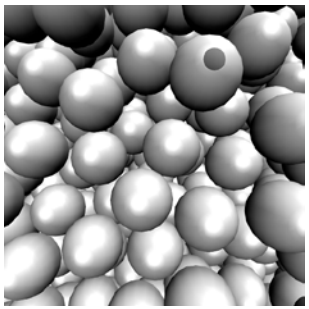
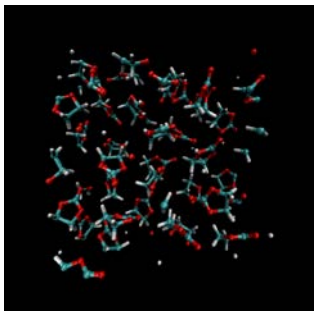
Quantifying Uncertainties in Equilibrium Particle Dynamics Simulations

Changho Kim (김창호 / 金彰鎬)

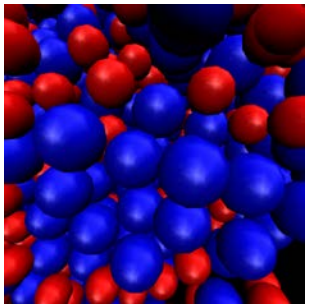
- Division of Applied Mathematics, Brown University, USA
- Center for Computational Sciences and Engineering (CCSE), Lawrence Berkeley National Laboratory, USA



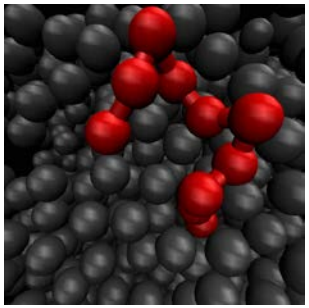
LJ fluid



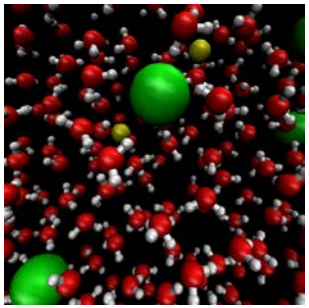
Ethylene carbonate liquid



Ar/Kr mixture

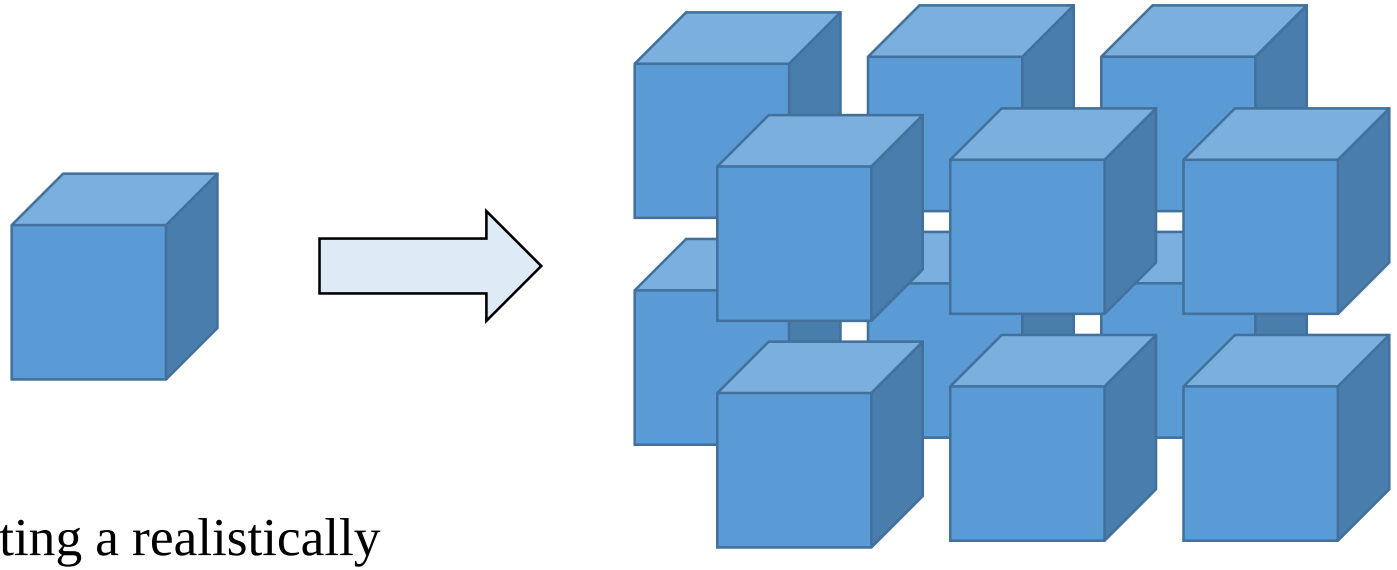


FENE chains in a WCA fluid



H_2O + NaCl

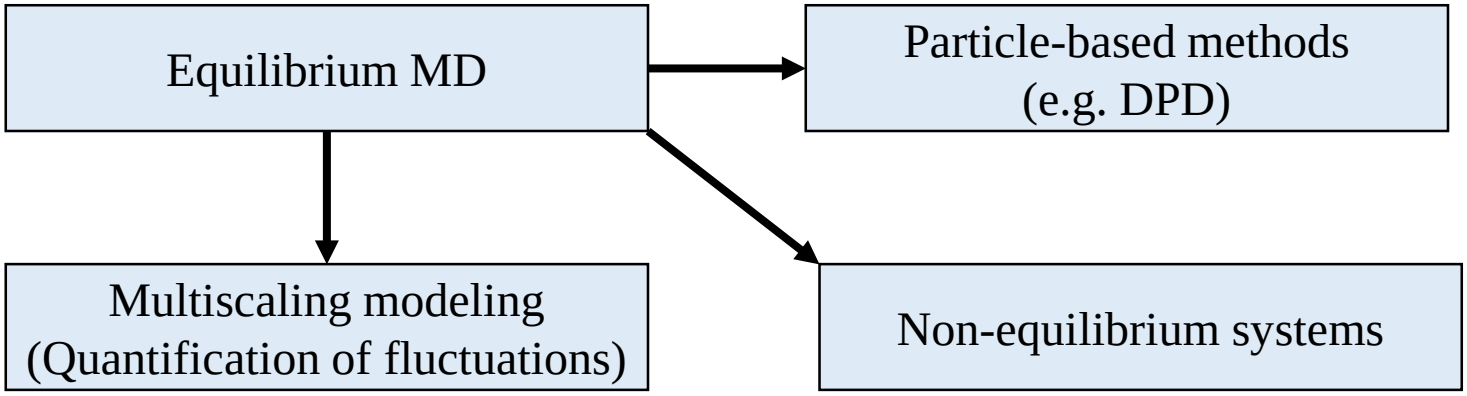
Underlying Idea



Simulating a realistically large system

Simulating an ***ensemble*** of a molecular system

It is time to analyze uncertainties of particle-based methods.

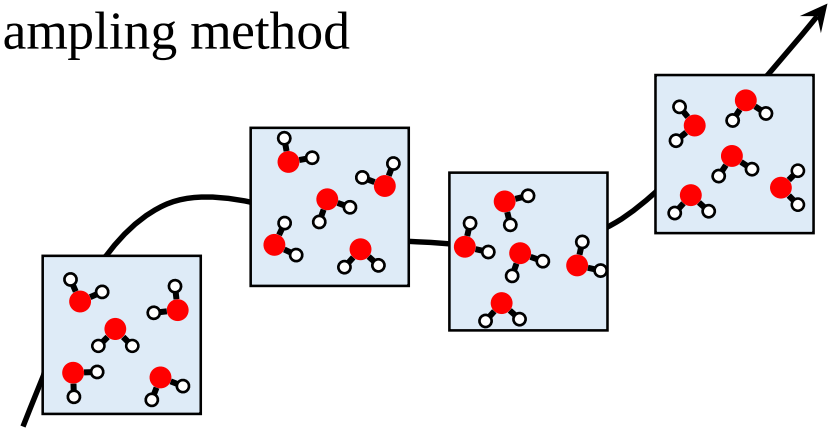


Acknowledgements

- US Army Research Laboratory
  - Alliance for Computationally-guided Design of Energy Efficient Electronic Materials (CDE3M)
  - Dr. Oleg Borodin
- US Department of Energy
  - CM4 center: Collaboratory on Mathematics for Mesoscopic Modeling of Materials (PI: Prof. Karniadakis)
  - INCITE project: computing time for clusters at the Argonne / Oak Ridge National Laboratories

