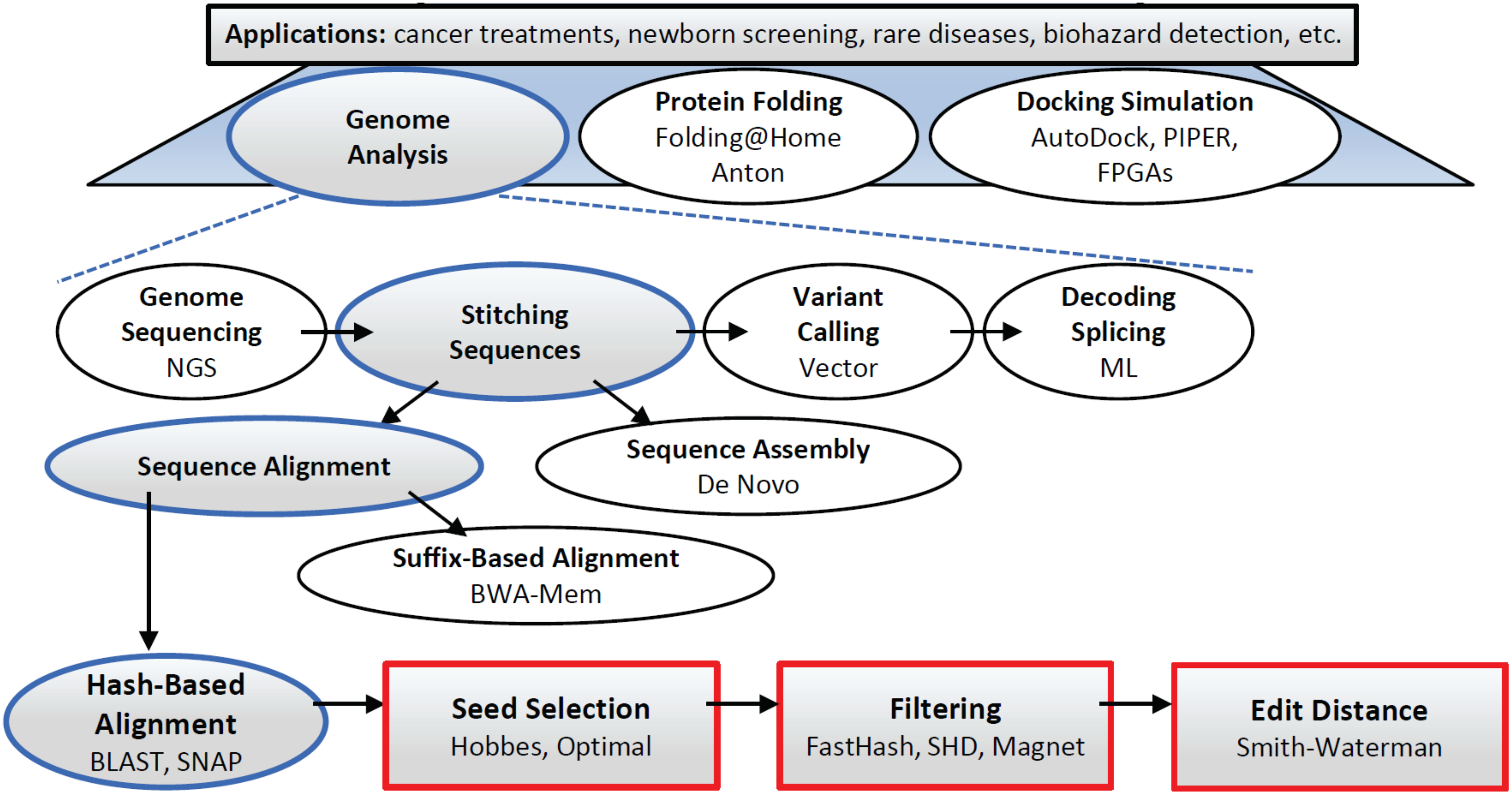
# Lecture: Molecular Dynamics

- Topics: sequence alignment wrap-up, Anton for moldyn

# Precision Medicine



**Applications:** cancer treatments, newborn screening, rare diseases, biohazard detection, etc.

**Genome Analysis**

**Protein Folding**
Folding@Home
Anton

**Docking Simulation**
AutoDock, PIPER,
FPGAs

**Genome Sequencing**
NGS

**Stitching Sequences**

**Variant Calling**
Vector

**Decoding Splicing**
ML

**Sequence Alignment**

**Sequence Assembly**
De Novo

**Suffix-Based Alignment**
BWA-Mem

**Hash-Based Alignment**
BLAST, SNAP

**Seed Selection**
Hobbes, Optimal

**Filtering**
FastHash, SHD, Magnet

**Edit Distance**
Smith-Waterman

