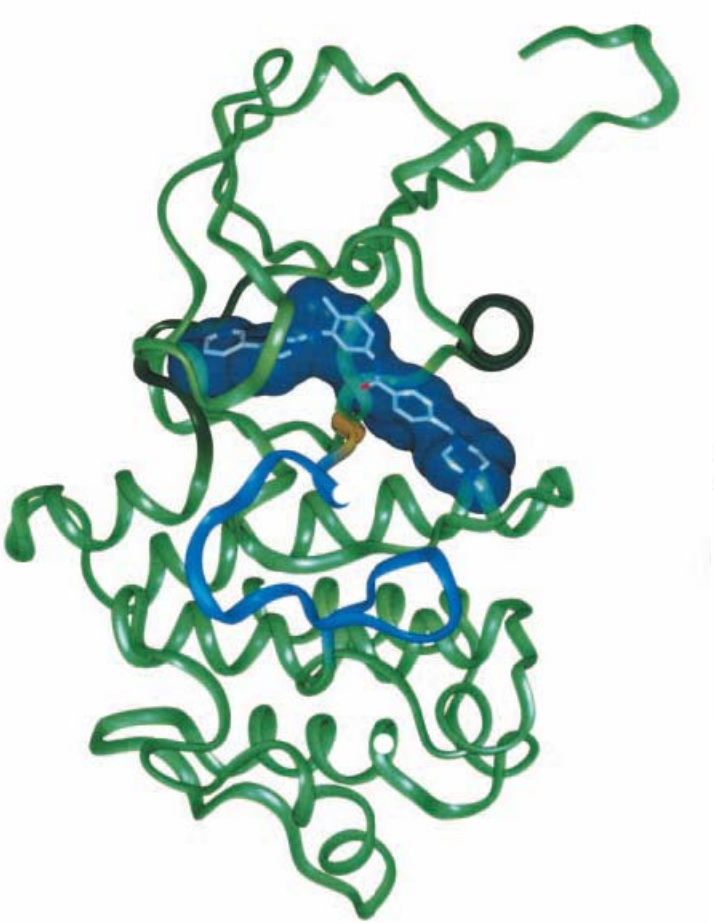
D. E. Shaw Research

- Independent research lab
  - Computational biochemistry
  - Focus on basic science, with long time horizon
- Central focus: Molecular dynamics simulations
  - Elucidate structural, dynamic behavior of
    - Proteins
    - Other biologically significant molecules
  - Provide tools for development of new drugs

Our Goal

- Single, millisecond-scale MD simulations
  - Protein with ~25K to 50K atoms, including water
- One long trajectory, not many short ones
- Why a millisecond?
  - That's a time scale at which many biologically interesting things start to happen.

Binding of Drugs to their Molecular Targets



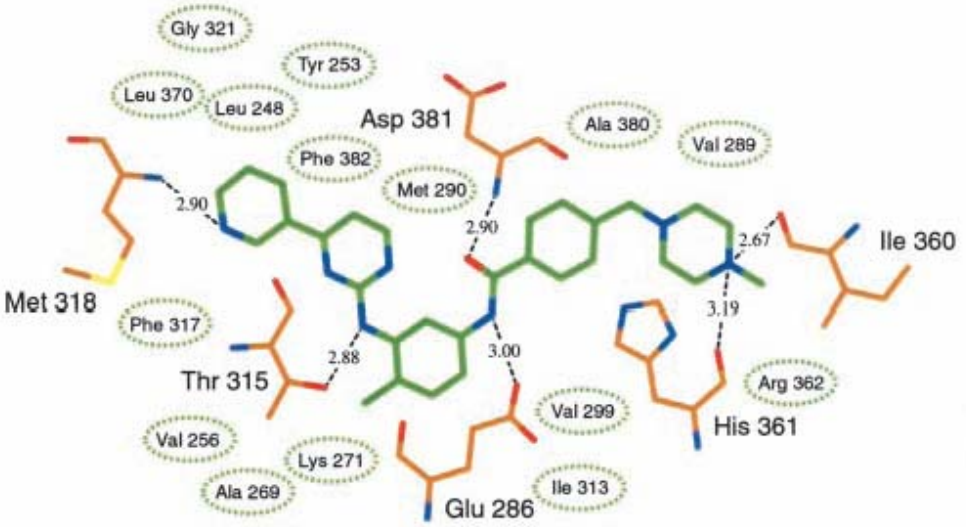
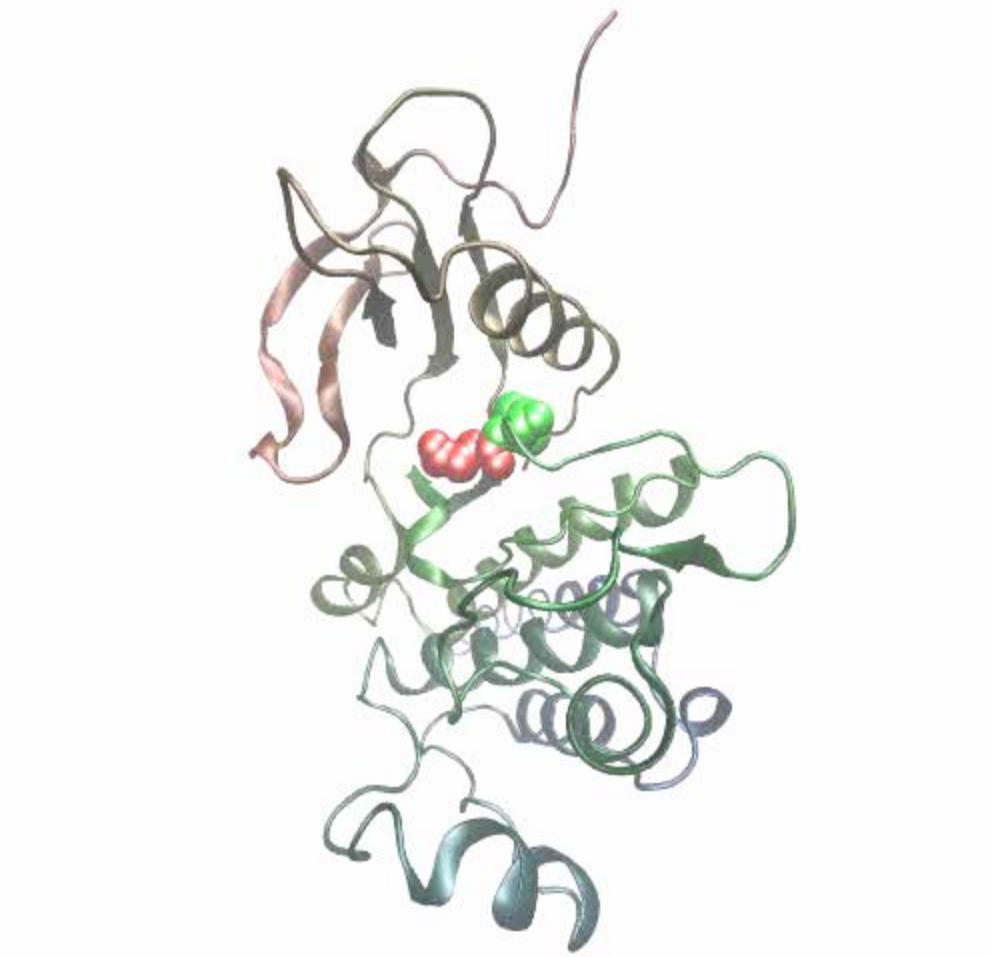


Image: Nagar, et al., *Cancer Res.* 62, 4236 (2002)

Dynamics of Conformational Changes



Interactions Between Proteins

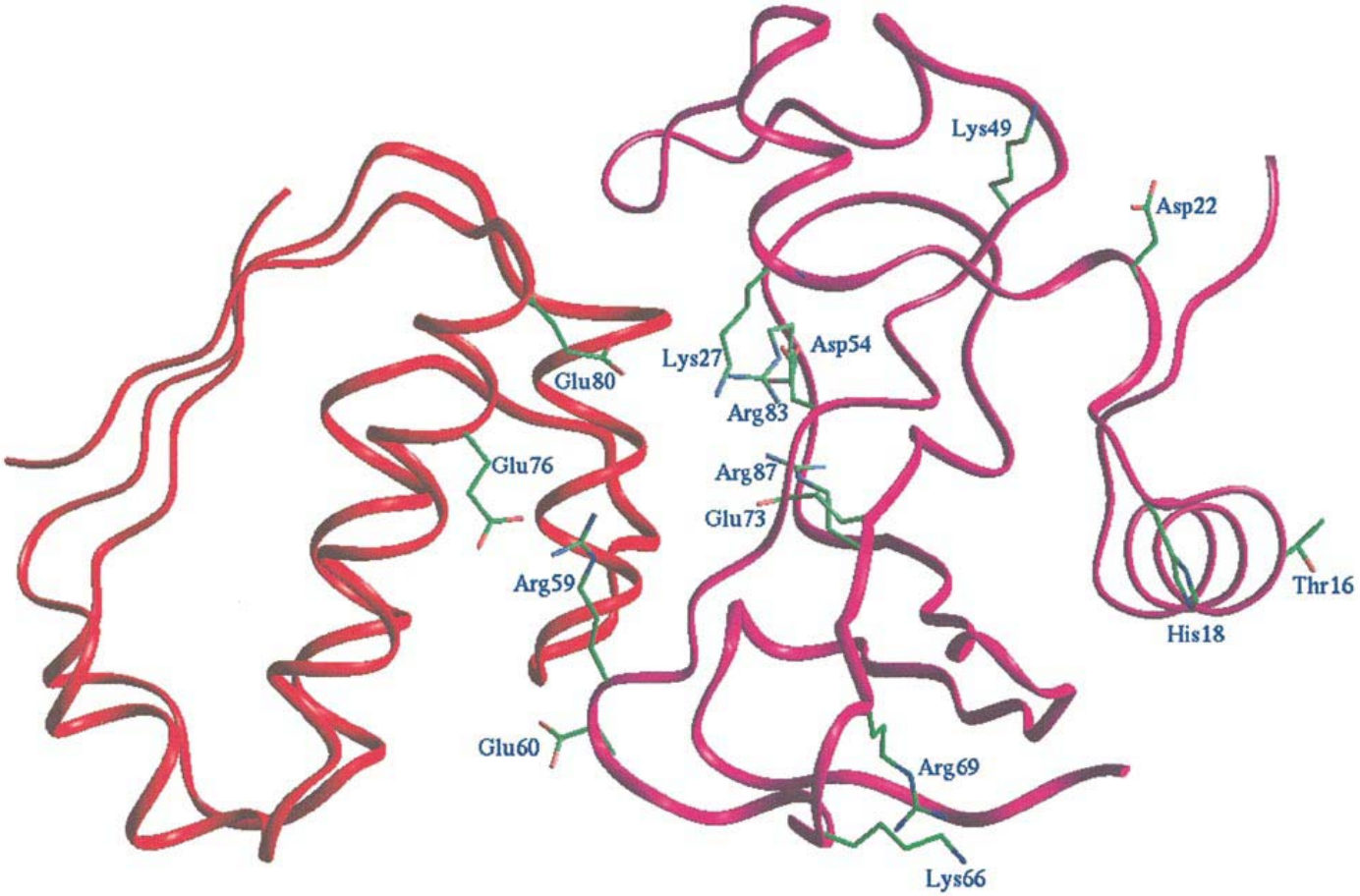


Image: Vijayakumar, et al., *J. Mol. Biol.* 278, 1015 (1998)

