The Computational Microscope

Computational microscope views at atomic resolution ...

... how living cells maintain health and battle disease

Our Microscope is Made of...

Chemistry

$$V(R) = \sum_{\text{inter}} E_{\text{LJ}}(r_{ij}) \ + \ \sum_{\text{elec}} E_{\text{elec}}(r_{ij}) \ + \ \sum_{\text{bonds}} E_{\text{bond}}(\theta_{ij})$$

$$= \sum_{\text{inter}} (4\epsilon_{\text{LJ}} [(\sigma_{\text{LJ}}/r_{ij})^{12} - (\sigma_{\text{LJ}}/r_{ij})^{6}] \ + \ \sum_{\text{elec}} (\epsilon_{\text{elec}} [\cos(\phi_{ij}) - 1] \ + \ \sum_{\text{bonds}} (\epsilon_{\text{bond}} [\cos(\theta_{ij}) - 1])$$

Physics

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\nabla U(R)$$

Math

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \vec{F}_i(t)$$

(repeat one billion times = microsecond)
NAMD impact is broad and deep

• Comprehensive, industrial-quality software
  – Integrated with VMD for simulation setup and analysis
  – Portable extensibility through Tcl scripts (also used in VMD)
  – Consistent user experience from laptop to supercomputer

• Large user base – 51,000 registered users
  – 9,100 (18%) are NIH-funded; many in other countries
  – 14,100 have downloaded more than one version

• Leading-edge simulations
  – “most-used software” on NICS Cray XT5 (largest NSF machine)
  – “by far the most used MD package” at TACC (2nd and 3rd largest)
  – NCSA Blue Waters early science projects and acceptance test
  – Argonne Blue Gene/Q early science project

Outside researchers choose NAMD and succeed


180K-atom 30 ns study of anesthetic binding to bacterial ligand-gated ion channel provided "complementary interpretations...that could not have been deduced from the static structure alone."

2100 external citations since 2007

Voth, et al., PNAS, 2010

500K-atom 500 ns investigation of effect of actin depolymerization factor/cofilin on mechanical properties and conformational dynamics of actin filament.

Recent NAMD Simulations in Nature

• M. Koeksal, et al., Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. (2011)
• C.-C. Su, et al., Crystal structure of the CusRA heavy-metal efflux complex of Escherichia coli. (2011)
• D. Slade, et al., The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. (2011)
• F. Rose, et al., Mechanism of copper(II)-induced misfolding of Parkinson’s disease protein. (2011)
• L. G. Cuello, et al., Structural basis for the coupling between activation and inactivation gates in K(+) channels. (2010)
• R. H. P. Law, et al., The structural basis for membrane binding and pore formation by lymphocyte perforin. (2010)
• P. Dalhaimer and T. D. Pollard, Molecular Dynamics Simulations of Arp2/3 Complex Activation. (2010)
The Molecular Dynamics Simulation Process

For textbooks see:

More at http://www.biomath.nyu.edu/index/course/99/textbooks.html

---

Classical Dynamics

*at 300K*

Energy function:  \[ U(\vec{r}_1, \vec{r}_2, \cdots \vec{r}_N) = U(\vec{R}) \]

used to determine the force on each atom:

\[ m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla}U(\vec{R}) \]

yields a set of 3N coupled 2\textsuperscript{nd}-order differential equations that can be propagated forward (or backward) in time.

Initial coordinates obtained from crystal structure, velocities taken at random from Boltzmann distribution.