# Burrows-Wheeler Transform

A data transformation that is reversible and amenable to compression

Transformation

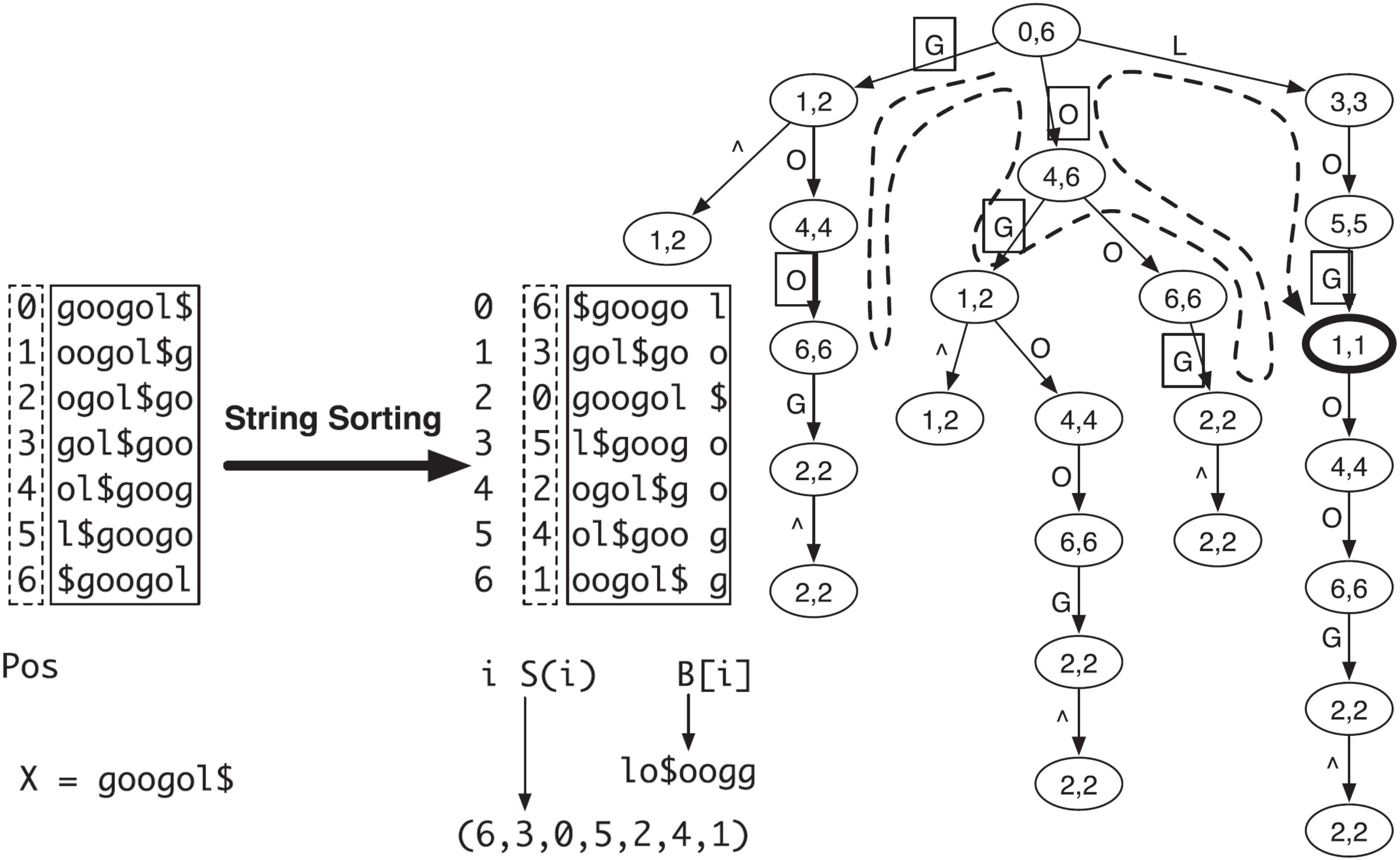
| Input | All rotations | Sorted into lexical order | Taking last column | Output last column |
|---|---|---|---|---|
| ^BANANA\| | ^BANANA\| | **A**NANA\|^B | ANANA\|^**B** | BNN^AA\|A |
| | \|^BANANA | **A**NA\|^BAN | ANA\|^BA**N** | |
| | A\|^BANAN | **A**\|^BANAN | A\|^BANA**N** | |
| | NA\|^BANA | **B**ANANA\|^ | BANANA\|**^** | |
| | ANA\|^BAN | **N**ANA\|^BA | NANA\|^B**A** | |
| | NANA\|^BA | **N**A\|^BANA | NA\|^BAN**A** | |
| | ANANA\|^B | **^**BANANA\| | ^BANANA**\|** | |
| | BANANA\|^ | **\|**^BANANA | \|^BANAN**A** | |