Protein Folding

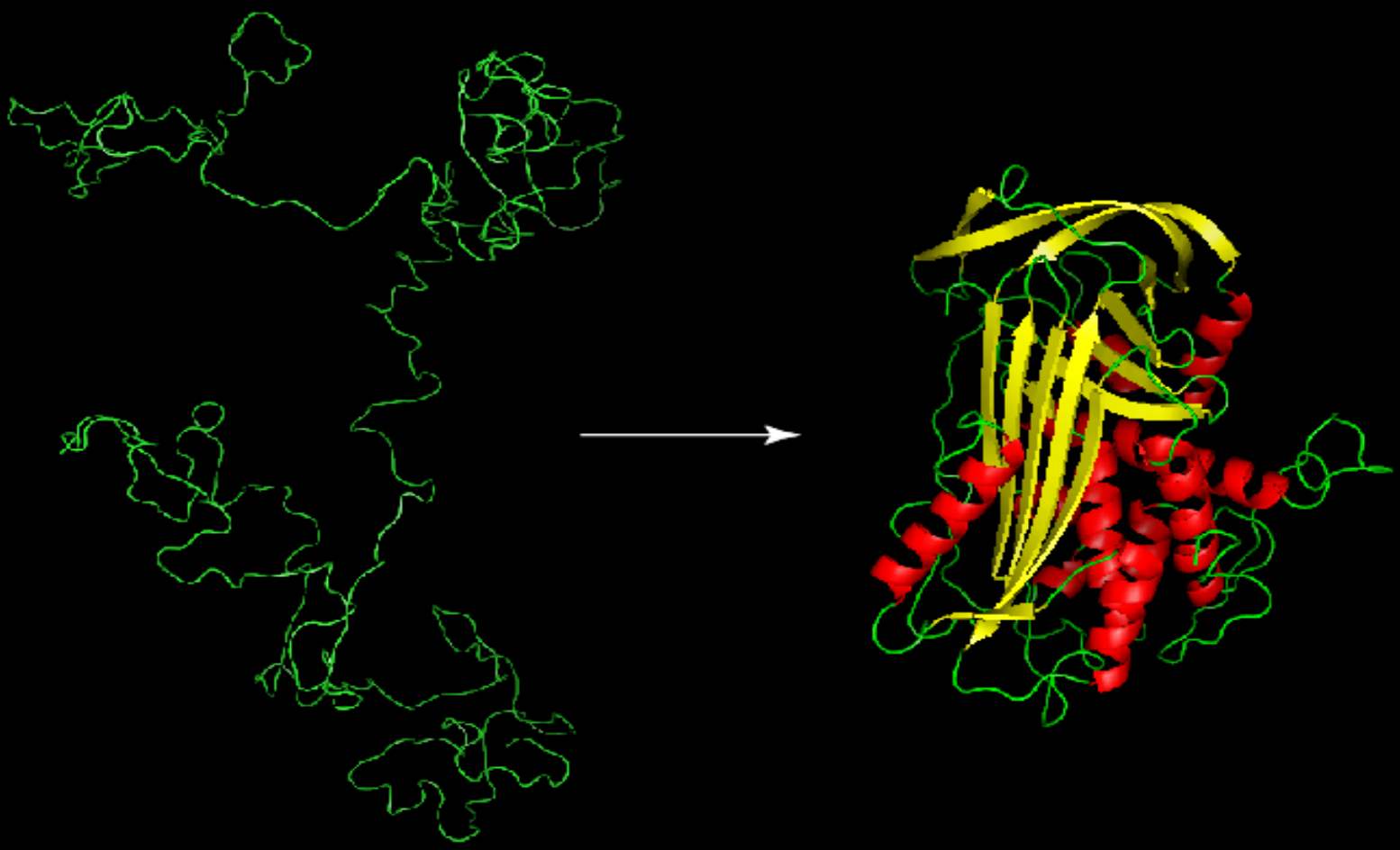
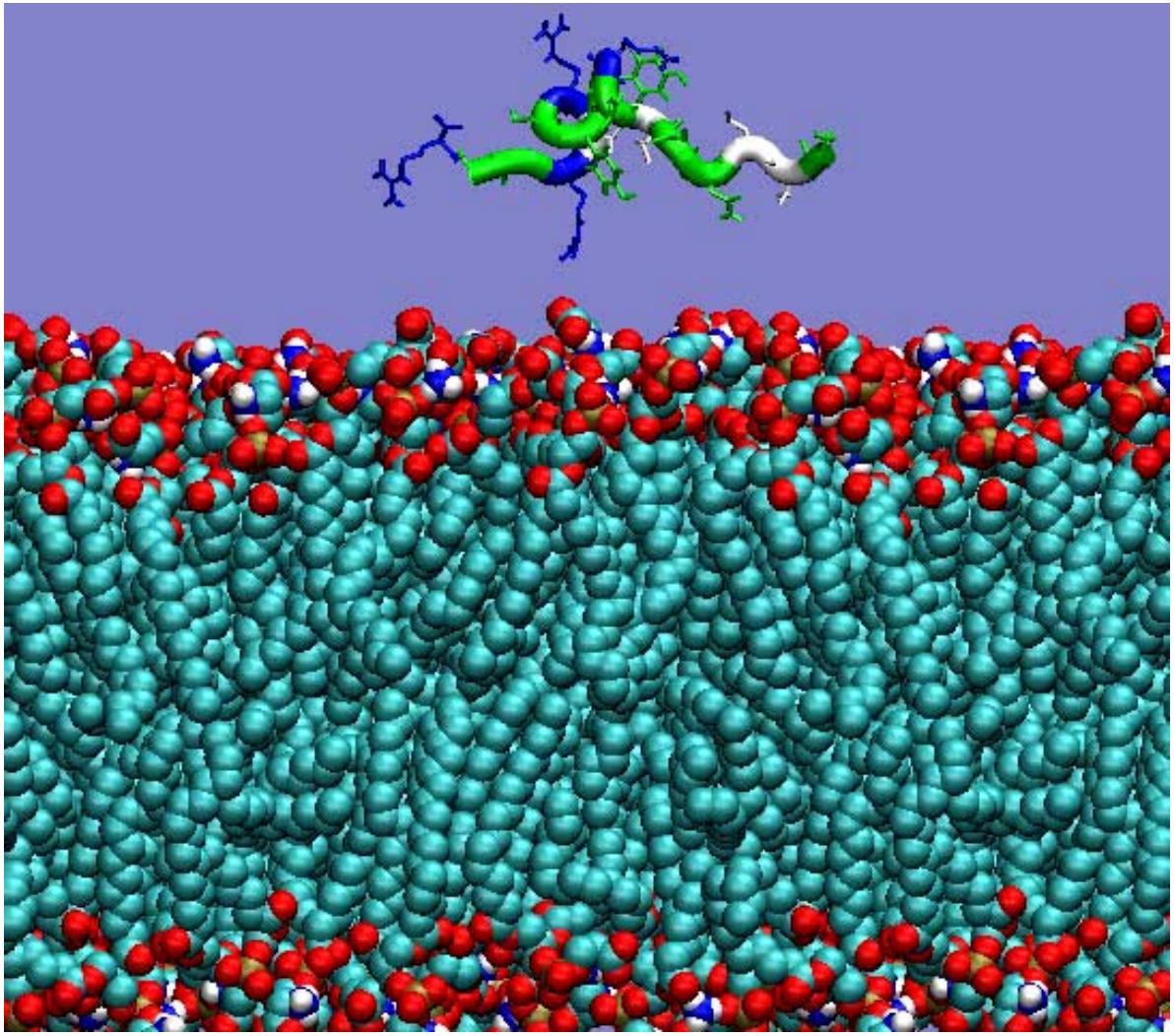


Image: Istvan Kolossvary & Annabel Todd, D. E. Shaw Research

What Will It Take to Simulate a Millisecond?

- Will require truly massive parallelism
  - Current single processor codes: ~ 1 ns / day
  - Fastest current parallel codes: ~ 100 ns/day
  - Goal will require ~ 10,000 ns / day
- Key to massive parallelism:
  - Avoid interprocessor communication bottlenecks