Maintain appropriate temperature by adjusting velocities.
Classical Dynamics

discretization in time for computing

\[ m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\nabla U(\vec{R}) \]

Use positions and accelerations at time t and the positions from time t-\( \delta t \) to calculate new positions at time t+\( \delta t \).

\[
\begin{align*}
\vec{r}(t + \delta t) & \approx \vec{r}(t) + \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)\delta t^2 \\
\vec{r}(t - \delta t) & \approx \vec{r}(t) - \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)\delta t^2
\end{align*}
\]

“Verlet algorithm”

\[ \vec{r}(t + \delta t) \approx 2\vec{r}(t) - \vec{r}(t - \delta t) + \vec{a}(t)\delta t^2 \]

Potential Energy Function

of Biopolymer

- Simple, fixed algebraic form for every type of interaction.
- Variable parameters depend on types of atoms involved.

\[
U(\vec{R}) = \sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2 + \sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2 + \sum_{\text{dihedrals}} k_i^{\text{dihedral}} [1 + \cos (n_i \phi_i + \delta_i)] + \sum_{i \neq j} 4\epsilon_{ij} \left[ \frac{\sigma_{ij}}{r_{ij}} \right]^{12} - \left[ \frac{\sigma_{ij}}{r_{ij}} \right]^6 + \sum_{i \neq j} q_i q_j \epsilon r_{ij}
\]

heuristic

from physics
Atomic polarizability increases computation by 2x…

…but, the additional computations are perfectly suited to the GPU!

For now, NAMD calculates atomic polarizability on CPUs only...soon we will also use GPUs.

Atomic polarizability of water, highly accurately simulated through additional particles (shown in green).

NAMD CPU performance scaling

<table>
<thead>
<tr>
<th></th>
<th>polarizable water</th>
<th>non-polarizable water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seconds per step</td>
<td>0.01</td>
<td>1</td>
</tr>
<tr>
<td>CPU cores</td>
<td>100</td>
<td>1000</td>
</tr>
</tbody>
</table>

Molecular Dynamics Ensembles

Constant energy, constant number of particles (NE)

Constant energy, constant volume (NVE)

Constant temperature, constant volume (NVT)

Constant temperature, constant pressure (NPT)

Choose the ensemble that best fits your system and start the simulations, but use NE to check on accuracy of the simulation.
Langevin Dynamics
for temperature control

Langevin dynamics deals with each atom separately, balancing a small friction term with Gaussian noise to control temperature:

\[ m \ddot{\vec{r}} = \vec{F}(\vec{r}) - \gamma m \dot{\vec{r}} + \vec{R}(t) \]

\[ \langle \vec{R}(t) \cdot \vec{R}(t') \rangle = 6k_B T \gamma \delta(t - t') \]

Langevin Dynamics
for pressure control

Underlying Langevin-Hoover barostat equation for all atoms:
Equations solved numerically in NAMD

\[ \frac{d^2 V(t)}{dt^2} = \frac{1}{W_{bs}} [P(t) - P_{\text{target}}] - \frac{1}{\tau_{bs}} \frac{dV(t)}{dt} + R_{bs}(t) \]

\[ P = \rho k_B T + \frac{1}{V} d \sum_{i<j} \left\langle r_{ij} \frac{dU_{\text{tot}}(r_{ij})}{dr_{ij}} \right\rangle \quad d = \text{dimension} \]

\[ \langle R_{bs}(t) R_{bs}(t') \rangle = \frac{2k_B T_{\text{target}} \delta(t - t')}{W_{bs} \tau_{bs}} \quad W_{bs} = d N_{\text{atoms}} k_B T_{\text{target}} \tau_{period}^2 \]

\[ \dot{\vec{r}}_i = \vec{v}_i + s \vec{r}_i \quad \dot{\vec{v}}_i = \vec{F}_i / m_i - s \vec{v}_i \]

\[ \dot{V} = dV_s \quad \dot{s} = dV(P - P_{\text{target}}) / W - s / \tau_{bs} + R(t) \]

\[ d - \text{dimension} \]
Figure 2.2: The size of biomolecular systems that can be studied using all-atom MD simulations has steadily increased from that of Lysozyme (40,000 atoms) in 1990 to the million-atom STMV virus capsid in 2006, and now 100 million atoms as in the spherical chromatophore model shown above. Atom counts include aqueous solvent, not shown.