When applied to text, you'll see many consecutive occurrences as below. Note that the last column is highly compressible.

he dog … T

he object … T

# BWT Applied to Sequence Alignment



| Pos | |
|---|---|
| 0 | googol$ |
| 1 | oogol$g |
| 2 | ogol$go |
| 3 | gol$goo |
| 4 | ol$goog |
| 5 | l$googo |
| 6 | $googol |

**String Sorting**

| i | S(i) | | B[i] |
|---|---|---|---|
| 0 | 6 | $googo | l |
| 1 | 3 | gol$go | o |
| 2 | 0 | googol | $ |
| 3 | 5 | l$goog | o |
| 4 | 2 | ogol$g | o |
| 5 | 4 | ol$goo | g |
| 6 | 1 | oogol$ | g |

X = googol$

(6,3,0,5,2,4,1)

lo$oogg

# Burrows Wheeler Alignment

- The trie structure reduces the storage requirement; strings with common suffixes share a portion of the trie

- Because the strings are sorted, each node must only store two numbers to capture the many possible matching locations

- In addition to the above "compression", the many matching locations are determined in O(w) time, where w is the string length; this is true for perfect matches (see example for GO on previous slide)

- For approximate matches, you may have to navigate a significant portion of the trie (see example for LOL on previous slide)

- Note that memory access pattern is quite random, but capacity is low

# Molecular Interactions

- DNA and RNA eventually produce the protein (a sequence of amino acids) – these proteins fold into a particular 3D structure that determines their behavior – gene mutations result in potentially incorrect protein expressions that can result in functionally incorrect folds

- Drug delivery – does it bind to carriers in the blood – does it bind to the appropriate receptors

# Anton

- Research supercomputer by DE Shaw Research to accelerate molecular dynamics simulations; Anton in 2008; Anton 2 in 2013; a “microscope” for biochemistry; specialization before it was cost-effective

# Scale of Mol Dyn

- Simulations are performed in time steps; each time step is a few femto-seconds ($10^{-15}$ seconds)
- Interesting interactions happen at the scale of milli-seconds; must therefore simulate $10^{12}$ steps
- Anton introduces a 1000x speedup, enabling milli-second simulations in a couple of months
- Anton 2 introduces an additional 5x speedup
- 42 kW/rack; 190 W Anton 2 chip; 408 $mm^2$ chip

# Anton Goals

- Trying to do one step involving 10s of thousands of atoms in 1000s of cycles
- New hardware, new algorithms for higher parallelism and lower inter- and intra-chip communication
- Low latency compute and communication
- Can’t use speculation because each step depends on the previous step
- Related: Folding@home explores many parallel trajectories on very short time scales

# Biochem Background

- Modeling a small (repeating) volume of protein, solvent, and other chemicals

- Two significant steps: force calculation and integration; integration applies Newton’s laws and is also non-trivial

- Forces are of two types: bonded and non-bonded

- Bonded: force between an atom and other atoms that share a covalent bond – an atom is typically bonded to <4 atoms, so calculating these forces takes linear time

- Non-bonded forces: van der Waals and electrostatic

# Biochem Background

- Non-bonded forces are the bottleneck because there are a quadratic number of interactions

- van der Waals: weak forces among atoms because of varying electron positions within the atom; these are weak enough that we can ignore them beyond a cutoff radius

- Electrostatic forces cannot be ignored, but we can approximate some of the long-range interactions

- Within a small radius, the pairwise interactions (vdw and es) are calculated; at long-range, they use FFT-based techniques to compute the es forces (charges are mapped to a mesh)

# Parallelization

- The space is partitioned into smaller boxes; each computational "node" is responsible for its home box

