Single vs. Double Precision in MD Simulations: Correlation Depends on System Length-Scale

Anqi Zou[1], Tyson J. Lipscomb[1], and Samuel S. Cho[1,2]

[1]Departments of Computer Science and [2]Physics

Wake Forest University, Winston-Salem, NC, USA

Introduction to Molecular Dynamics (MD) Simulations

Understanding how biomolecules transition from an unstructured unfolded state to a well-defined structured folded state is one of the grand challenges in biology because biomolecules sometimes misfold to cause diseases such as Alzheimer's and Parkinson's disease. Other times, the biomolecules themselves cause a deleterious effect, such as HIV-1 protease, a protein that is integral in HIV replication. The motions of biomolecules can be viewed at a microscopic scale using molecular dynamics simulations (MD) if the underlying computations are accurately calculated. By current standards, experiments can directly observe only the unfolded and folded states, and MD simulations can help fill in the critical blanks.

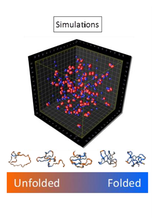
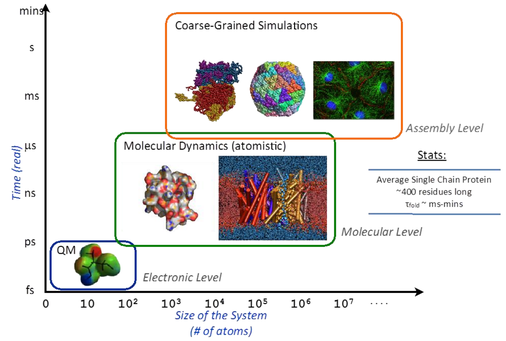


Table 1 Neurodegenerative diseases: proteins and pathology

Disease	Etiology	Regions most affected	Characteristic pathology	Disease proteins deposited
Huntington's disease	Huntingtin (dominant)	Striatum, other basal ganglia, cortex, other regions	Intranuclear inclusions and cytoplasmic aggregates	Huntingtin with polyglutamine expansion
Other polyglutamine diseases (DRPLA, SCA1–3, etc., SBMA)	Atrophin-1, ataxin-3, -3, etc.; androgen receptor (AR) (dominant)	Basal ganglia, brain stem cerebellum, and spinal cord	Intranuclear inclusions	Atrophin-1, ataxins or AR
Alzheimer's disease (AD)	Sporadic (ApoE risk factor)	Cortex, hippocampus, basal forebrain, brain stem	Neuritic plaques and neurofibrillary tangles	Aβ peptide (from APP) and hyperphosphorylated tau
	Amyloid precursor protein (APP) (dominant)	Same as sporadic	Same as sporadic	Same as sporadic
	Presenilin 1, -2 (dominant)	Same as sporadic	Same as sporadic	Same as sporadic
Fronto-temporal dementia with Parkinsonism	Tau mutations (dominant)	Frontal and temporal cortex, hippocampus	Pick bodies	Hyperphosphorylated tau protein
Parkinson's disease (PD)	Sporadic	Substantia nigra, cortex, locus caeruleus, raphe, etc.	Lewy bodies and Lewy neurites	α-Synuclein
	α-Synuclein (dominant)	Similar to sporadic, but more widespread	Similar to sporadic	α-Synuclein
	Parkin (also DJ-1, PINK1) recessive (some dominant)	Substantia nigra	Lewy bodies absent (or much less frequent)	α-Synuclein (when present)
Amyotrophic lateral sclerosis (ALS)	Sporadic	Spinal motor neurons and motor cortex	Bunina bodies and axonal spheroids	Unknown (neurofilaments)
	Superoxide dismutase-1 (dominant)	Same as sporadic	Same	Unknown
Prion diseases (kuru, CJD, GSS disease, fatal familial insomnia, new variant CJD)	Sporadic, genetic and infectious	Cortex, thalamus, brain stem, cerebellum, other areas	Spongiform degeneration, amyloid, other aggregates	Prion protein

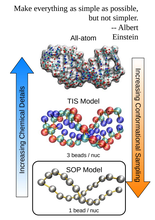
ApoE, apolipoprotein E; APP, amyloid precursor protein; CJD, Creutzfeldt–Jakob disease; DRPLA, dentatorubral and pallidoLuysian atrophy; GSS, Gerstmann–Straussler–Scheinker; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia.