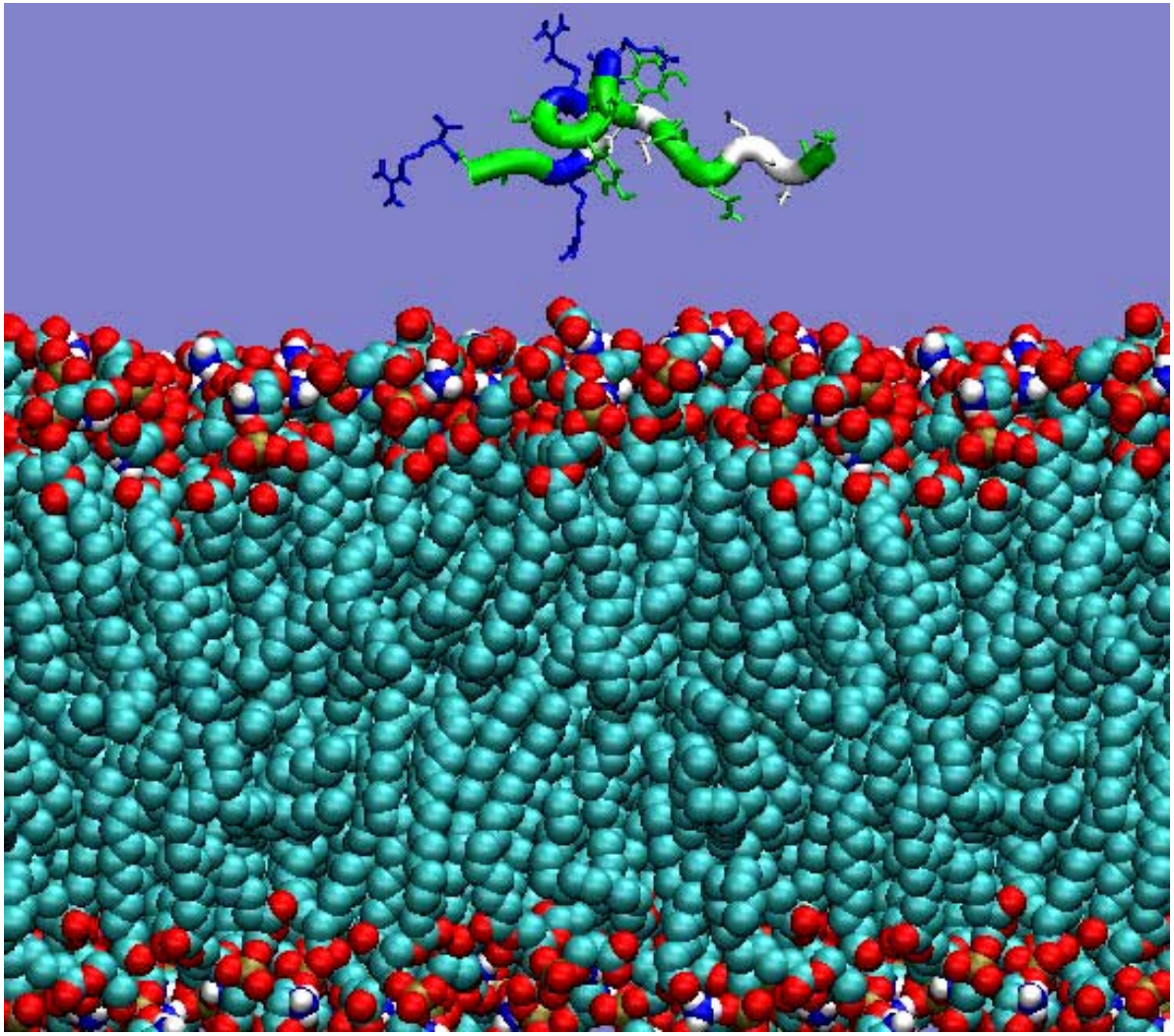
Illustration of Required Speedup



2 ns
simulated time

3.4 days to
simulate now
on single CPU

Illustration of Required Speedup



2 ns
simulated time

3.4 days to
simulate now
on single CPU

13 seconds at
target speed

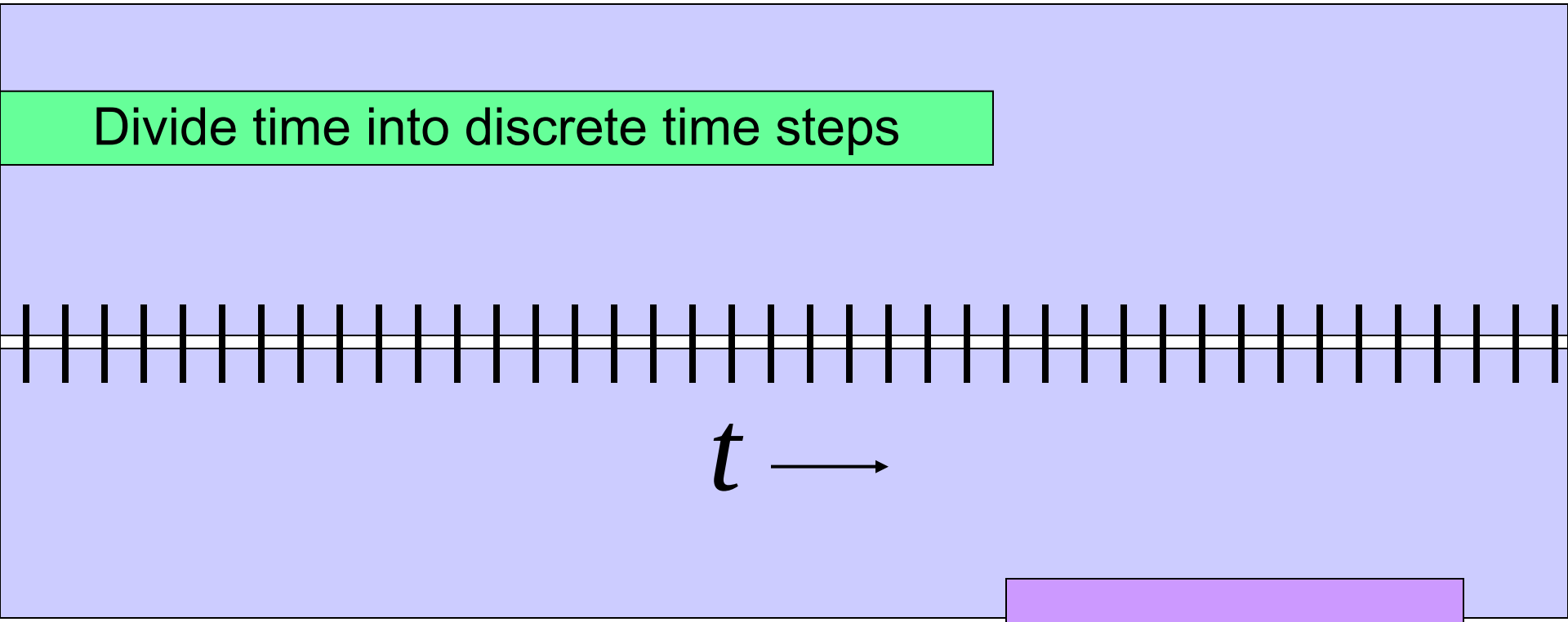
Our Approach

- New architectures
  - Designing specialized machine: Anton
  - Enormously parallel architecture
  - Based on special-purpose ASICs
  - Dramatically faster for MD, but less flexible
  - Projected completion: end of 2008 (chips working now)

- New algorithms
  - Desmond: Massively parallel MD code
  - Applicable to
    - Conventional clusters (Desmond C)
      - Available without cost to academic users
    - Anton (Desmond A)