- They use a "neutral territory (NT)" method where a node computes close-range interactions for its home box and parts of two neighboring boxes

- The particles in a node are split into "tower" (home box and one neighbor box) and "plate" (home box and the other neighbor box) groups; every element of tower is interacted with every element of plate

- This is the key step that requires the most time and hw

# Amdahl’s Law Analysis

- Must target everything to get 1000x

**Table 1. Profile of a single-processor run using the GROMACS MD package [33].**

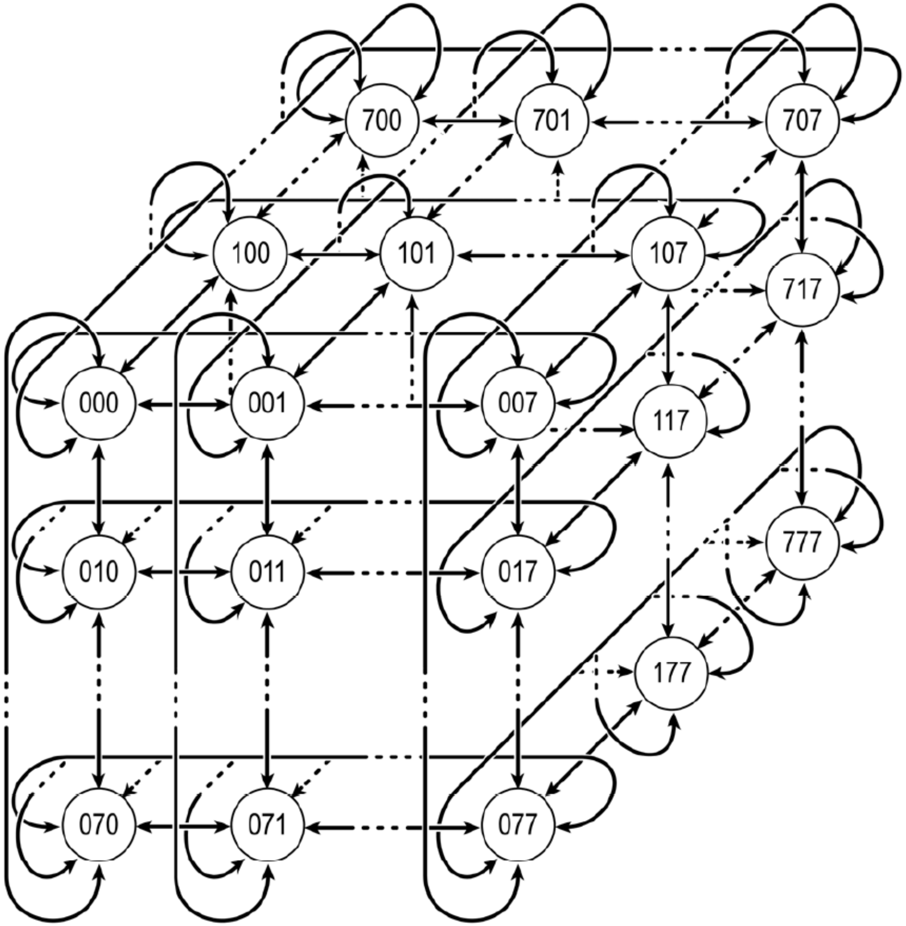
| Phase | Task | % execution time |
|---|---|---|
| Force Calculation | Explicit pairwise nonbonded interactions | 60 |
| | FFT, inverse FFT, & Fourier space multiplication | 17 |
| | Charge spreading and force interpolation | 13 |
| | Bonded force terms | 1 |
| | Correction force terms | 4 |
| Integration | Position & velocity updates | 2 |
| | Constraints | 2 |
| | Pressure computation | 1 |