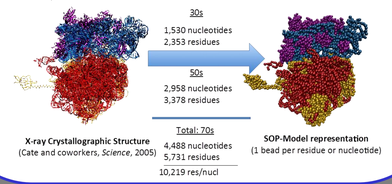
Length- and Time-Scales of MD Simulations



Coarse-Grained Simulations

We performed coarse-grained simulations of the ribosome, a biomolecular machine responsible for synthesizing proteins in the cell. The general idea is to reduce the degrees of freedom that are negligible for its folding and assembly while still retaining the main features of the biomolecule. That way, we can simulate biologically relevant sized systems on assembly timescales. We implemented the Self-Organized Polymer (SOP) model, in which each residue or nucleotide is represented by a bead.

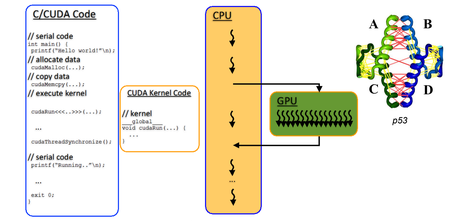




X-ray Crystallographic Structure
(Cate and coworkers, *Science*, 2005)

SOP-Model representation
(1 bead per residue or nucleotide)

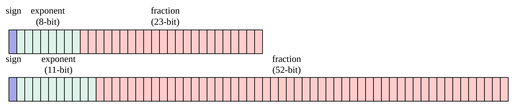
GPU-Optimized MD Simulations



To further increase the performance of our simulations, we developed a GPU-optimized MD simulation software. The code is implemented in the CUDA programming language that performs parallel computations on the GPU via kernel calls that execute a single instruction on multiple data. This paradigm is ideally suited for MD simulations because the interactions between pairs of interacting residues or nucleotides are independent calculations.

A fundamentally important issue is the accuracy and precision of the calculations that determine the fidelity of the computations in MD simulations.

Single vs. Double Precision Overview



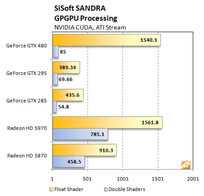
It is well known that computers can only represent numbers using a limited number of significant figures. Thus irrational numbers, such as π, cannot be represented as its exact mathematical value in computer memory. The discrepancy caused by this representation limitation is called "round-off error". Often computers have the capacity to represent numbers in two types of precisions, single (32-bit) and double (64-bit) precision.

Single precision uses 23 bits to present the significant digits and can represent to about 7 decimal places, while double precision uses 52 bits to represent about 15 decimal digits. The finite number of available bits limits the precision of a numerical representation. For example, π is represented as 3.141593 in single precision, and 3.141592653589793 in double precision. In general, using double precision reduces the consequences of rounding error. However, memory and computational cost associated with double precision could be very expensive, so single precision calculations are preferred when the precision of the calculation is not required to be as high.

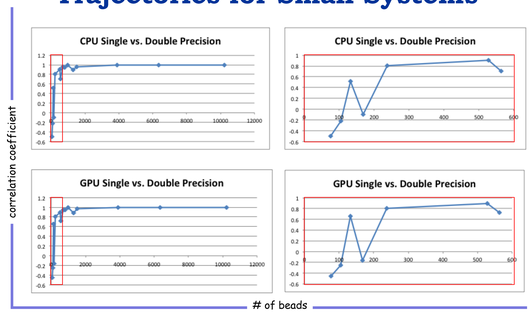
Nearly all modern computer architectures, including NVIDIA GPUs, implement the IEEE 754 standard for binary floating point arithmetic. The IEEE 754 standard guarantees that addition, subtraction, multiplication, division, comparisons, etc. are the same for a given format and rounding mode. Even still, round-off errors are inherent in all computer architectures because the order of operations can affect the accuracy of the final result.

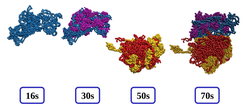
Motivation:

To evaluate how single vs. double precision operations affect MD simulations, we performed coarse-grained MD simulations of many biologically relevant systems of various size. Our main goal is to determine when single precision calculations would be appropriate and when they would not.