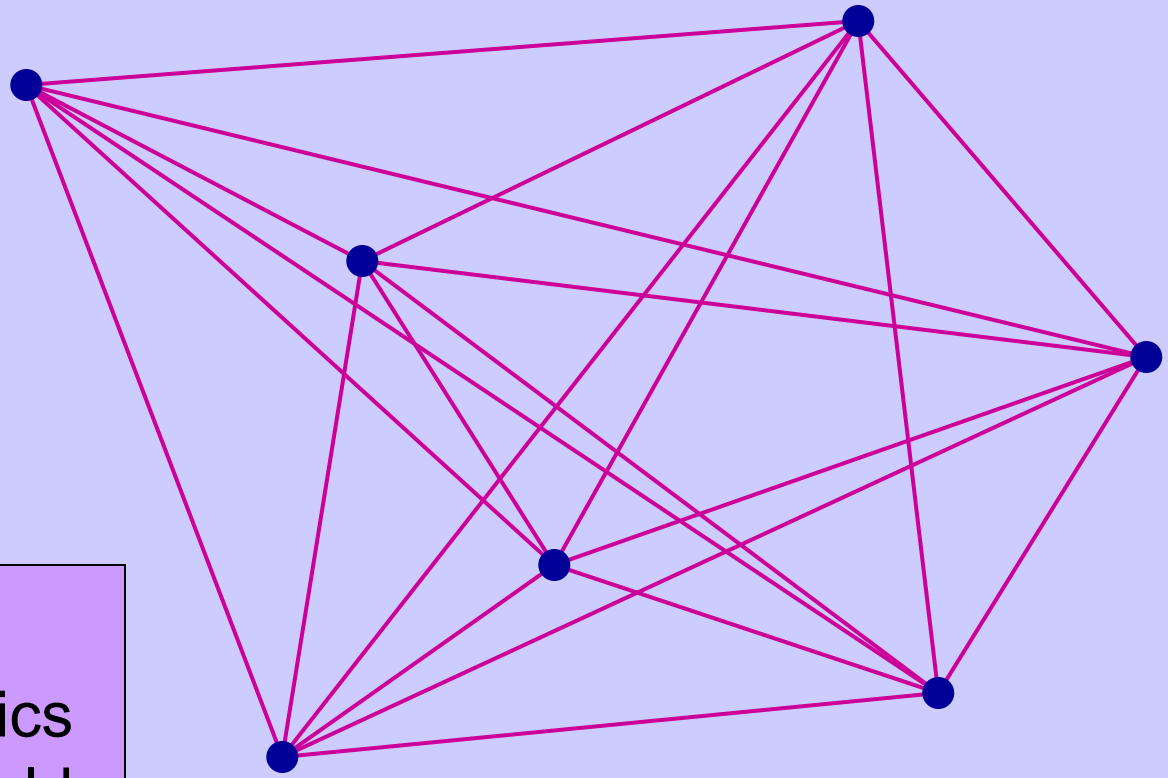
Molecular Dynamics

Molecular Dynamics



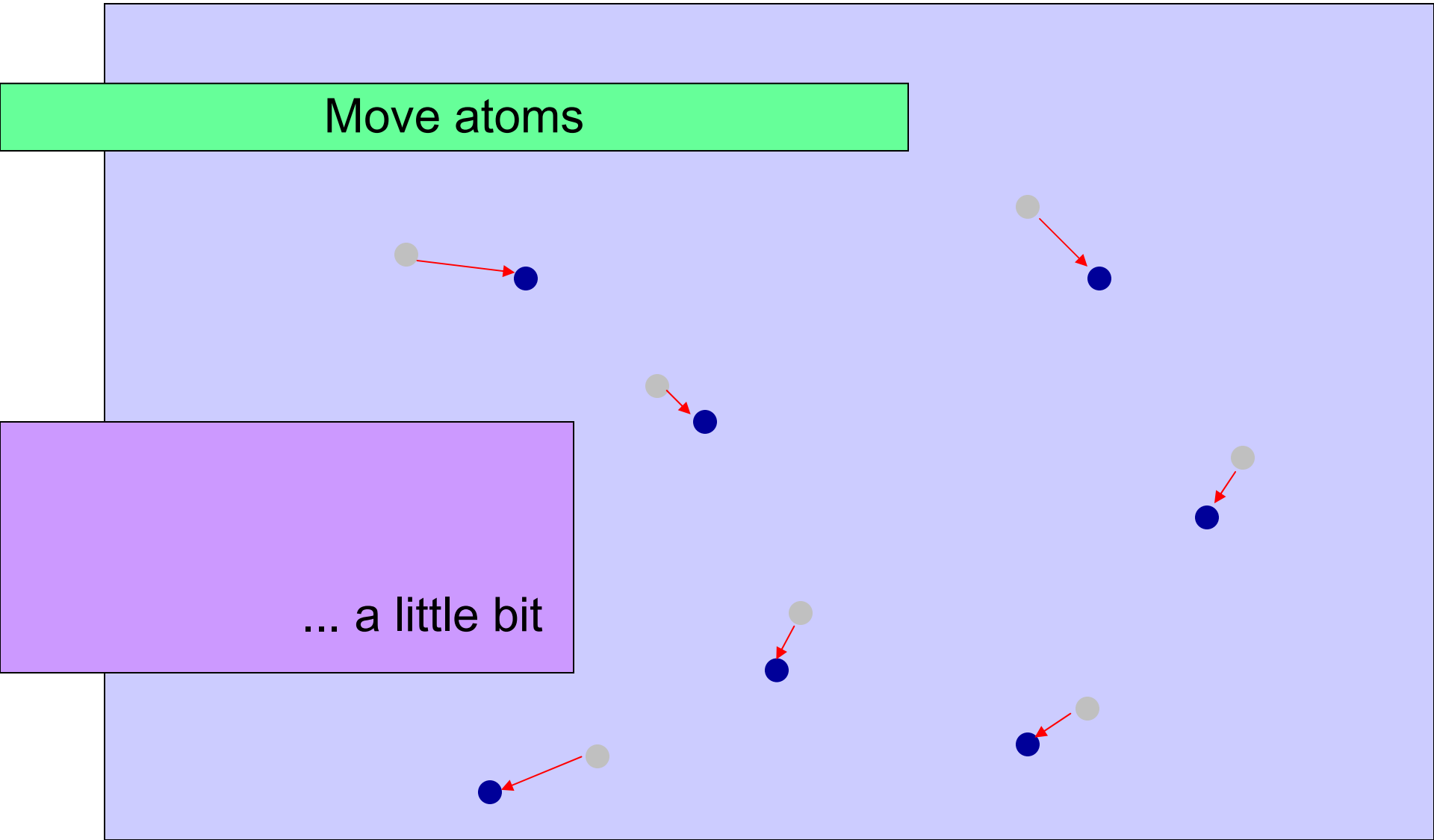
Molecular Dynamics

Calculate forces



Molecular mechanics force field

Molecular Dynamics

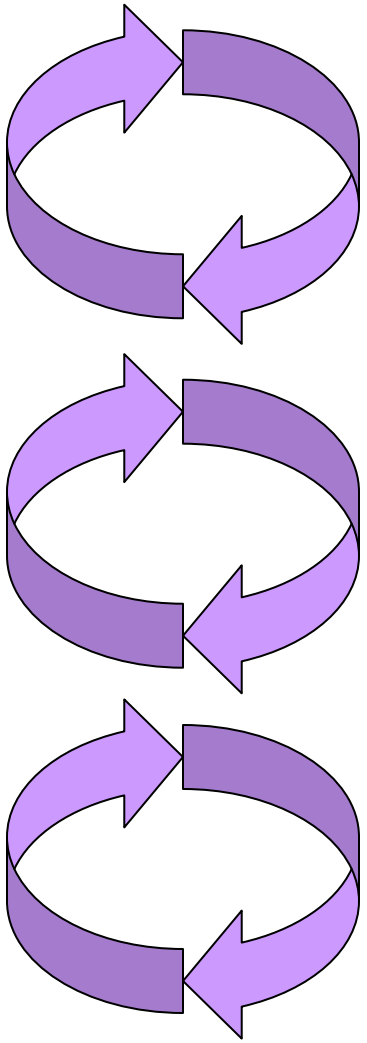


Molecular Dynamics

Iterate

... and iterate

... and iterate



1 ms ≈ 10^{12} time steps

Molecular Mechanics Force Field

$$E = \sum_{\text{bonds}} k_b \left(r - r_0\right)^2$$ Stretch

$$+ \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2$$ Bend

$$+ \sum_{\text{torsions}} A[1 + \cos(n\tau - \varphi)]$$ Torsion

Bonded

$$+ \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}}$$ Electrostatic

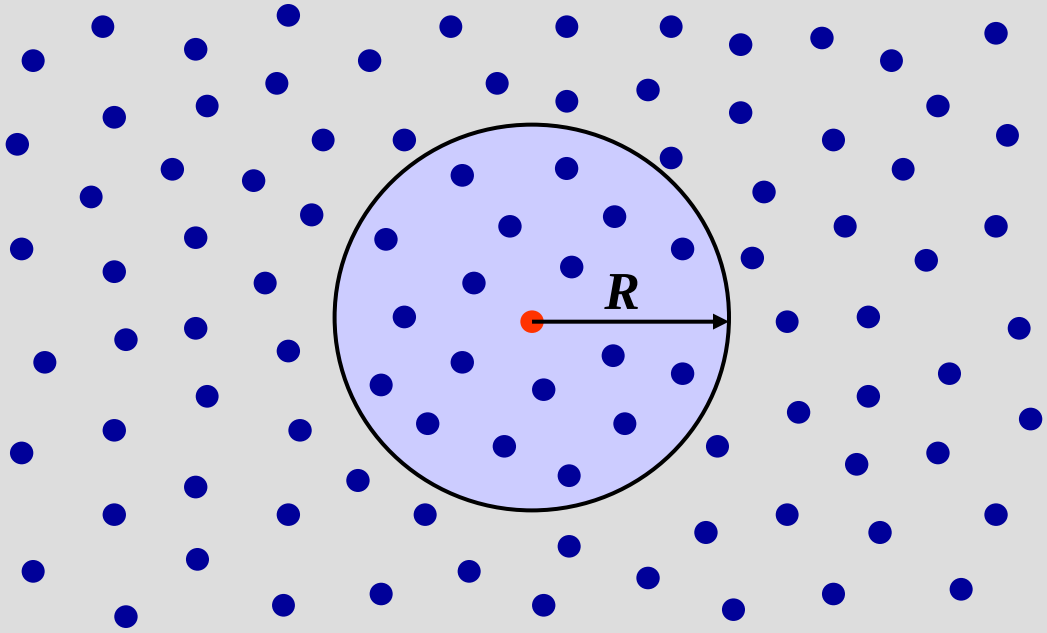
$$+ \sum_i \sum_{j>i} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$ Van der Waals

Non-Bonded

Non-bonded calculations account for most of the work

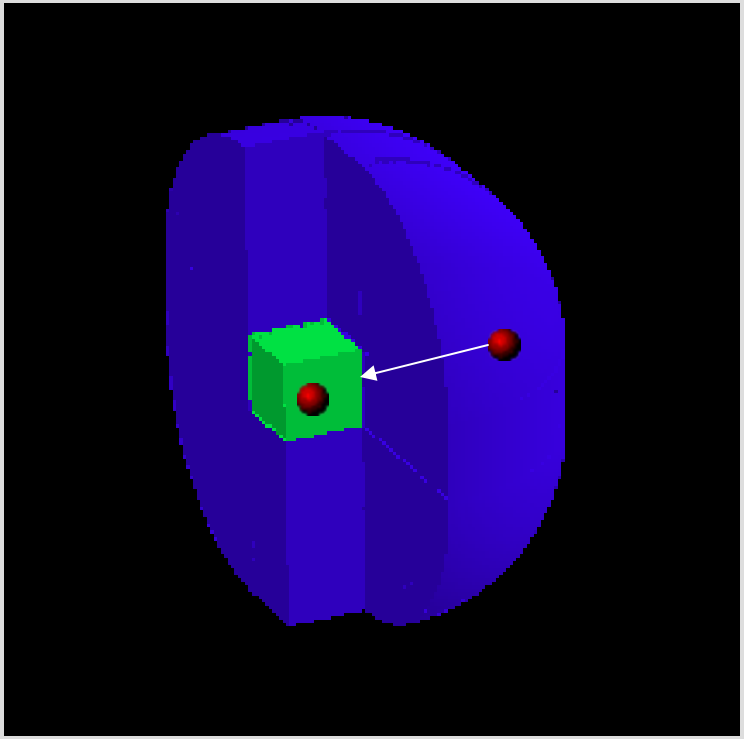
Algorithms:
Desmond and the NT Method

Separating Near and Distant Interactions

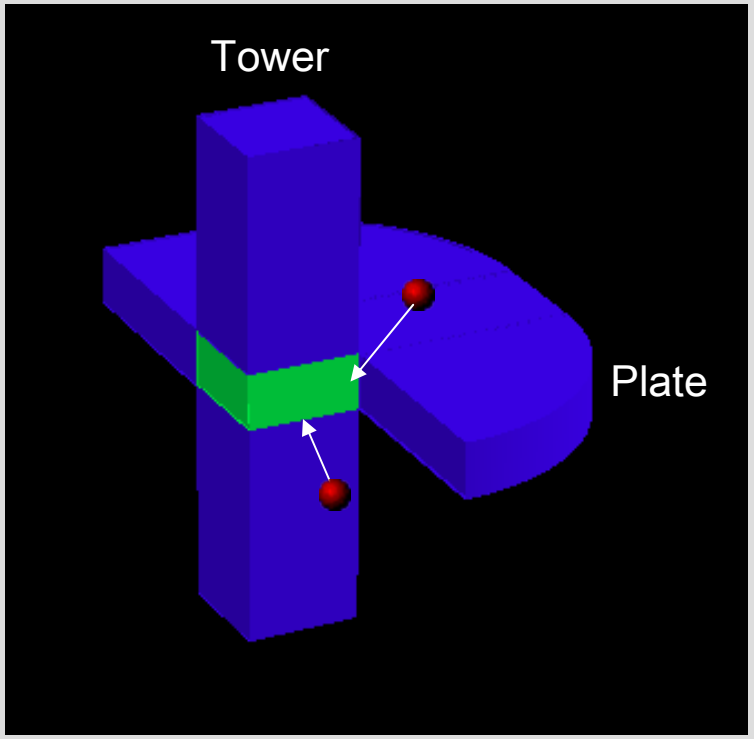


- Efficient methods known for distant interactions
- What's left is a range-limited N-body problem