# Contributions

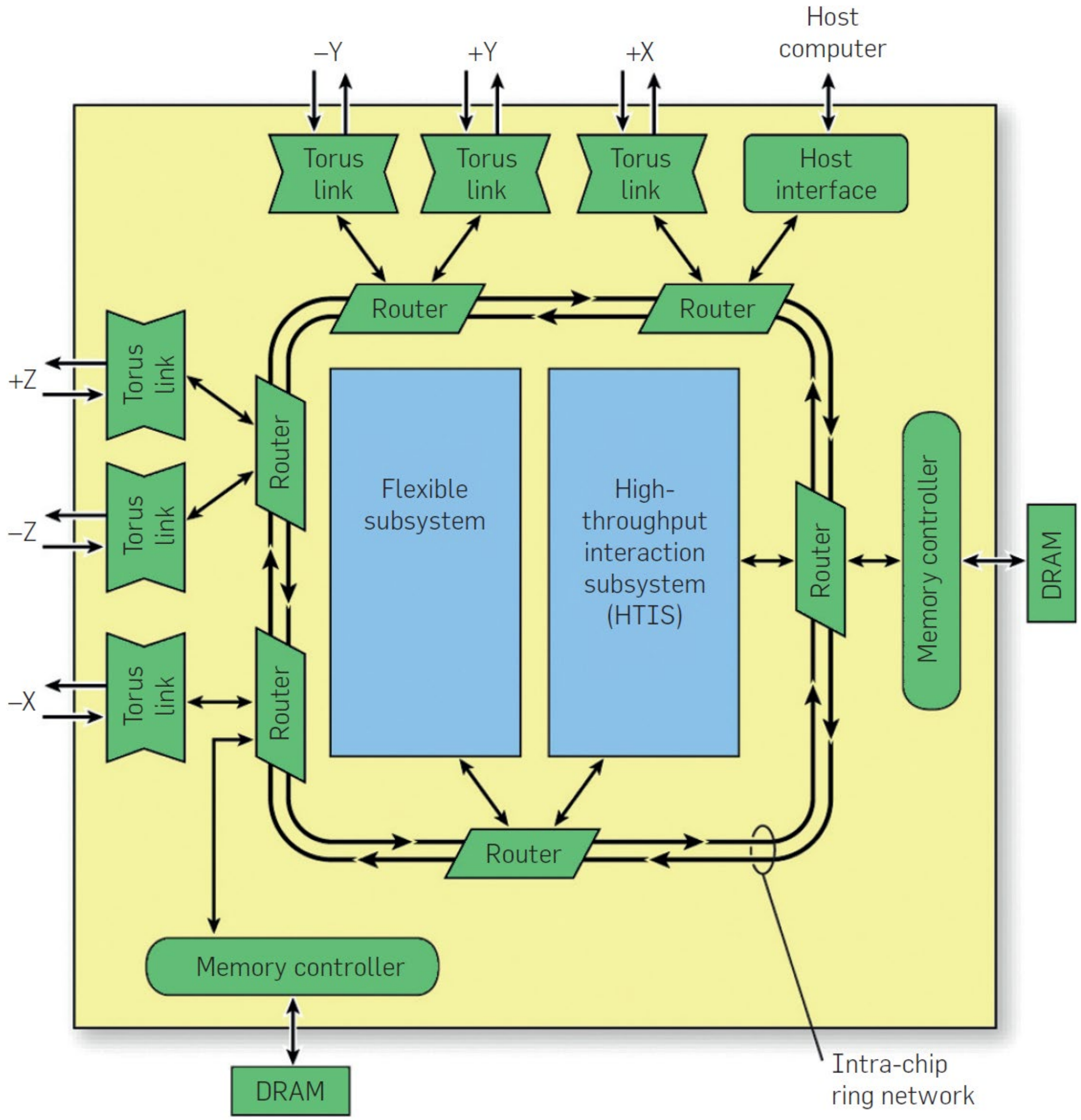
- Specialized pipeline for non-bonded nearby interactions
- Programmable processor with custom ISA for all other operations
- Accumulation at the memory system; little on-chip storage
- Network support for m-cast and compressed sparse packets
- Synchronization and push-based communication