Equilibrium MD Simulation

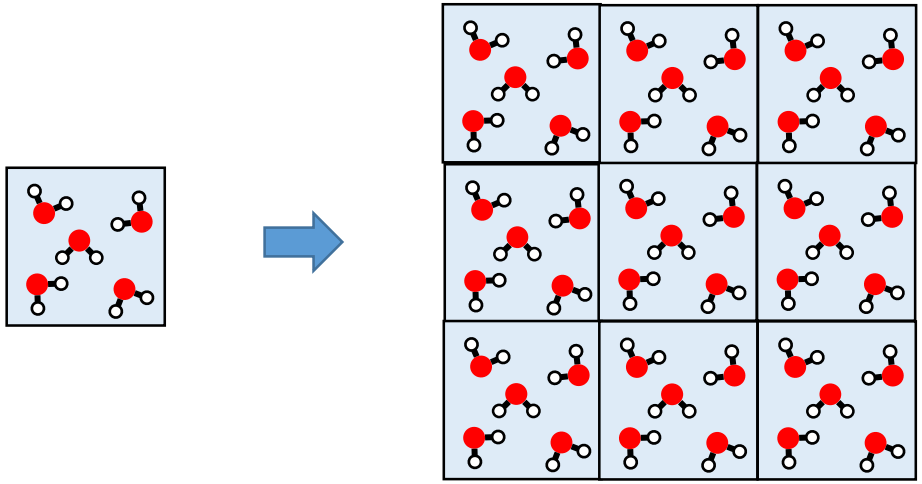
- Sampling method



$$\langle A \rangle \approx \frac{1}{N} \sum_{i=1}^{N} A_i$$

sampling errors
or statistical errors

- Periodic boundary conditions



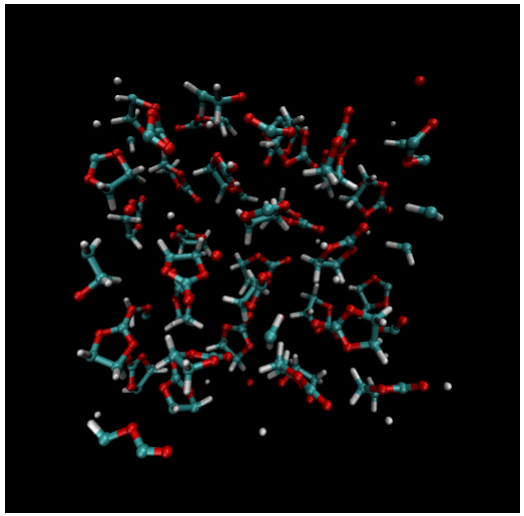
finite-system-size
effects

Uncertainty Quantification for MD

- Parameter uncertainty quantification

force field → material properties

- Two types of **intrinsic uncertainty in MD simulation results**
  - **Statistical errors**
  - **Finite-system-size effects**

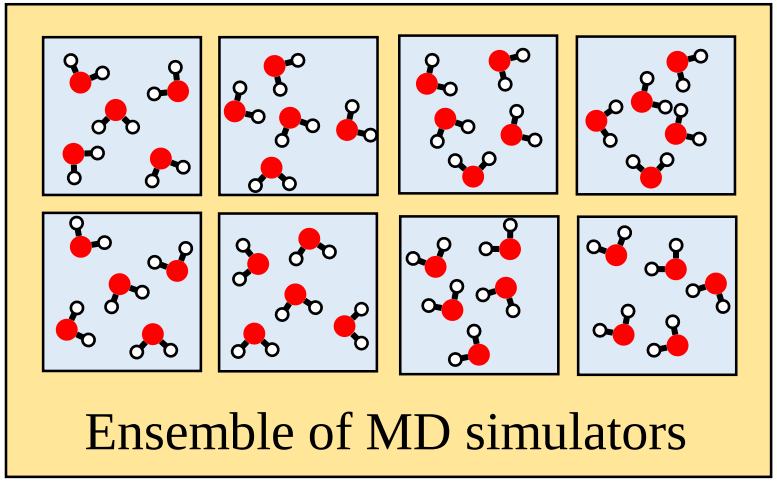


Statistical Errors → **Finite-system-size effects** → **Parameter UQ**

→ MD simulation of ethylene carbonate liquid

Approaches

- Large-sized ensemble MD runs
  - Accurate statistics
  - Direct evaluation of statistical errors



- Analysis (theoretical approaches)
  - Statistical mechanics
  - Probability theory
  - Continuum mechanics

Outline

Part 1. Statistical errors in the estimation of self-diffusion coefficients

- VACF method versus MSD method
- Part 1.1. Time-averaging and ensemble-averaging
- Part 1.2. Particle-averaging

Part 2. Finite-system-size corrections for self-diffusion coefficients

Part 3. Estimation of shear viscosity

Part 1.

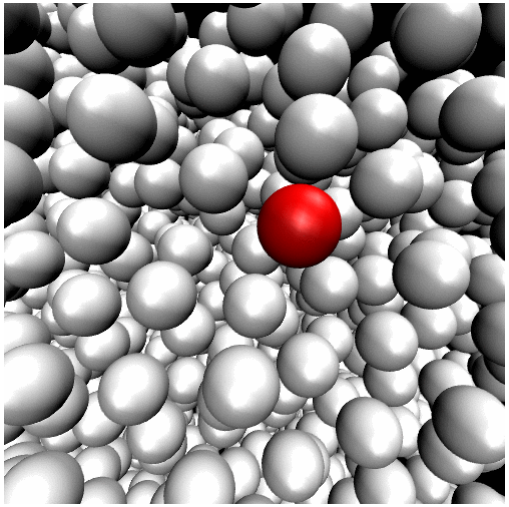
Statistical Errors in the Estimation of Self-Diffusion Coefficients

0. Background and motivating questions
1. Ensemble-averaging and time-averaging
2. Particle-averaging

Self-Diffusion Coefficient

- Definition through the mean-square-displacement (MSD)

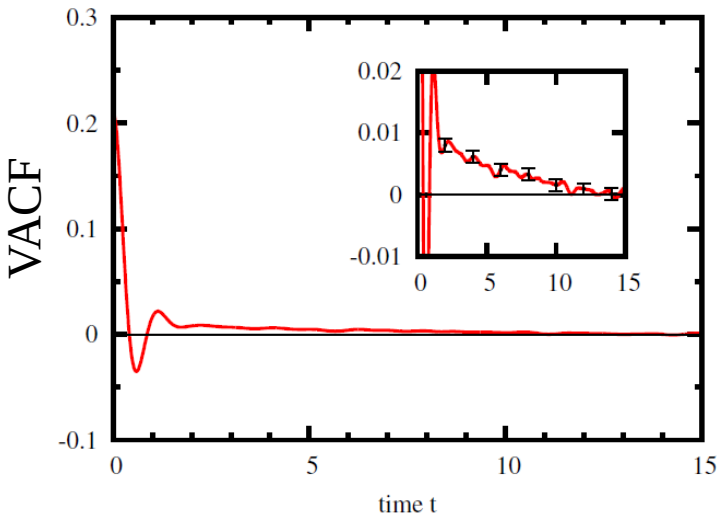
$$D = \lim_{t\to\infty} \frac{\langle [\mathbf{x}(t) - \mathbf{x}(0)]^2 \rangle}{6t}$$