Meeting on Neutral Territory



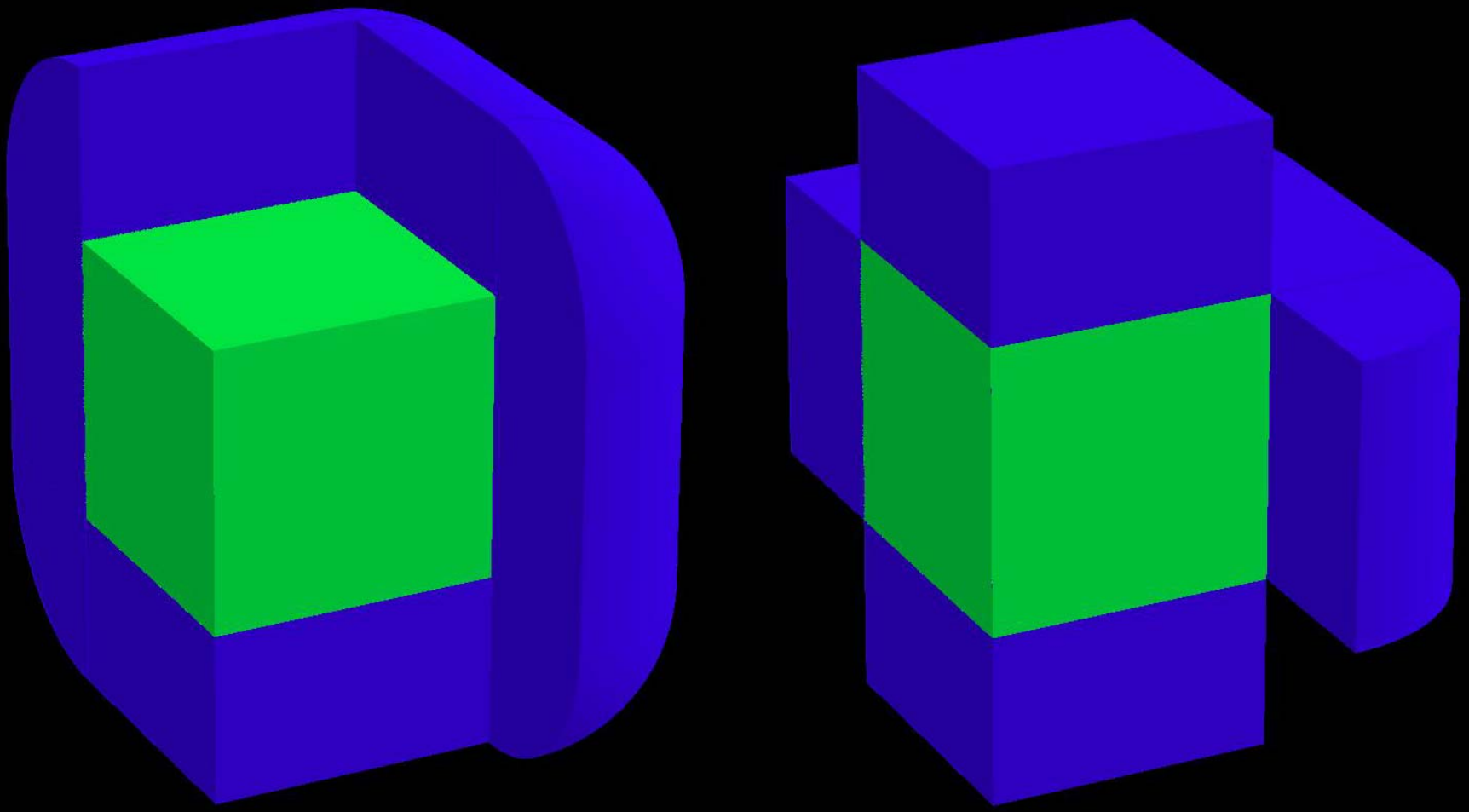
Traditional Method



NT Method

Scaling of Traditional vs. NT Method

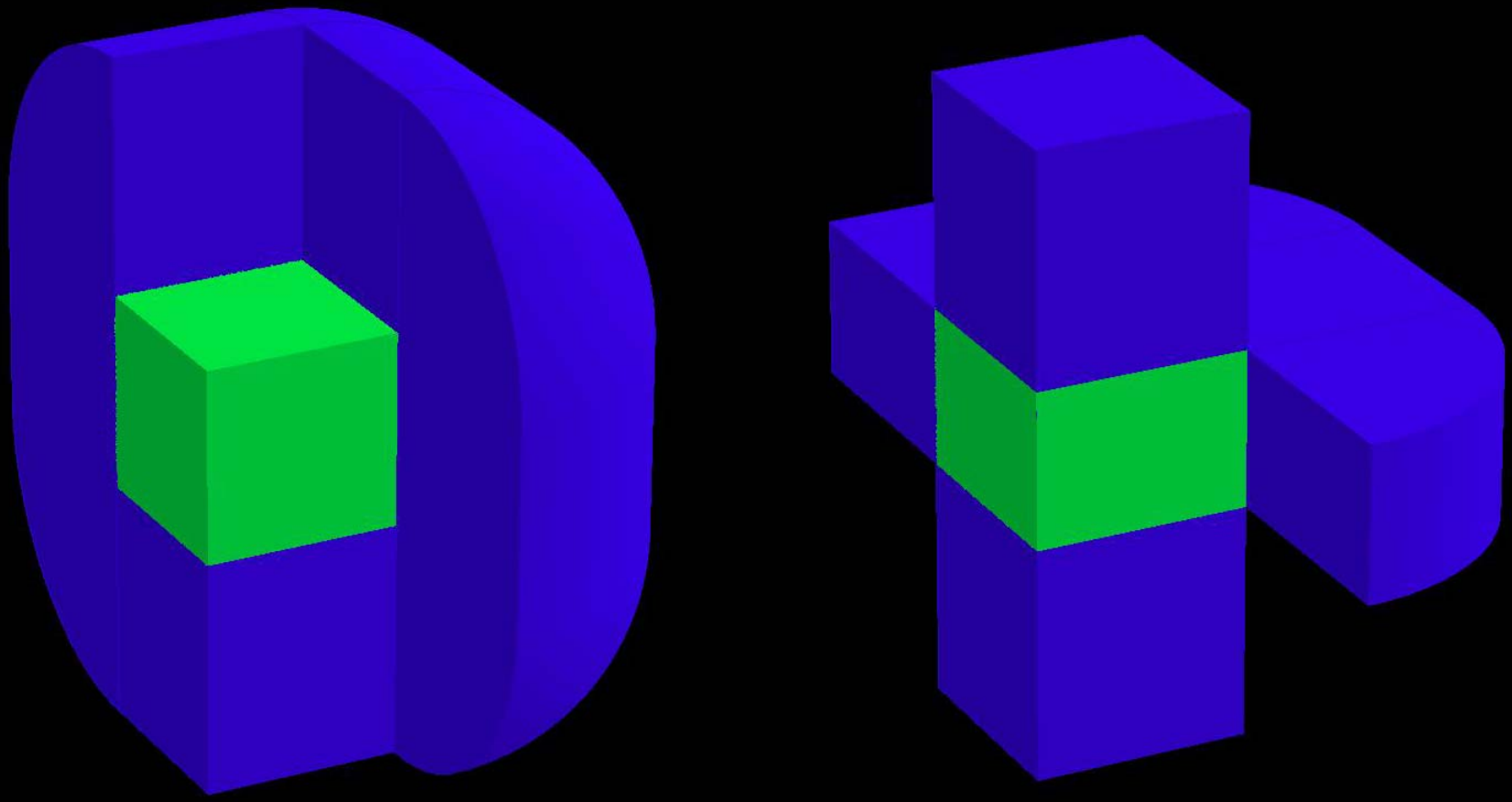
64 Processors



Assumes 50,000 atoms, interaction radius = 12A, density = 0.1 atom/A^3

Scaling of Traditional vs. NT Method

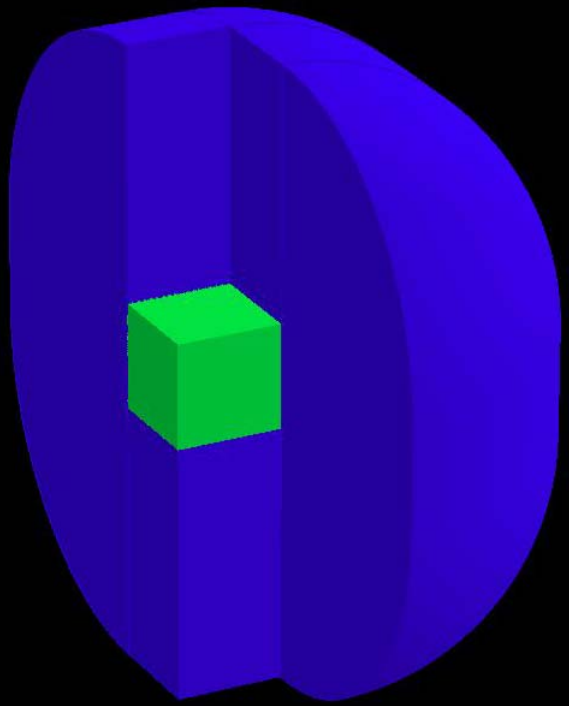
512 Processors

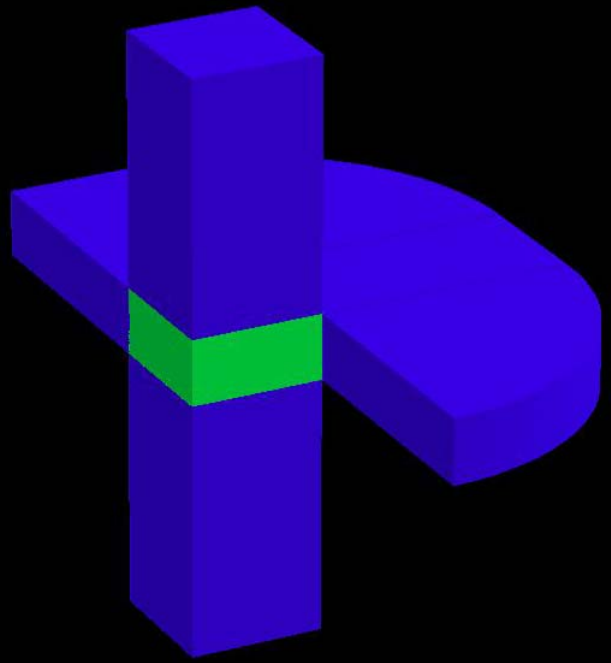


Assumes 50,000 atoms, interaction radius = 12A, density = 0.1 atom/A^3

Scaling of Traditional vs. NT Method

4K Processors

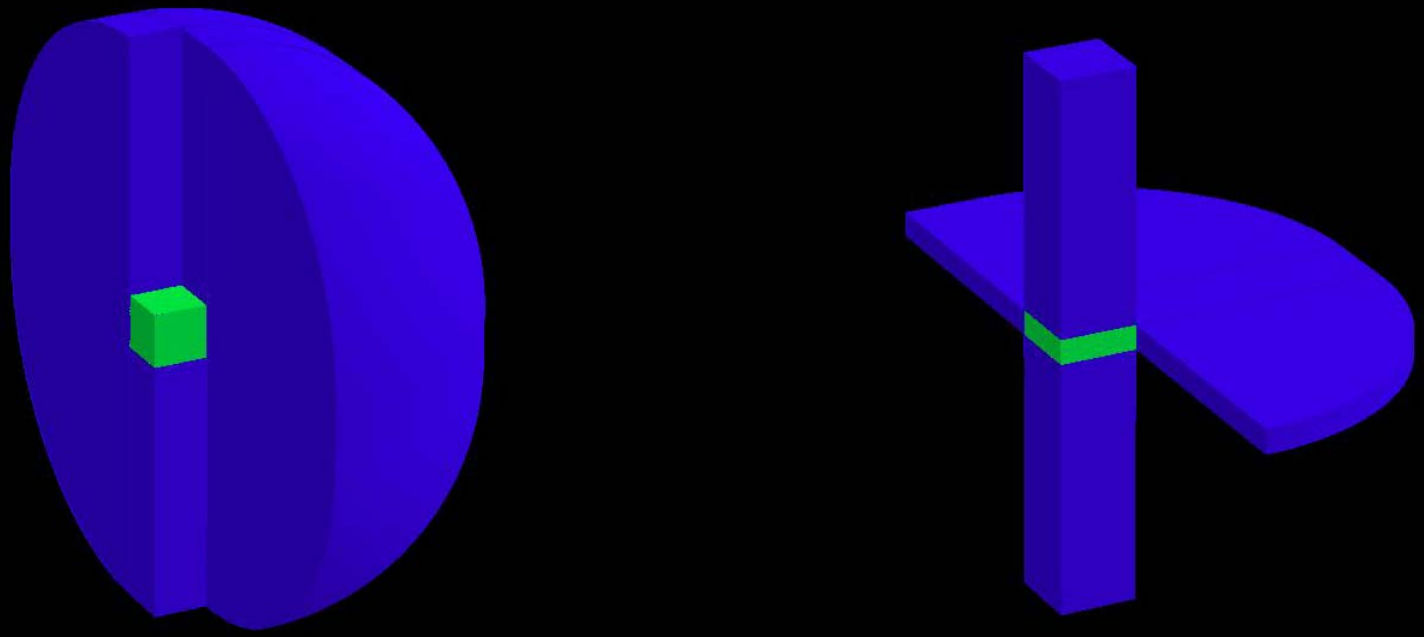




Assumes 50,000 atoms, interaction radius = 12A, density = 0.1 atom/A^3