# Architecture

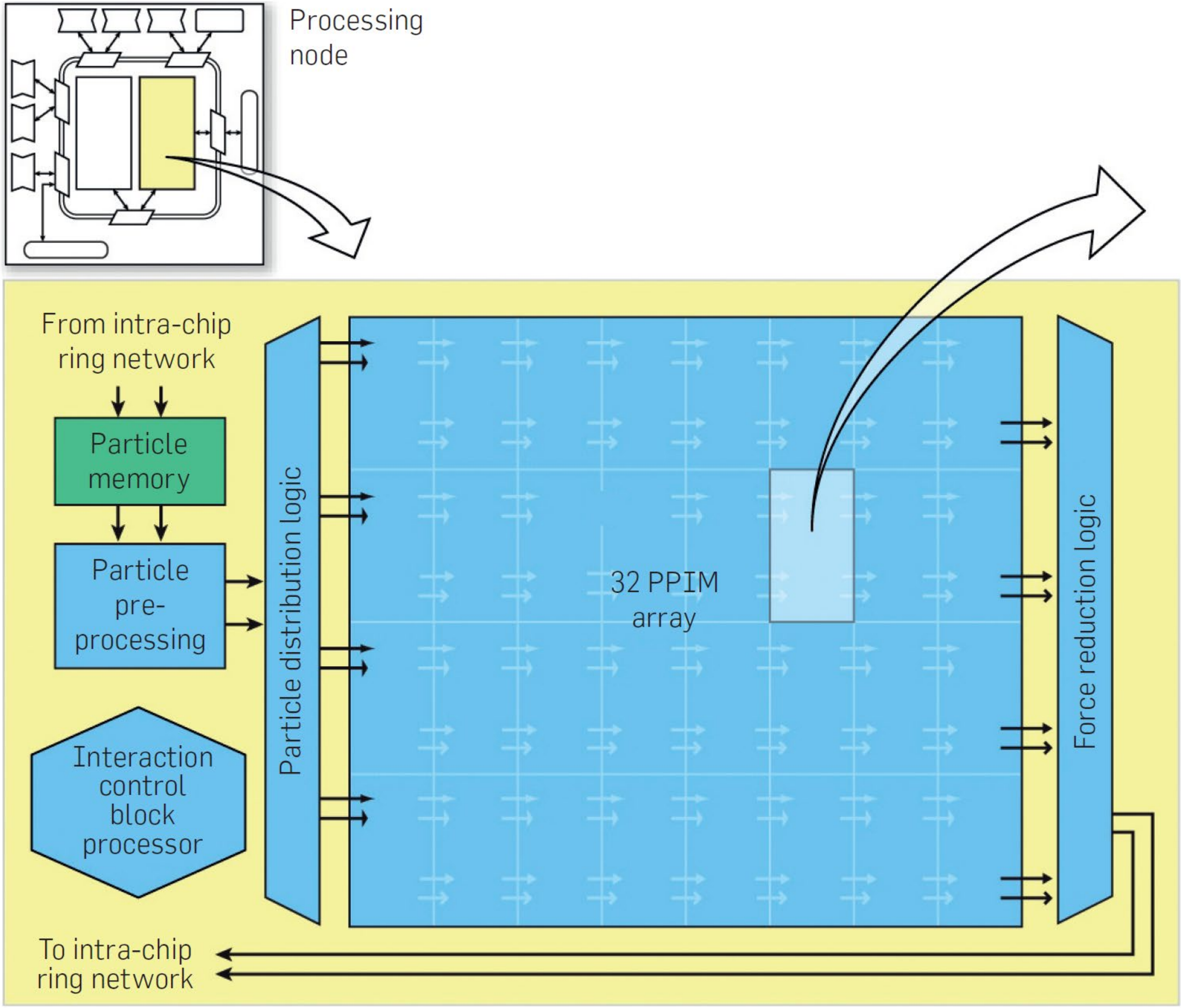
- 512-node torus; a node has a chip, attached DRAM, and 6 ports; 4 nodes per board; 32 boards per rack; 4 racks