Single vs. Double Precision MD Simulations Give Different Trajectories for Small Systems





We calculated the Pearson correlation coefficient (r) of the coordinates from our MD simulation trajectories. When r = 1, the correlation is highest, indicating that they are exactly the same, while r = 0 indicates no correlation and therefore significant differences. For systems larger than ~1,000 residues/nucleotides, there exist almost no difference between single and double precision MD simulations. However, for smaller systems, there exists significant differences, and a clear trend does not exist using this measure.

Prion (104 residues)
Putative Agent of Mad Cow Disease

HIV-Protease (132 residues)
HIV replication

P53 Tumor Suppressor (168 residues)
Blocks tumor growth

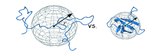
Green Fluorescent Protein (238 residues)
Biosensor Protein

Acetylcholinesterase (528 residues)
Target of Alzheimer Dementia Drugs

Deoxyhemoglobin (564 residues)
Sickle Cell Anemia

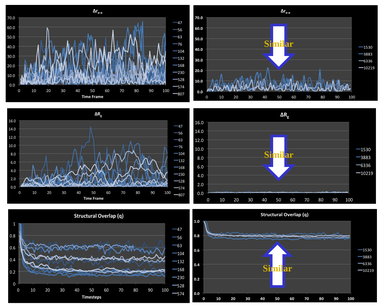
Structural Analysis of Trajectories

Radius of Gyration (R_g):

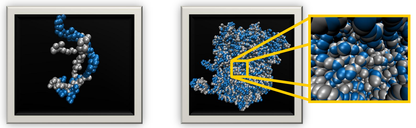


Structural Overlap (q)

$$q = \frac{1}{(N-1)(N-2)} \sum_{i \neq j-1}^{N} \exp\left[-\frac{(r_{ij} - r_{ij}^0)^2}{\sigma_{ij}^2}\right]$$



We performed MD simulations using single vs. double precision computations, and we quantified the difference between the single vs. double precisions frames using three different structural measures: difference in end-to-end distance ($\Delta r_{e\text{-}e}$), difference in radius of gyration (ΔR_g), and structural overlap (q). For each measure, the difference between each frame from the single precision and double precision MD simulations are pronounced for smaller systems but negligible for systems consisting for more than ~1000 beads. For smaller systems, there is not a clear relationship between the structural similarity and system size.



Prion (104 beads)
RMSD: 21.7 Å
$\Delta r_{e\text{-}e}$: 27.4 Å
ΔR_g: 7.95 Å
qscore: 0.14

70s Ribosome (10,219 beads)
RMSD: 4.4 Å
$\Delta r_{e\text{-}e}$: 0.79 Å
ΔR_g: 0.01 Å
qscore: 0.80

Conclusions

We performed coarse-grained MD simulations of biologically relevant systems of various size using single and double precision variables to perform the computations. The structures from the MD simulations are quantified using three different measures of structural similarity over the course of the simulations. By comparing each pairs of frames from the single precision vs. double precision MD simulations, we observe that, for large systems, the difference in precision results in negligible difference in the coordinates of the MD simulations during 1,000,000 timesteps of simulations. For smaller systems, however, we observe significant differences in the computations that can result in very different structures, even after only a relatively few number of timesteps. As such, we recommend double-precision calculations for studying small systems that will result in a performance penalty but accurate trajectories. For larger systems, single-precision calculations give nearly identical results so double-precision calculations are unnecessary for MD simulations. Therefore, the increased performance of single-precision implementations of MD simulations makes no significant difference in the accuracy and precision of MD simulations if the system size is sufficiently large. Although we use coarse-grained MD simulations in our present study, our results are general for all MD simulations.

Future Directions

- Round-off error correction for smaller systems.
- Evaluate other MD simulation types to determine the size-dependent accuracy and precision limitations.

Acknowledgements

- (TJL) Wake Forest University Computer Science Graduate Fellowship for Excellence
- (SSC) NVIDIA Academic Partnership
- (SSC) Wake Forest University Science Research Fund