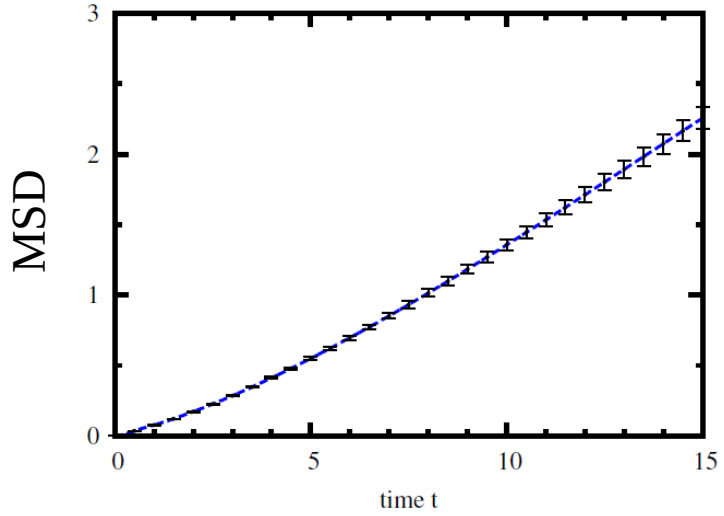


- Alternatively, through the velocity autocorrelation function (VACF)

$$D = \frac{1}{3}\int_0^\infty \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle dt$$

Statistical Errors in VACF, MSD, and $D(t)$





A colloidal particle suspended in a simple molecular fluid

VACF method

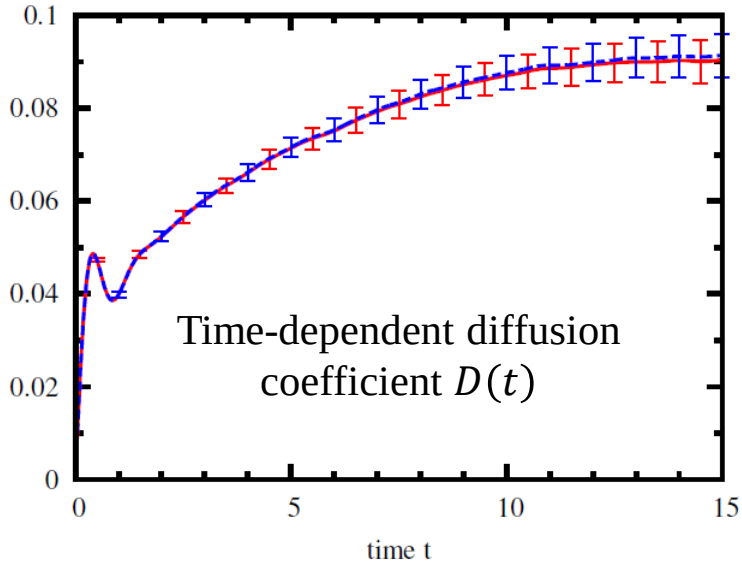
integration

$$D(t) = \int_0^t \langle v(0)v(t')\rangle dt'$$

MSD method

differentiation

$$D(t) = \frac{1}{2}\frac{d}{dt}\langle [x(t) - x(0)]^2 \rangle$$



1. Are the two methods equivalent?
2. Can we calculate the error bars of $D(t)$ from those of VACF or MSD?
3. Can we estimate the error bars from VACF under reasonable assumptions?

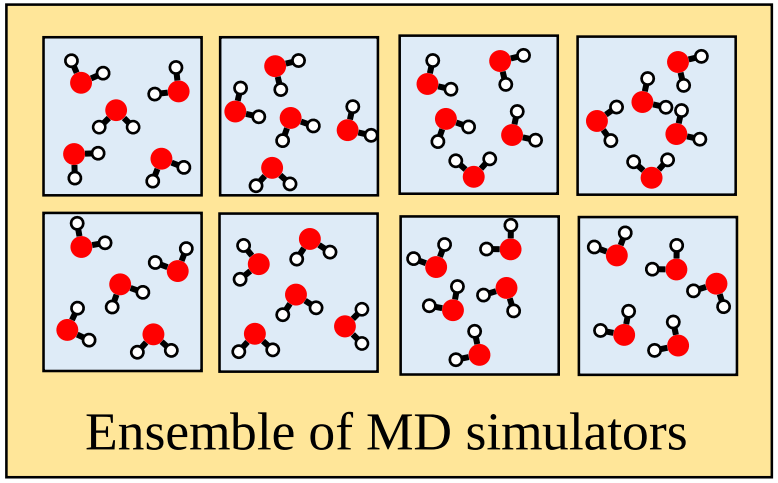
Three Types of Averaging Procedures

- Ensemble-averaging over (i.i.d.) samples
- Time-averaging over trajectory
- Particle-averaging over identical particles

Ensemble-Averaging

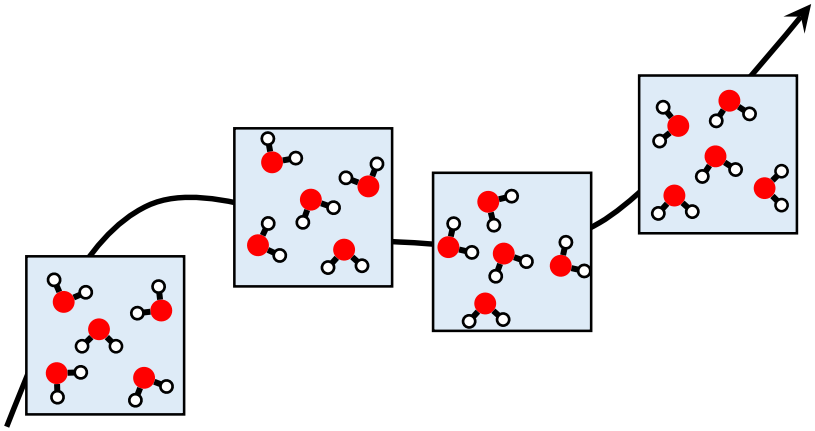
$$\langle v_1(0)v_1(t)\rangle = \int v_1 v_1(t)\,\rho(x_1,\cdots,x_N,v_1,\cdots,v_N)dx_1\cdots dx_N dv_1\cdots dv_N$$

$$\frac{1}{N_{sample}}\sum_{i=1}^{N_{sample}} v_1^{(i)}(0)v_1^{(i)}(t)$$

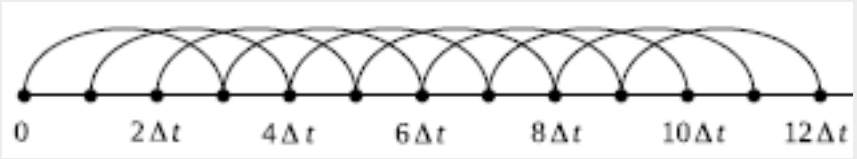


- Generating an equilibrium sample is usually time-consuming.

Time-Averaging



$$\frac{1}{N_t}\sum_{i=1}^{N_t} v_1(t_i)v_1(t_i+t)$$



← Calculation of $\langle v_1(0)v_1(3\Delta t)\rangle$

- Calculating $\langle v_1(0)v_1(n\Delta t)\rangle$ $(0 \le n \le n_{max})$ on the fly requires storing trajectory of length $n_{max}+1$.

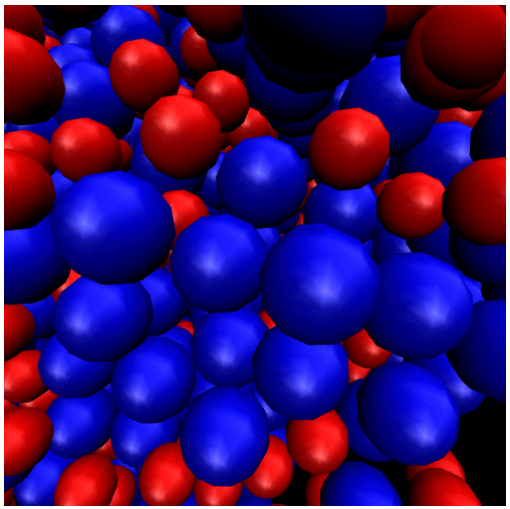
$$\langle v_1(0)v_1(t)\rangle = \lim_{\mathcal{T}\to\infty}\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} v_1(t')v_1(t'+t)dt'$$

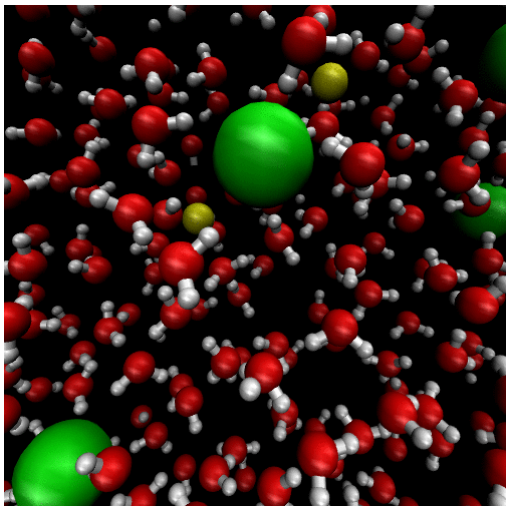
Ergodic hypothesis

Particle-Averaging

If there are $N_{particle}$ identical particles in the system, one can also take particle-averaging over n particles ($1 \leq n \leq N_{particle}$).

$$\langle v(0)v(t)\rangle = \frac{1}{n}\sum_{i=1}^{n}\langle v_i(0)v_i(t)\rangle$$





❖ Three types of averaging procedures can be combined.