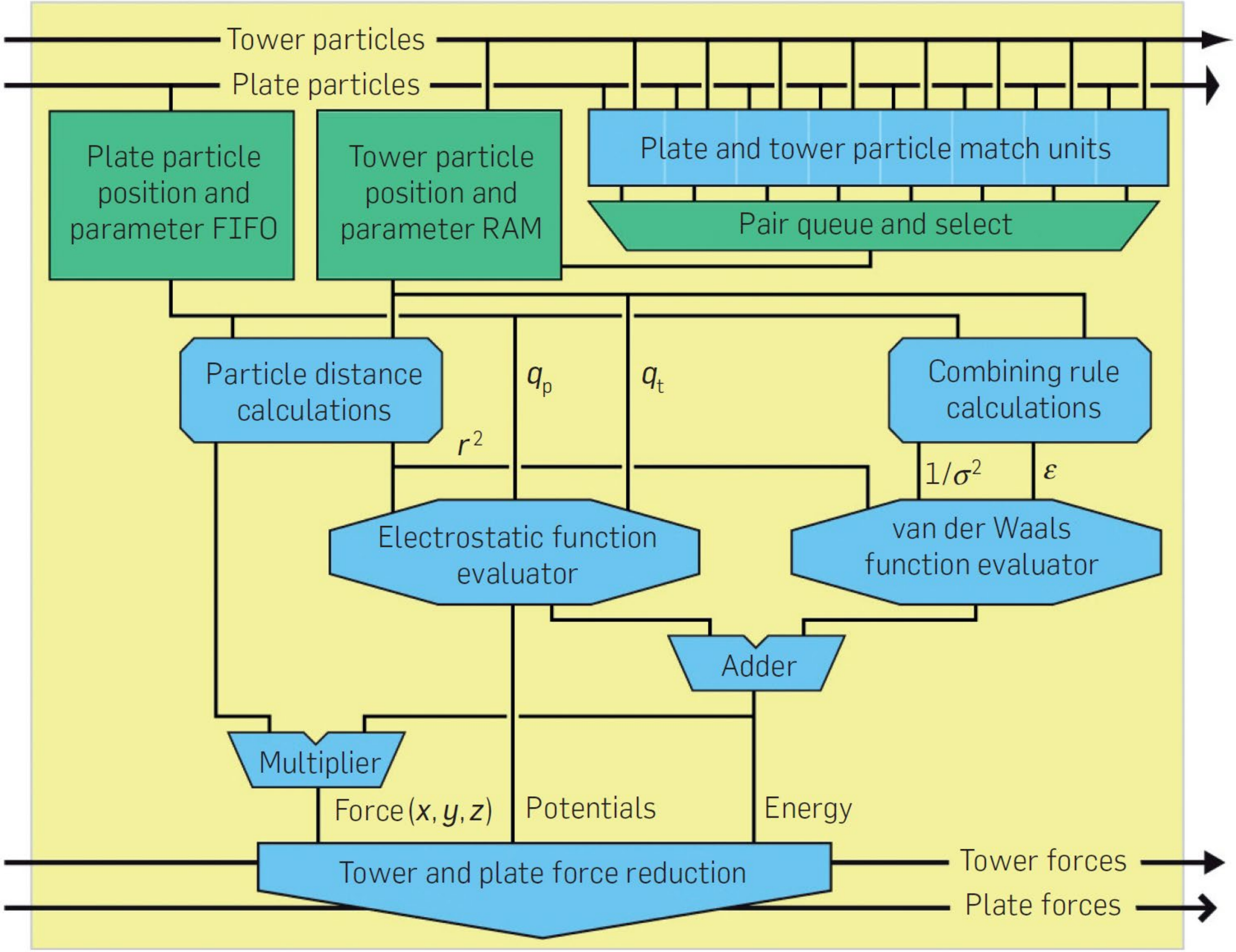
# Anton Chip


Host computer
–Y
+Y
+X
Torus link
Torus link
Torus link
Host interface
Router
Router
+Z
Torus link
–Z
Torus link
–X
Torus link
Router
Router
Flexible subsystem
High-throughput interaction subsystem (HTIS)
Router
Memory controller
DRAM
Router
Memory controller
DRAM
Intra-chip ring network