Scaling of Traditional vs. NT Method

32K Processors



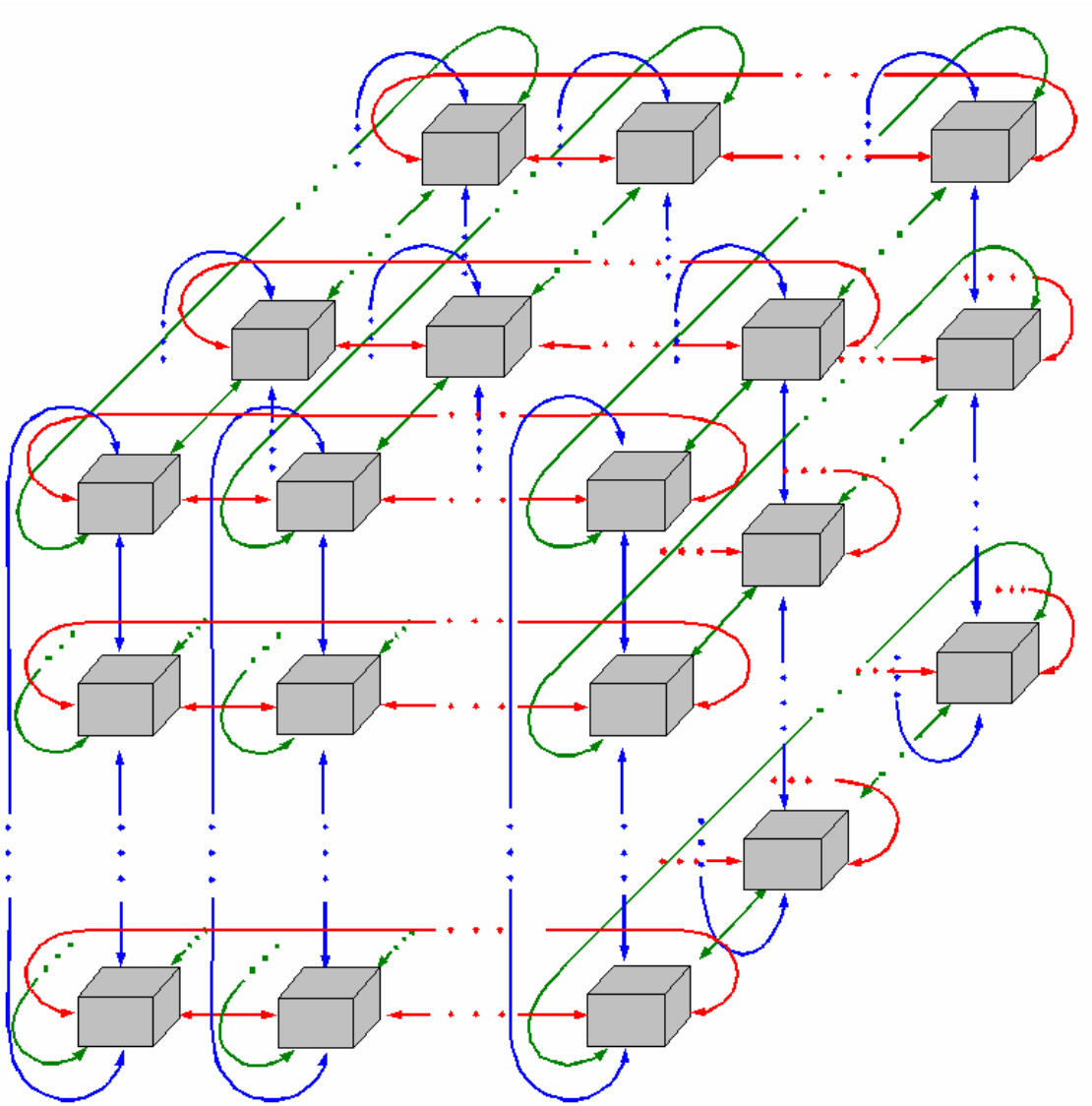
Assumes 50,000 atoms, interaction radius = 12A, density = 0.1 atom/A^3

Architectures:
The Anton Machine

Source of Speedup on Anton

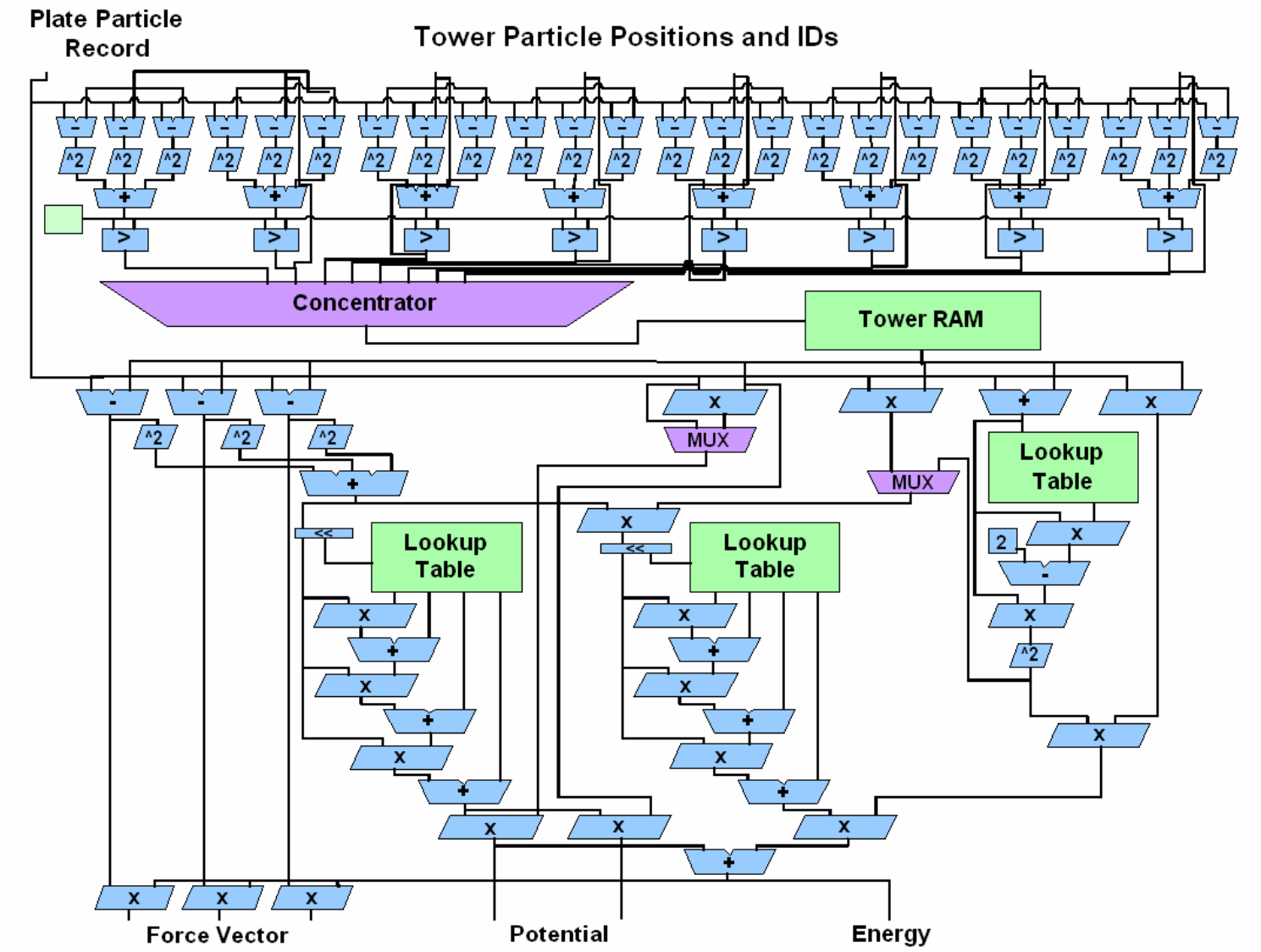
- Judicious use of arithmetic specialization
  - Flexibility, programmability only where needed
  - Elsewhere, hardware tailored for speed
    - Tables and parameters, but not programmable
- Carefully choreographed communication
  - Data flows to just where it's needed
  - Almost never need to access off-chip memory