Scaling Behavior of Statistical Error

$$\langle \varepsilon^2(t) \rangle \approx \frac{a(t)}{\mathcal{N}\mathcal{T}n^*}$$

- $\mathcal{N}$ = number of independent trajectories (ensemble-averaging)
- $\mathcal{T}$ = length of a trajectory (time-averaging)
- n^* = effective number for particle averaging
  - For sufficiently small n, $n^* \approx n$
  - Otherwise, $n^* \ll n$

Part 1.1.
Ensemble Averaging and Time Averaging

- Theoretical error estimates
- MD simulation results: LJ fluid / EC liquid
- Further analysis: Langevin equation driven by GWN/PWSN

Kim, Borodin, and Karniadakis, “Quantification of Sampling Uncertainty for Molecular Dynamics Simulation: Time-dependent Diffusion Coefficient in Simple Fluids”, in press, *J. Comput. Phys*. (http://dx.doi.org/10.1016/j.jcp.2015.09.021)

Main Results

For both VACF and MSD methods, the standard errors of $D(t)$ are the same.

- For ensemble-averaging,

$$\langle \varepsilon^2(t) \rangle = \frac{1}{\mathcal{N}} \int_0^t dt' \int_0^t dt'' \left[f(0) f(t'' - t') + f(t') f(t'') \right]$$

- For time-averaging,

$$\langle \varepsilon^2(t) \rangle = \frac{1}{\mathcal{T}} \int_{-\infty}^{\infty} d\alpha \left[f(\alpha) \int_0^t dt' \int_{\alpha}^{\alpha+t} dt'' \, f(t' - t'') + \int_{\alpha}^{\alpha+t} f(t') dt' \int_{\alpha-t}^{\alpha} f(t') dt' \right]$$

where $f(t)$ is the VACF.

- Once the VACF has been (roughly) estimated, the standard errors of $D(t)$ as well as the VACF and the MSD are available.
- These results are obtained under the assumption that the velocity process $v(t)$ is a Gaussian process (GPA = Gaussian process approximation).

Derivation, Step 1: Error Correlation Functions

Error correlation function of VACF $\langle \varepsilon_{\mathrm{VACF}}(t')\varepsilon_{\mathrm{VACF}}(t'')\rangle$

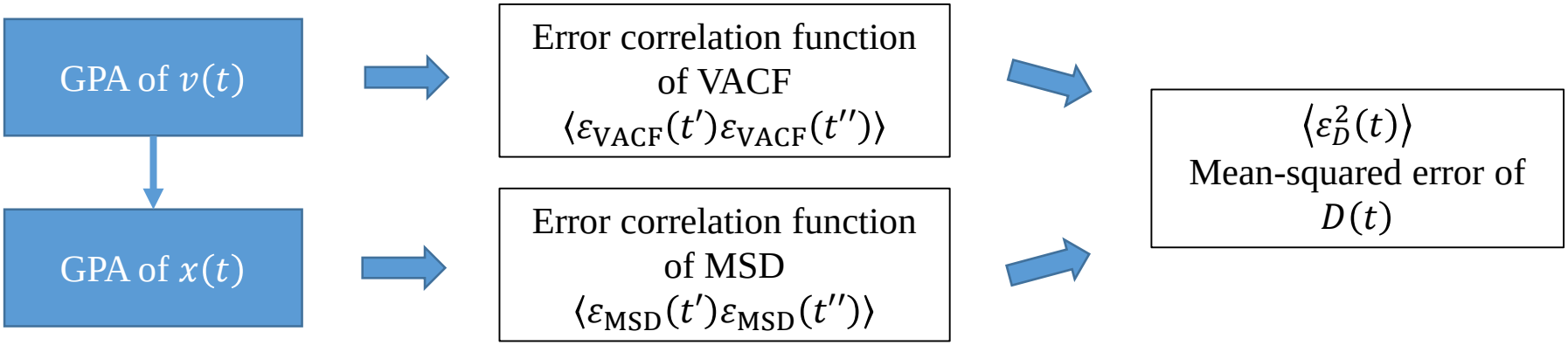
Error correlation function of MSD $\langle \varepsilon_{\mathrm{MSD}}(t')\varepsilon_{\mathrm{MSD}}(t'')\rangle$

$\langle \varepsilon_D^2(t)\rangle$ Mean-squared error of $D(t)$

$$D(t) = \int_0^t \langle v(0)v(t')\rangle dt' \longrightarrow \langle \varepsilon_D^2(t)\rangle = \int_0^t dt' \int_0^t dt'' \langle \varepsilon_{VACF}(t')\varepsilon_{VACF}(t'')\rangle$$

$$D(t) = \frac{1}{2}\frac{d}{dt}\langle [x(t) - x(0)]^2\rangle \longrightarrow \langle \varepsilon_D^2(t)\rangle = \ldots$$

Derivation, Step 2: GPA

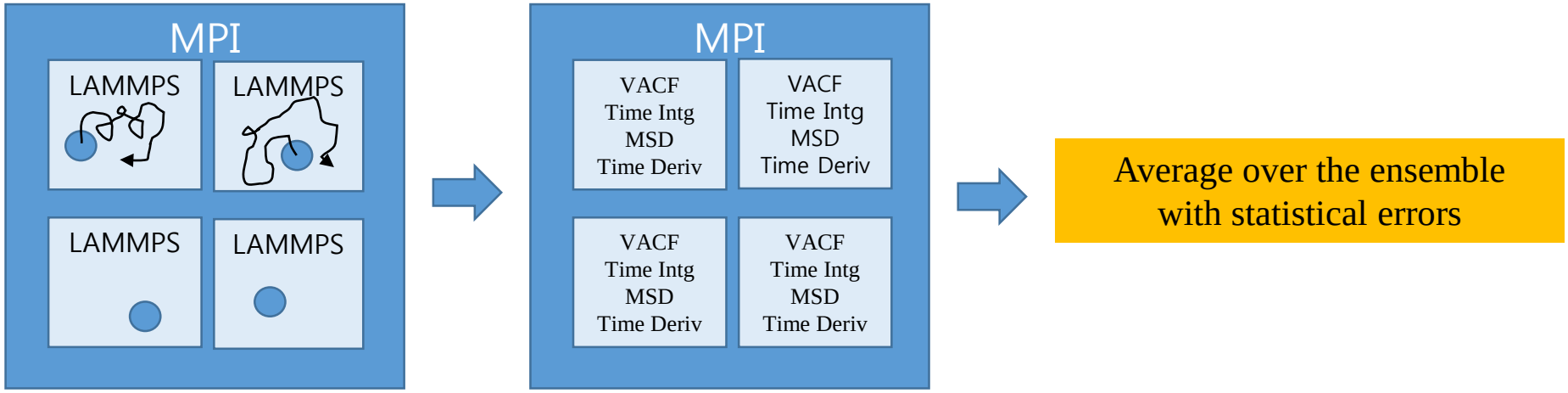


- Error correlation function $\langle \varepsilon_{\mathrm{VACF}}(t')\varepsilon_{\mathrm{VACF}}(t'')\rangle$ is expressed in terms of four time correlation function $\langle v(0)v(t_1)v(t_2)v(t_3)\rangle$.
- Under the GPA, $\langle \boldsymbol{v(0)v(t_1)v(t_2)v(t_3)}\rangle$ is decomposed into

$$\langle v(0)v(t_1)\rangle\langle v(t_2)v(t_3)\rangle + \langle v(0)v(t_2)\rangle\langle v(t_1)v(t_3)\rangle + \langle v(0)v(t_3)\rangle\langle v(t_2)v(t_4)\rangle.$$

- The GPA of $x(t)$ is implied by that of $v(t)$.