NAMD is an easy-to-use parallel MD software program developed by the University of Illinois at Urbana-Champaign. The non-commercial version of NAMD is distributed as free software, with source code included. The core NAMD software is written in C and employs a “stand-alone” turnkey approach that requires only a few input files and a standard mpiexec program to simplify launching NAMD within the cluster’s batch environment. In addition, NAMD provides cross-platform extensibility through Tcl scripting, a language users are familiar with from its use in VMD, allowing custom simulation protocols to be implemented without recompiling the program. So prized is the usability of NAMD that its commands, input formats, and Tcl interface are reproduced (as allowed by the NAMD open-source license) in the (closed-source) GPU-based ACEMD software [108], providing NAMD users with an alternate MD engine for, e.g., studies of protein-ligand binding [109].

Significant parallel computing resources are available to NAMD users. The most popular platform is small Linux clusters (less than 100 nodes), used by 48% of survey respondents, followed by 26% using large clusters (including supercomputers), and a significant minority of 4% each using NAMD on Cray or IBM Blue Gene supercomputers. NAMD is maintained by the Center on the machines of the NSF supercomputer centers, on which NAMD users receive significant allocations. NAMD is the “most used software” on the largest NSF-funded supercomputer, representing 9% of total usage on the 112,896-core Cray XT5 of the National Institute for Computational Sciences (see NICS letter), and “by far the most used molecular dynamics package” at the Texas Advanced Computing Center (see TACC letter), which hosts the second and third largest NSF machines.

The Center has long been an early adopter and advocate of low-cost commodity computing technology, acquiring and using networks of workstations in 1993, Linux clusters in 1998, and graphics processors since 2007 [155]. Graphics processor acceleration is now ranked the most important area of future development by NAMD users, an amazing reversal from 2006 in which hardware acceleration was ranked last. NAMD loads the most time-consuming operations to NVIDIA GPUs while preprocessing.

Codes: NAMD/VMD 260,000 registered users, same user interface from laptop to BW, busiest code NSF centers

12 ns / day with GPU acceleration
100 M atom simulation
NAMD Scalability

![Graph showing the scalability of NAMD with increasing number of cores and time per day.](image)

From 10,000 to 100,000 Atom MD in 2000

100k atom MD reached in 2000
- then a factor 10 increase in computation;
- **needed to describe membrane processes**;
- was achieved through cluster computing;
- produced good quality results for aquaporin;
- is now standard.


From 100,000 to 64,000,000 Atom MD Now

- All-atom structure of mature HIV capsid
- 216 hexamers + 12 pentamers, pdb 3J3Q
- 64 million atoms total
- Run on 2000 Cray-XK nodes (GPU accelerated) at 12 ns / day

All-atom MD Simulation of HIV-1 Capsid

- 216 hexamers + 12 pentamers, pdb 3J3Q (available May 29)
- 64 million atoms total
- Over 100 ns of MD on NSF Blue Waters – 5000 Nodes, 160,000 cores - 10ns/day

Capsid structure stable without constraints!
1M Atom Virus on TitanDev GPU

100M Atoms on TitanDev
Large is no problem. But …

Molecular dynamics simulation of alpha-hemolysin with about 300,000 atoms; 1 million atom simulations are routine today, 20 million atom simulations are possible.
But long is still a problem!

*biomolecular timescale and timestep limits*

Rotation of buried sidechains
Local denaturation
Allosteric transitions

small protein folding

Hinge bending
Rotation of surface sidechains
Elastic vibrations

Bond stretching
Molecular dynamics timestep

$\delta t = 1 \text{ fs}$

<table>
<thead>
<tr>
<th>Unit</th>
<th>Value</th>
<th>Time Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>$10^{15}$</td>
<td>(NSF center, Shaw Res.)</td>
</tr>
<tr>
<td>ms</td>
<td>$10^{12}$</td>
<td>(30 years, 2 months)</td>
</tr>
<tr>
<td>µs</td>
<td>$10^{9}$</td>
<td>(10 days, 2hrs)</td>
</tr>
<tr>
<td>ns</td>
<td>$10^{6}$</td>
<td>(15 min)</td>
</tr>
<tr>
<td>ps</td>
<td>$10^{3}$</td>
<td></td>
</tr>
<tr>
<td>fs</td>
<td>$10^0$</td>
<td></td>
</tr>
</tbody>
</table>