One Anton Segment (512 ASICs)

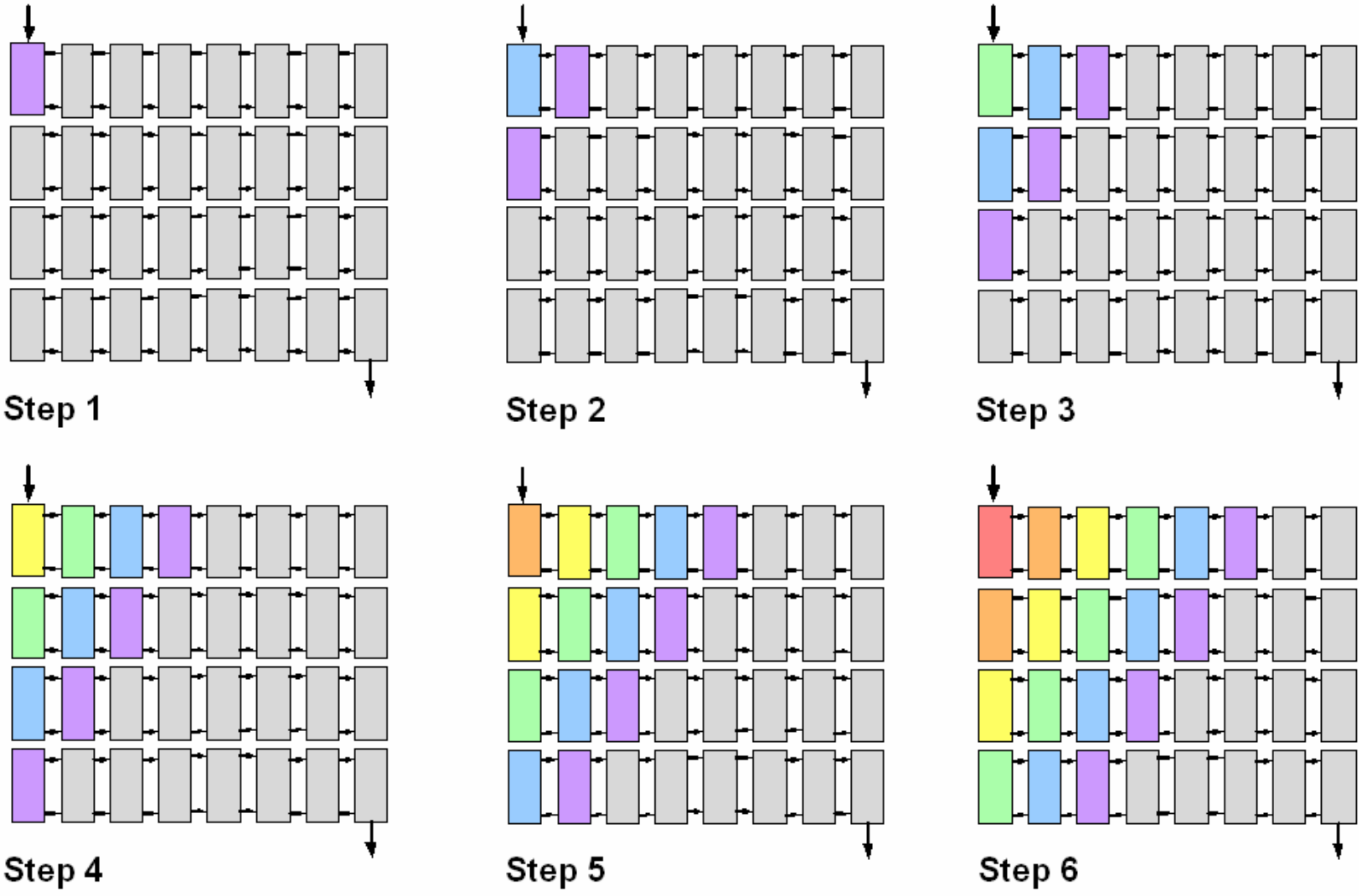


Anton PC Board

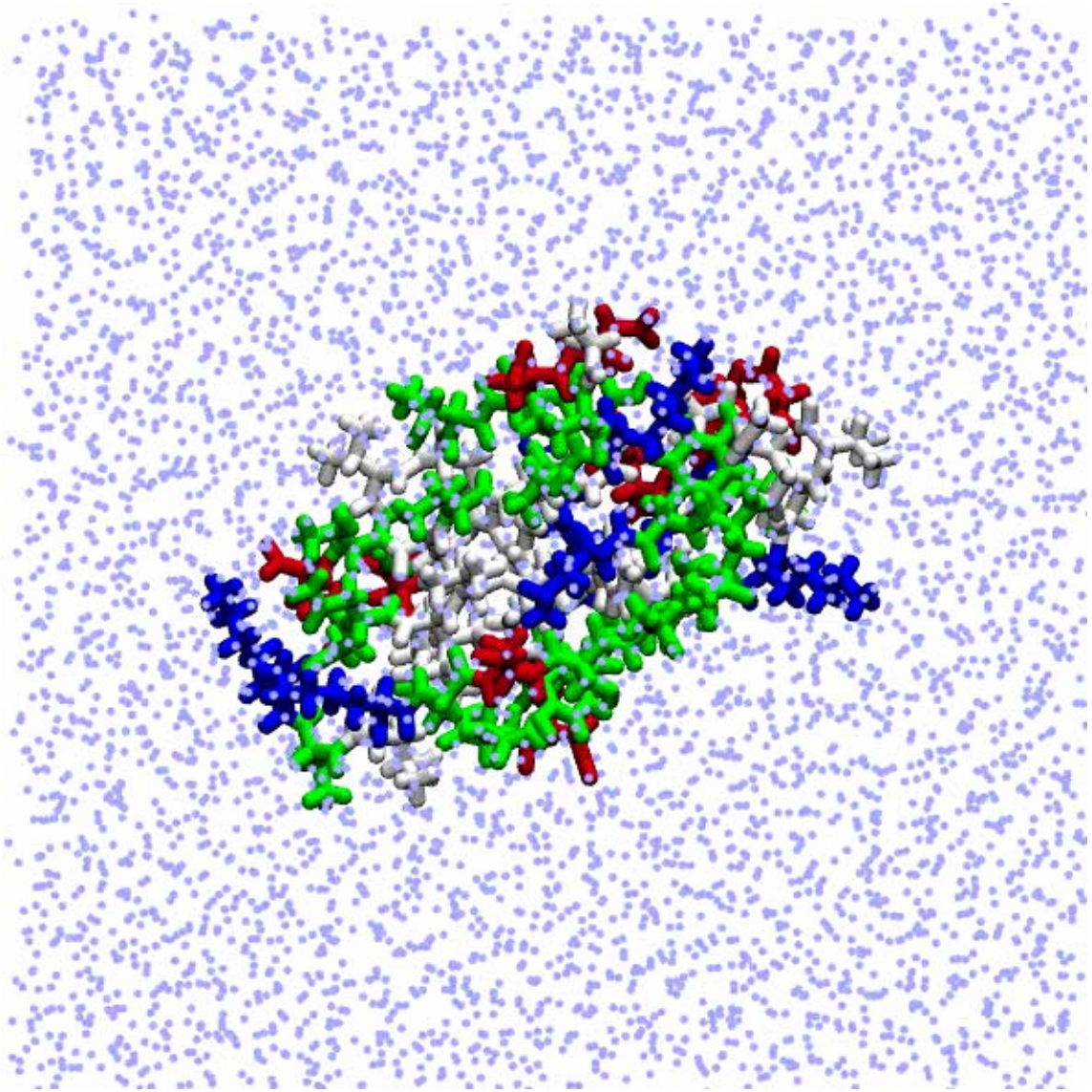
Pairwise Point Interaction Pipeline (32 per chip)



Array of 32 Pairwise Point Interaction Pipelines



Test Run (GB3) on a Real Anton Chip



Comparative Performance

Desmond's Performance on a Commodity Cluster

(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0

Desmond's Performance on a Commodity Cluster

(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3

Desmond's Performance on a Commodity Cluster

(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3
IBM Blue Matter on Blue Gene/L	512 processor cores	4.8 / 9.5

Desmond's Performance on a Commodity Cluster

(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3
IBM Blue Matter on Blue Gene/L	512 processor cores	4.8 / 9.5
GROMACS 4 (new parallel implementation)	128 processor cores	~42

Desmond's Performance on a Commodity Cluster

(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3
IBM Blue Matter on Blue Gene/L	512 processor cores	4.8 / 9.5
GROMACS 4 (new parallel implementation)	128 processor cores	~42
NAMD on cluster (benchmark params)	512 processor cores	77

Desmond's Performance on a Commodity Cluster

(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3
IBM Blue Matter on Blue Gene/L	512 processor cores	4.8 / 9.5
GROMACS 4 (new parallel implementation)	128 processor cores	~42
NAMD on cluster (benchmark params)	512 processor cores	77
Desmond on cluster (benchmark params)	512 processor cores	132

Ref: Bowers ... Shaw, *Proc. SC06*, 2006

Desmond's Performance on a Commodity Cluster

(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3
IBM Blue Matter on Blue Gene/L	512 processor cores	4.8 / 9.5
GROMACS 4 (new parallel implementation)	128 processor cores	~42
NAMD on cluster (benchmark params)	512 processor cores	77
Desmond on cluster (benchmark params)	512 processor cores	132
Desmond on cluster (production params)	512 processor cores	280

Ref: Bowers ... Shaw, *Proc. SC06*, 2006

Comparative Performance of Anton
(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3
IBM Blue Matter on Blue Gene/L	512 processor cores	4.8 / 9.5
GROMACS 4 (new parallel implementation)	128 processor cores	~42
NAMD on cluster (benchmark params)	512 processor cores	77
Desmond on cluster (benchmark params)	512 processor cores	132
Desmond on cluster (production params)	512 processor cores	280
Anton (benchmark params)	512 ASICs	4,547