Large-Sized-Ensemble MD Run: LMP_ENS

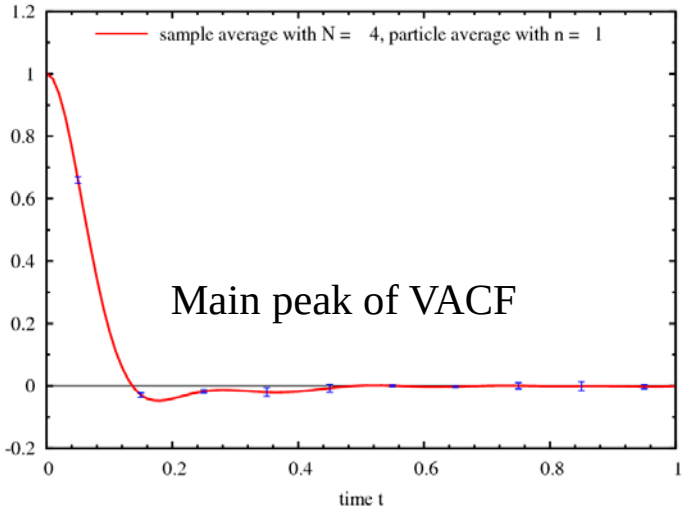


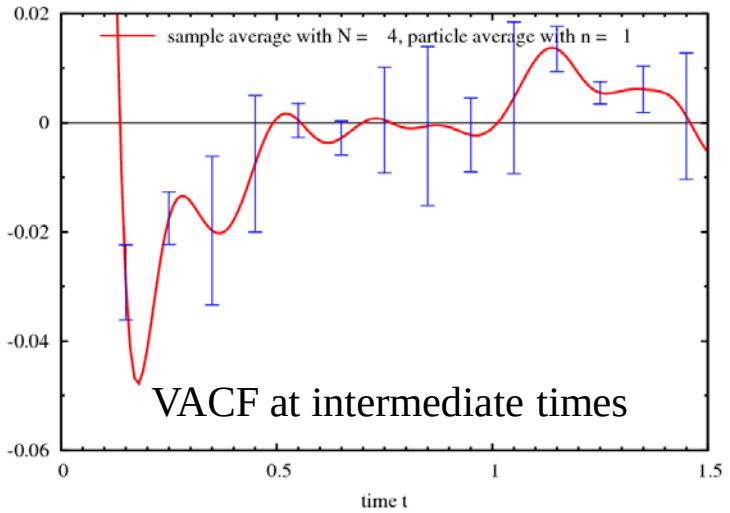
Typical size of ensemble ~10^4 samples

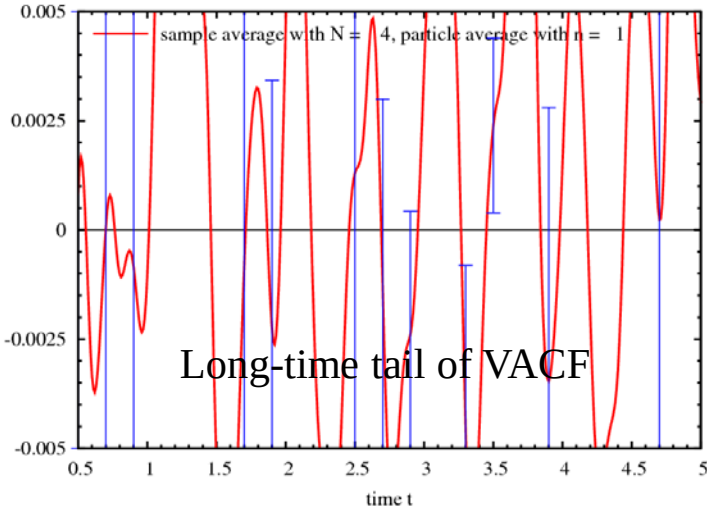
- Statistical errors are suppressed by factor of 100.

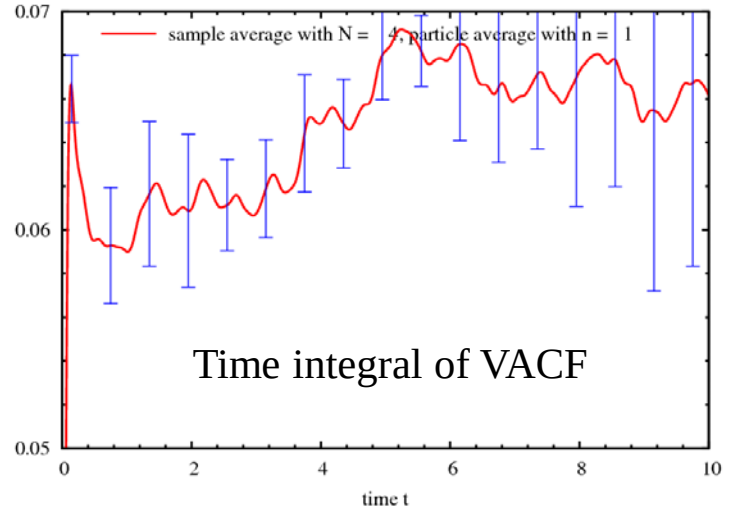
Computing time provided by INCITE project: BG/Q machine at Argonne Lab