# HTIS – Hi-Thruput Interaction Subsystem



# HTIS – Hi-Thruput Interaction Subsystem

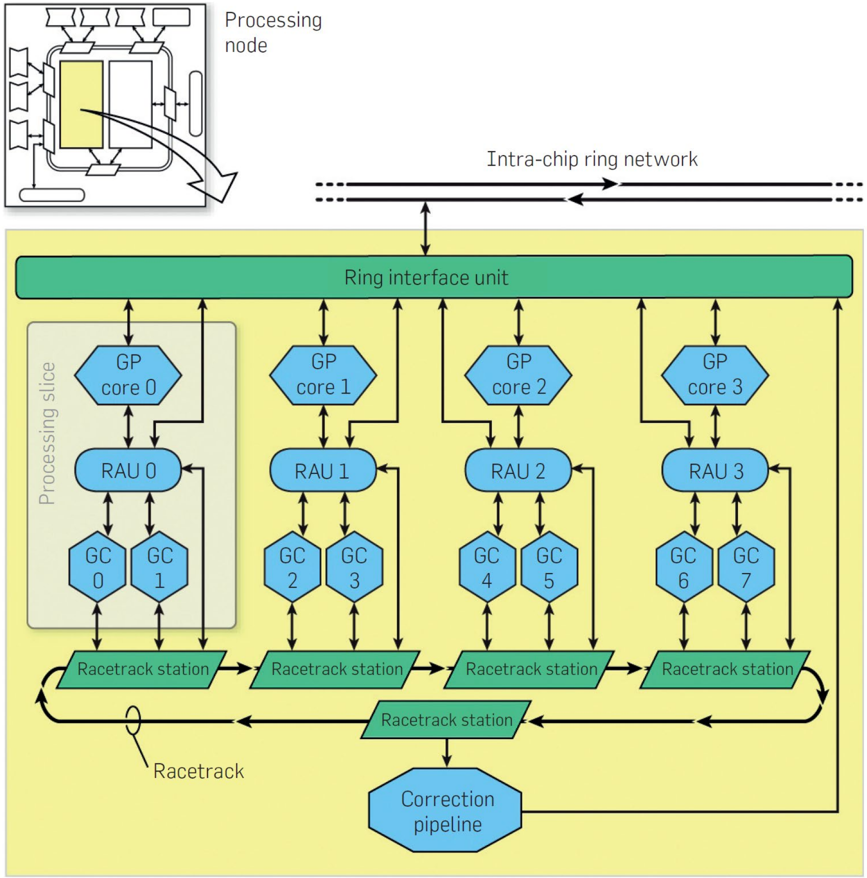


# HTIS – Hi-Thruput Interaction Subsystem

- Particles are received through push-communication from neighbors
- ICB buffers these and controls the systolic array computations
- 32 PPIMs store parts of the tower particles; the plates are replicated four times and are streamed to each of the rows
- At each step, 8 matches are performed to identify the pairs that must be processed; these are then sent thru the rest of the pipeline
- 26-stage PPIM pipeline at 800 MHz; rest of the chip operates at 400 MHz; the pipeline is fixed-function and uses variable precision
- Each PPIM stores the forces collected for the tower; the plates carry their forces as they move to the right; the forces from the four rows are merged at the end and sent to the appropriate buffers; at the end, the tower forces are streamed out as well
- DRAM is only used when spilling very large simulations; on-chip cache is enough to handle most working sets

# Flexible Subsystem

- Responsible for several complex operations; the GP cores handle control flow; the remote access units have scratchpads and take care of data movement; the geometry cores take care of SIMD calculations; the racetrack is used for intra-node communication



# Speedups

| Task | Anton 512-node speedup over GROMACS | Divided by 512 (single-node speedup) |
|---|---:|---:|
| Range-limited non-bonded forces | 50,000 | 98.5 |
| Charge spreading and force interpolation | 5,400 | 10.5 |
| FFT and Fourier space multiplication | 3,700 | 7.2 |
| Bond forces | 4,600 | 9.0 |
| Correction forces | 3,700 | 7.2 |
| Position and velocity updates | 1,200 | 2.4 |
| Constraint calculations | 1,400 | 2.7 |
| Temperature computation | 400 | 0.8 |
| Total | 9,000 | 17.6 |

# Anton 2

| | ANTON | ANTON2 |
|---|---|---|
| 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 |

| | | |
|---|---|---|
| Power: | 75 W | 190 W |
| Size: | 300 mm$^2$ | 408 mm$^2$ |

# The Molecular Dynamics Computation

1. Divide time into uniform timesteps (~ 1 fs)
2. Compute forces on all particles in system in a timestep

E= *Bonded:* *(Flexible Subsystem)*

$\Sigma\ k_b(r-r_0)^2$ Stretch

$+\ \Sigma\ k_\theta(\theta-\theta_0)^2$ Bend

$+\ \Sigma\ A[1+\cos(n\tau-\varphi)]$ Torsion

*Non-Bonded:* *(High Throughput Interaction Subsystem)*

$+\ \Sigma\ \Sigma\ q_i q_j / r_{ij}$ Electrostsatic

$+\ \Sigma\ \Sigma\ q_i q_j / r_{ij}$ Van der Waals

3. Based on these forces, compute new velocities and positions for all particles

$V = V_0 + F * \Delta T / M$ *(Flexible Subsystem)*

$S = S_0 + V_0 * \Delta T + F * (\Delta T)^2 / M$

4. Rinse and repeat... About 1 TRILLION times to simulate a millisecond!

*Also, do a bunch of ugly housekeeping functions along the way to maintain (as needed for experiment) constant temperature and / or pressure, etc.*

# References

- Anton papers and presentations at: ISCA’07, HotChips’08, SC’09
- Anton 2 papers at: HotChips’14, SC’14

# Title

- Bullet