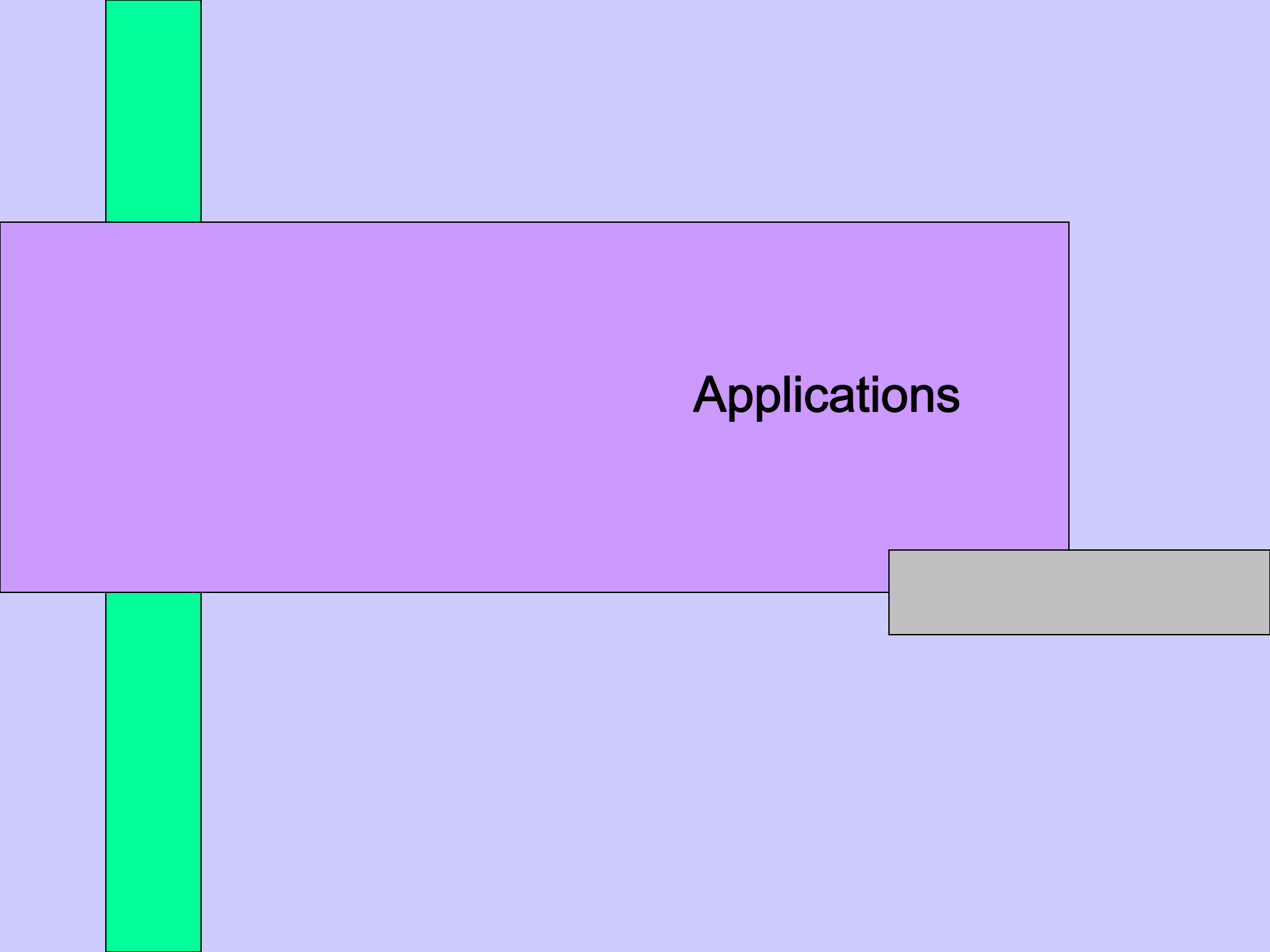
Ref: Shaw, et al., *Proc. ISCA07*, 2007

Comparative Performance of Anton

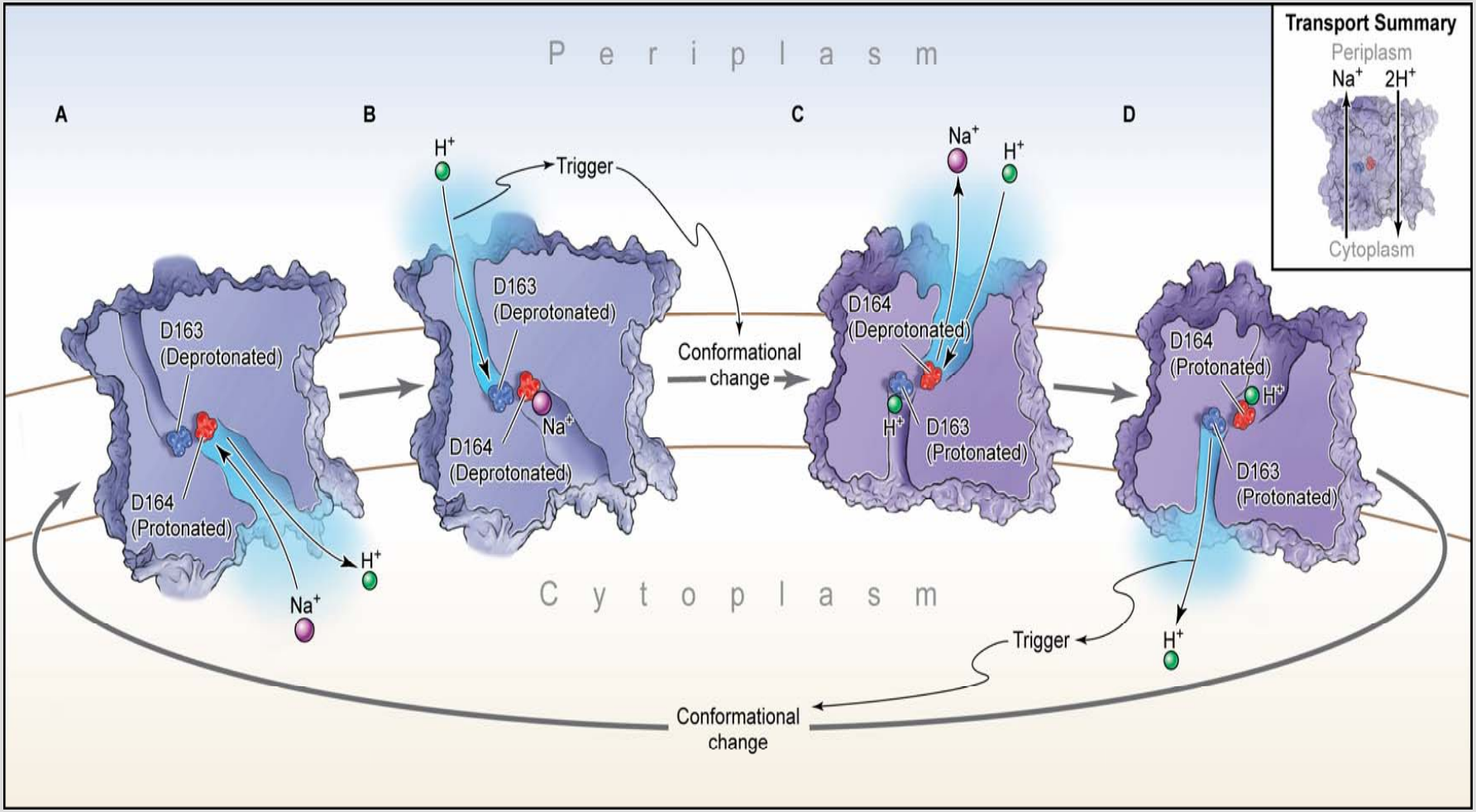
(Dihydrofolate reductase benchmark; 23,558 atoms)

Platform	Size	ns/day
GROMACS on single processor	1 processor core	~1.0
MDGRAPE-3	12 ASICs	3.3
IBM Blue Matter on Blue Gene/L	512 processor cores	4.8 / 9.5
GROMACS 4 (new parallel implementation)	128 processor cores	~42
NAMD on cluster (benchmark params)	512 processor cores	77
Desmond on cluster (benchmark params)	512 processor cores	132
Desmond on cluster (production params)	512 processor cores	280
Anton (benchmark params)	512 ASICs	4,547
Anton (production params)	512 ASICs	14,115

Ref: Shaw, et al., *Proc. ISCA07*, 2007

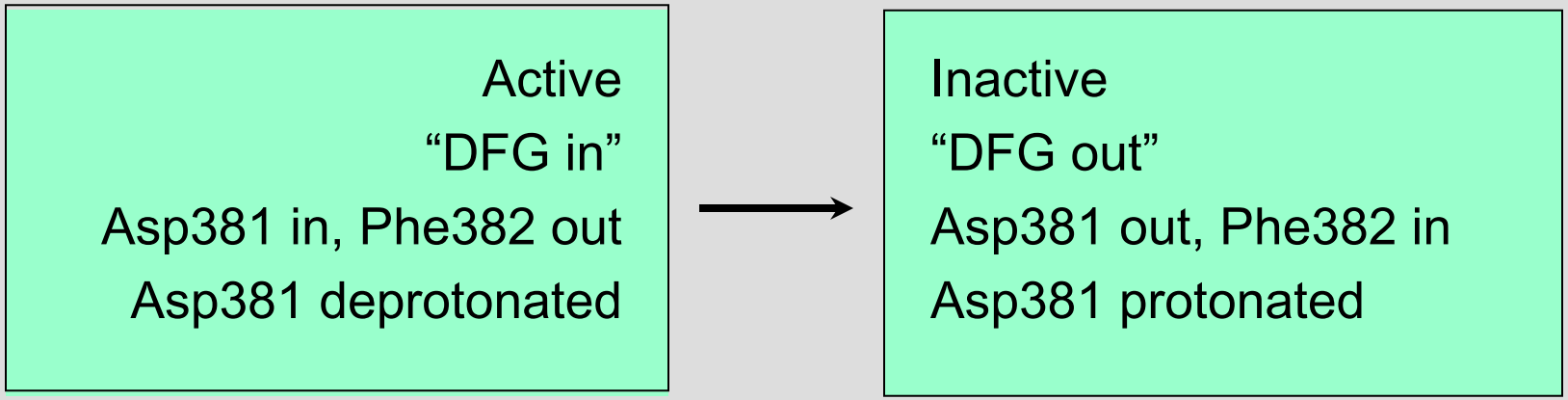

Applications

Mechanism of NhaA Antiporting



Arkin, ..., Shaw, *Science* 317, 799 (2007)

DFG Flip in Abl Kinase



- Experimental predictions:
  - Effects of pH variations
  - Effects of mutations
- Experimental results consistent with predictions
  - John Kuriyan's lab, Berkeley

Folding of Villin Headpiece