Sampling Size and Quality of MD Data



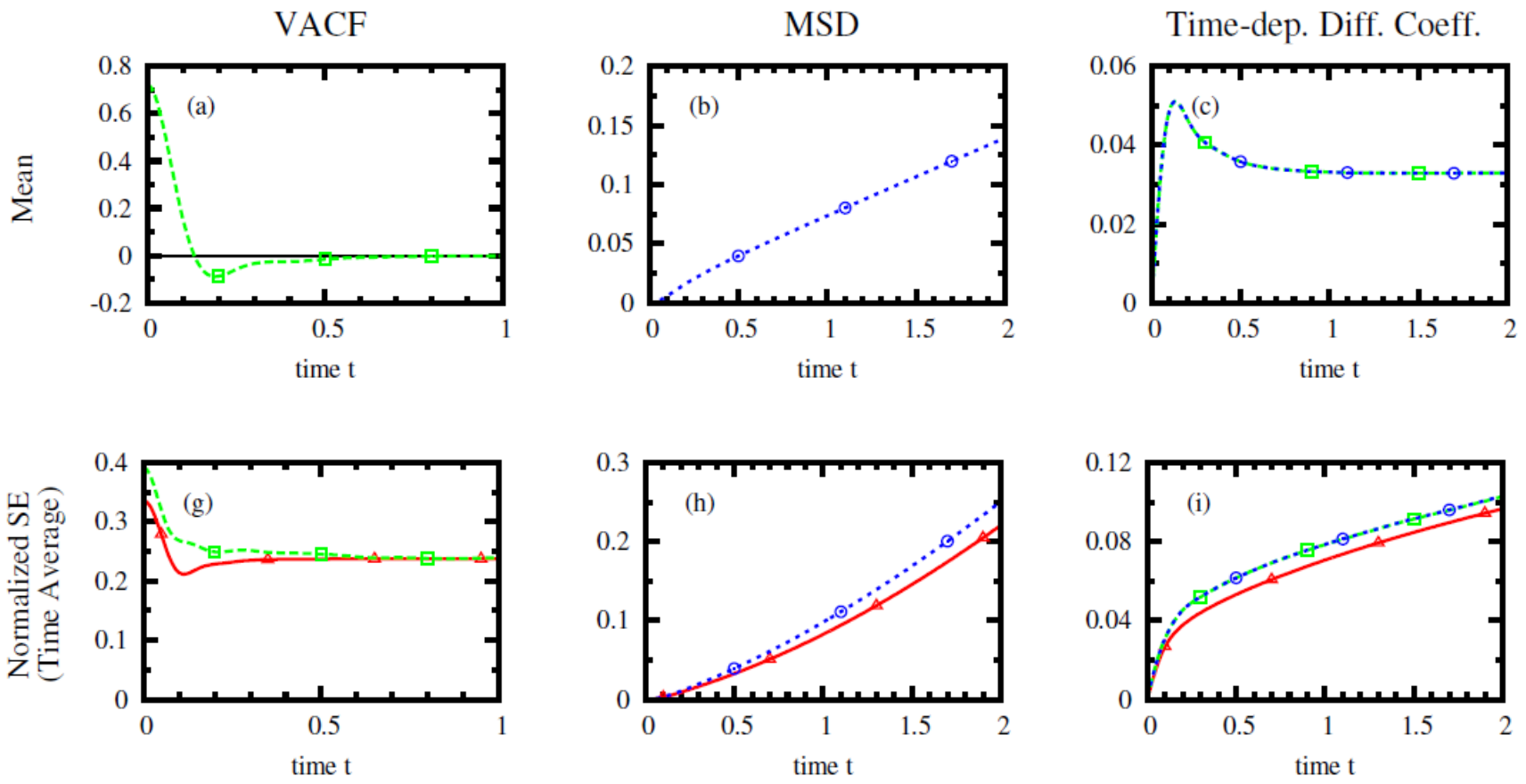




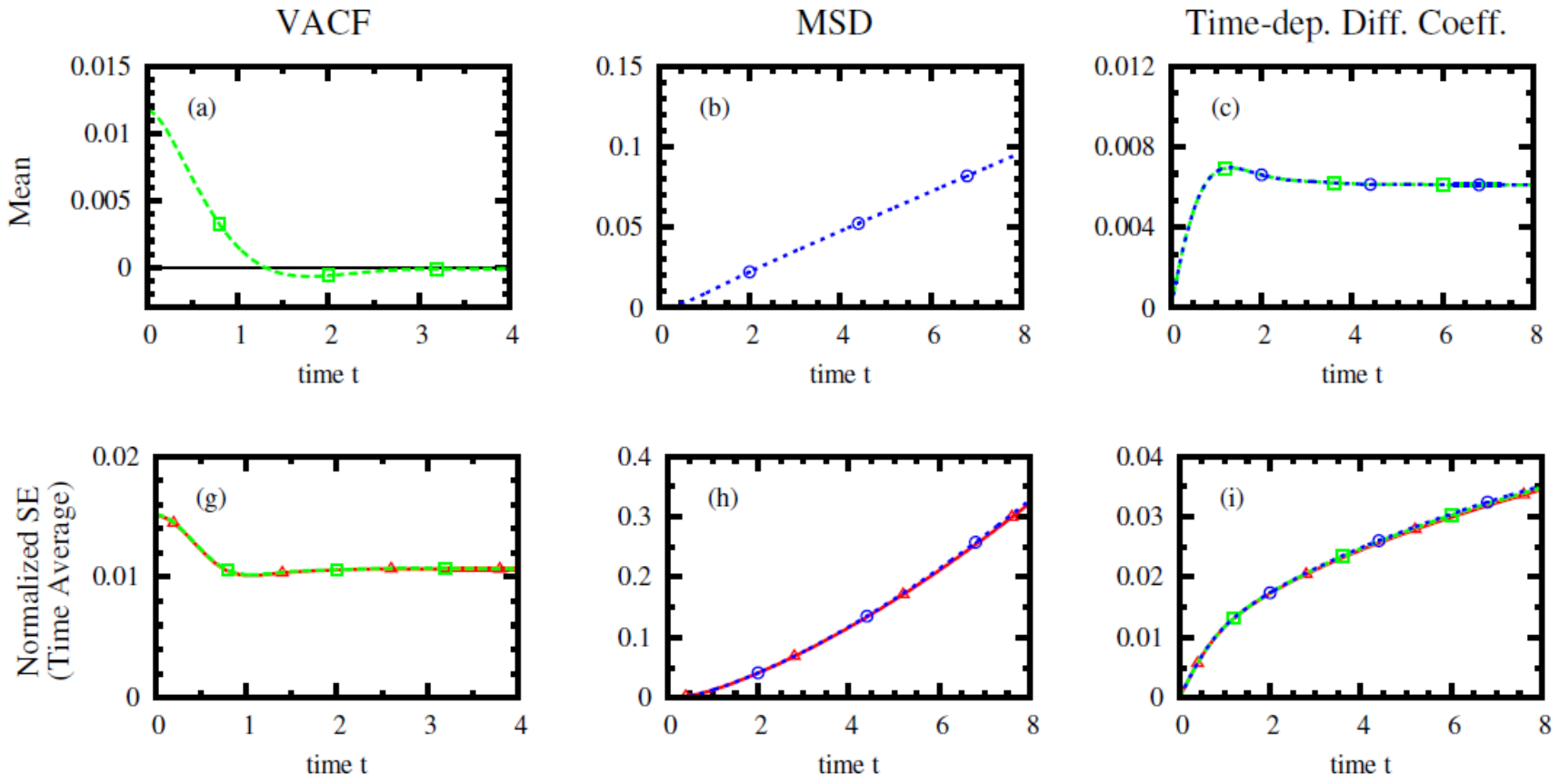


Sample size increases up to by factor of 2^{18}
Standard error decreases up to by factor of $2^9 = 512$

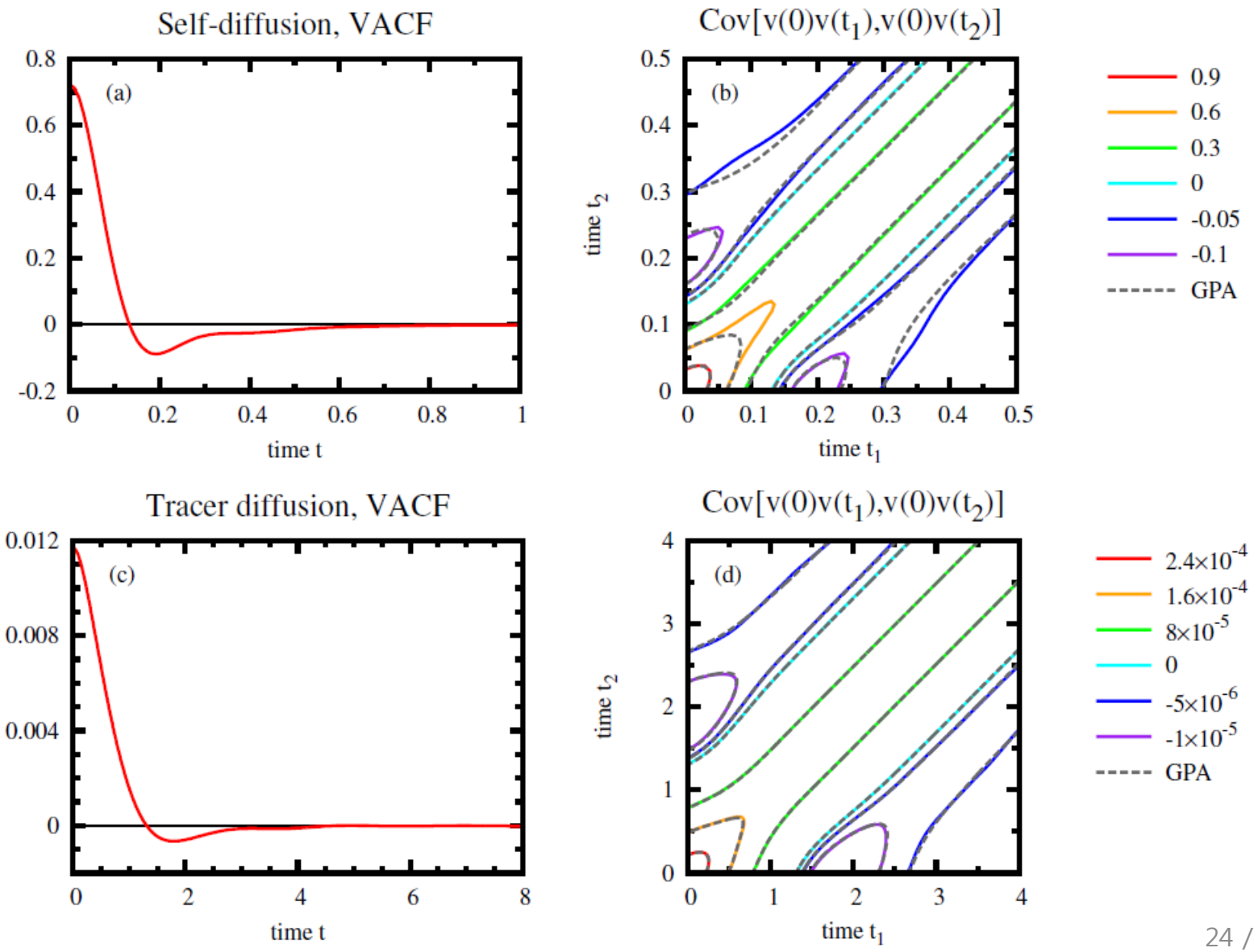
Self-Diffusion of a Solvent Particle



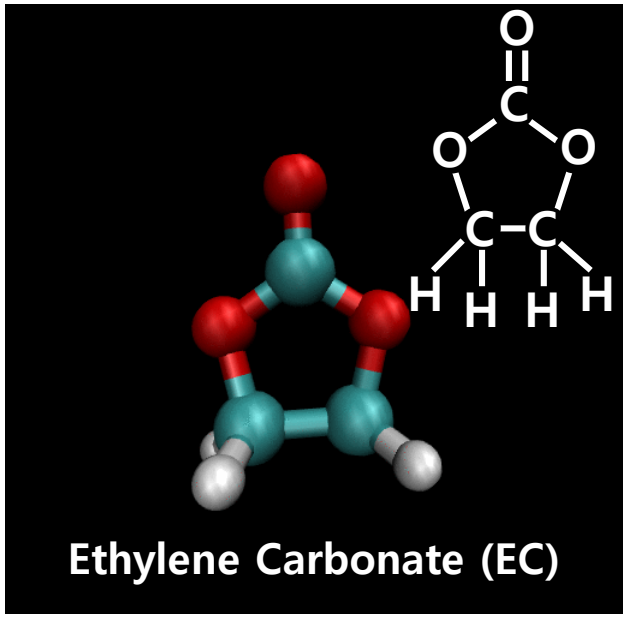
Tracer-diffusion of a Colloidal Particle

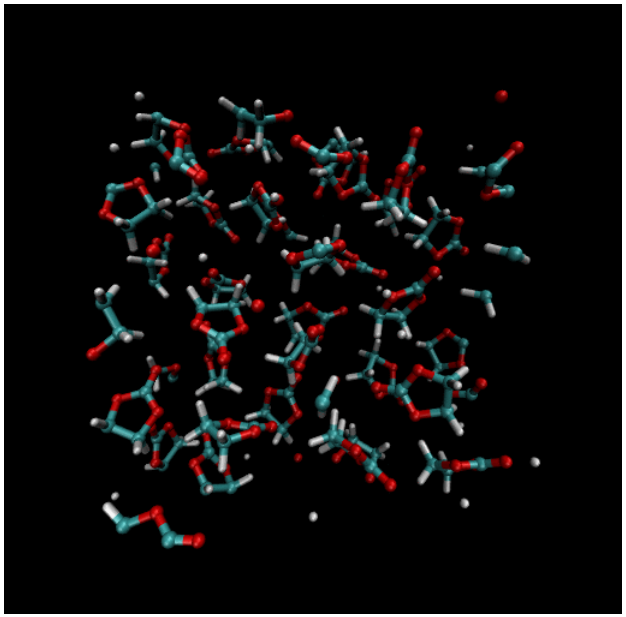


The Validity of GPA



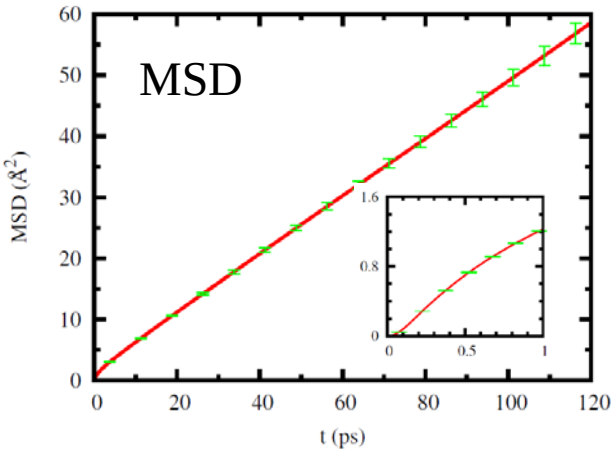
MD Simulation of EC Liquid

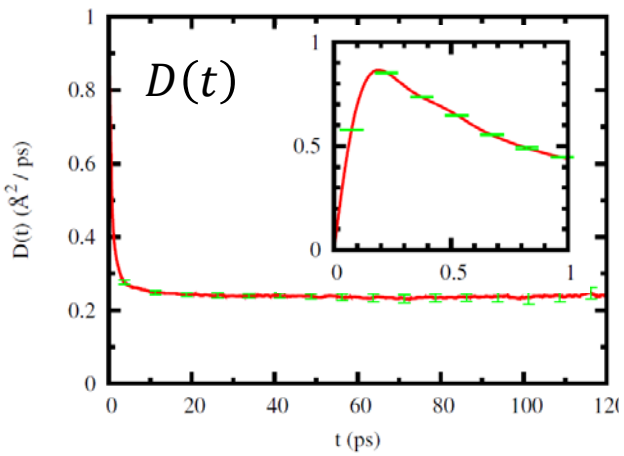


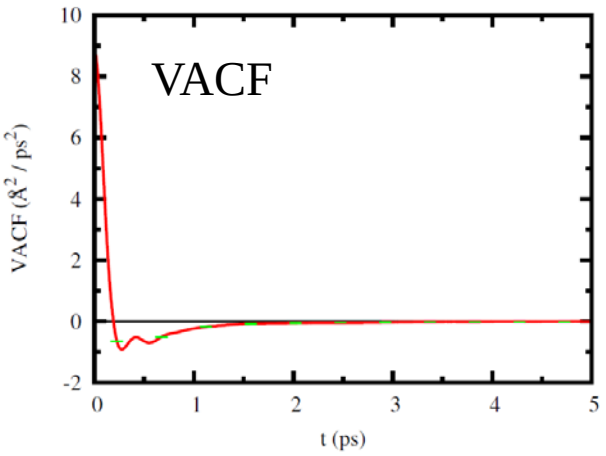


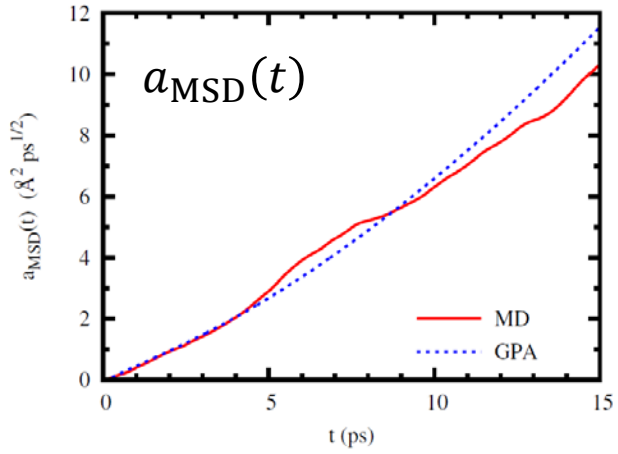
- Quantum-chemistry based, highly transferable atomistic force field
  - APPLE&P (Atomistic Polarizable Potential for Liquids, Electrolytes, and Polymers) developed by Oleg Borodin

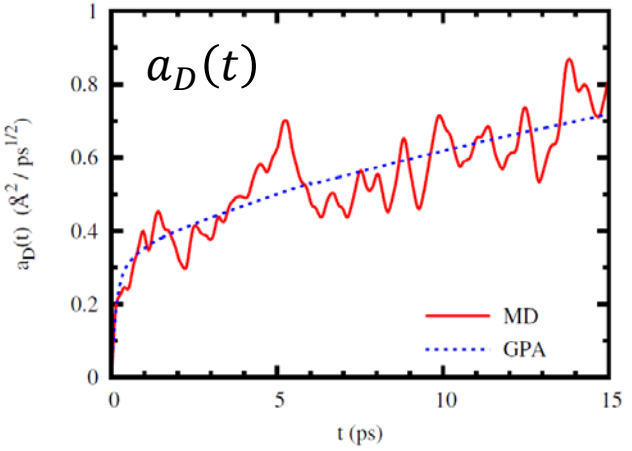
MD Results for EC System

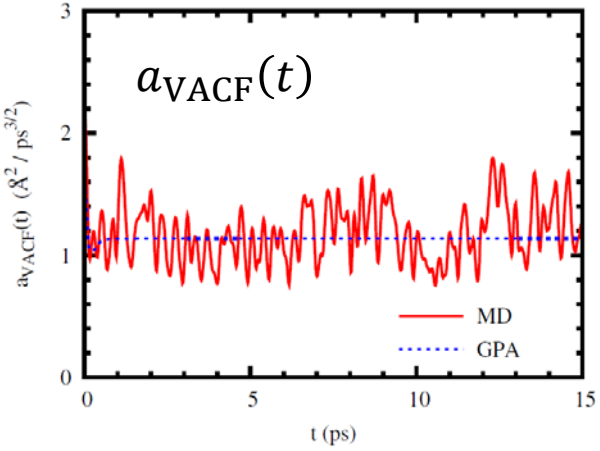












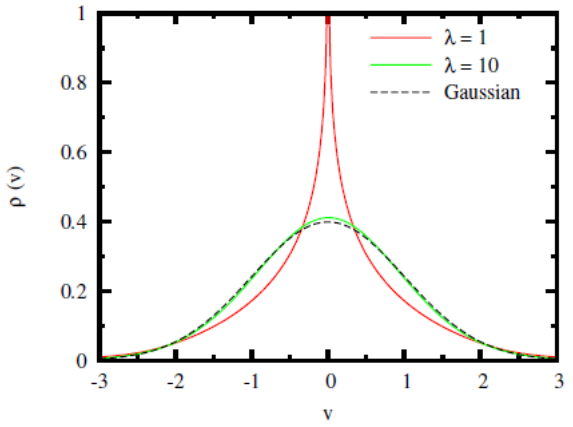
Non-Gaussianity Indicator

Langevin equation

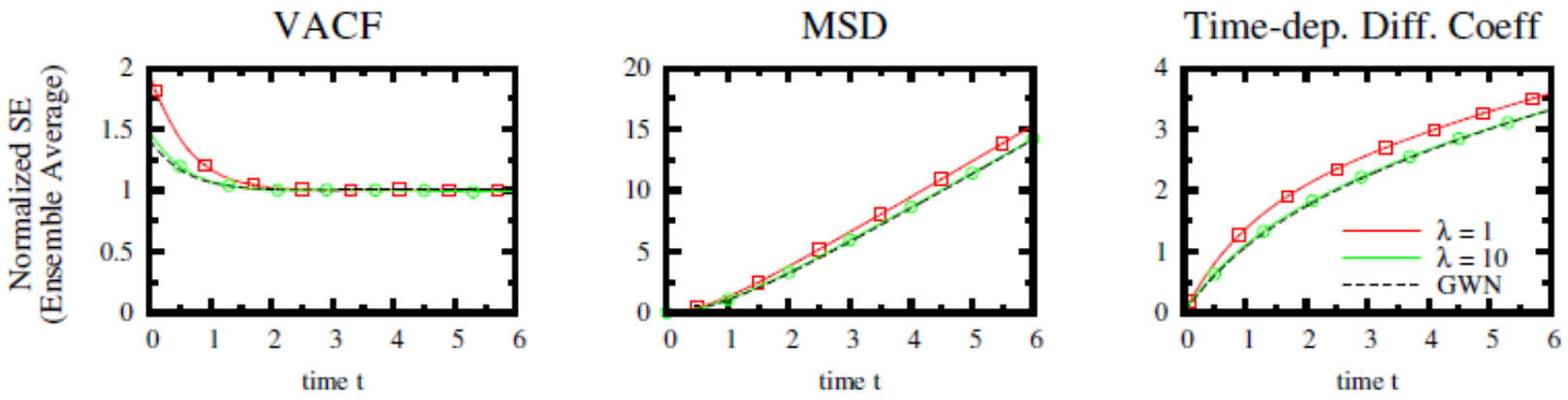
$$\dot{x}(t) = v(t)$$

$$\dot{v}(t) = -\gamma v(t) + \xi(t)$$

$$\langle \xi(t')\xi(t'') \rangle = 2k_B T\gamma\delta(t' - t'')$$



$\xi(t)$ is either Gaussian white noise (GWN) or Poissonian white shot noise (PWSN).



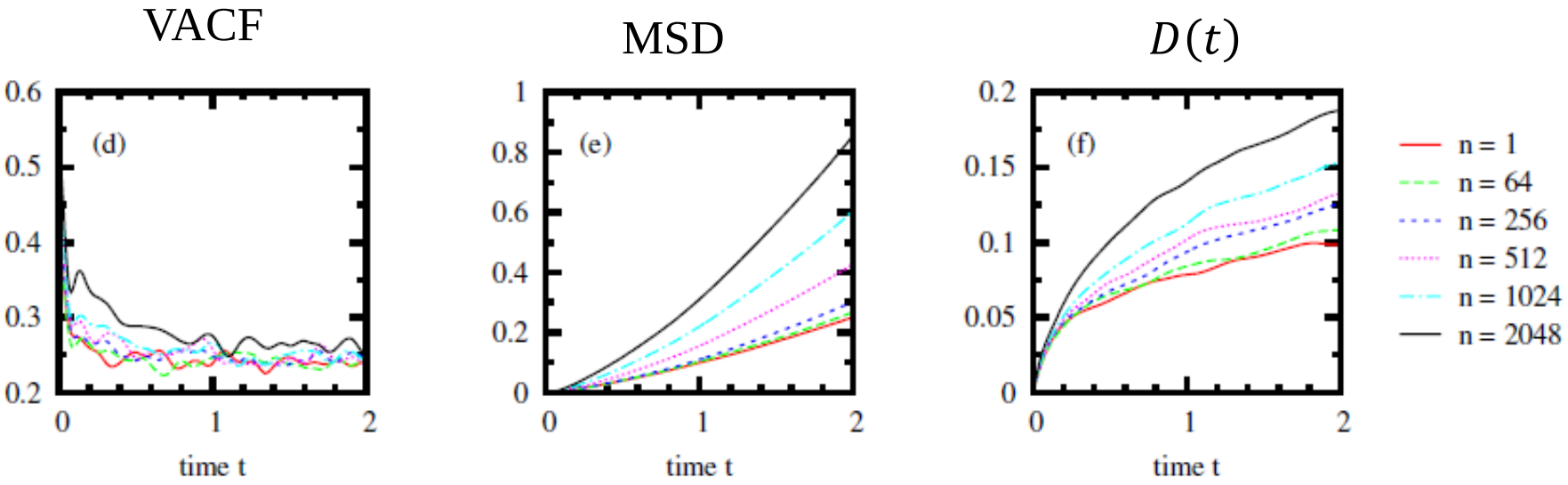
Deviations from theoretical error estimates are proportional to the fourth-order cumulant $\kappa[\xi(t_1), \xi(t_2), \xi(t_3), \xi(t_4)]$.

Part 1.2.

Particle-Averaging

Observation: LJ Fluid ($N_{\mathrm{atom}} = 2048$)

Normalized standard errors $n\mathcal{N}\mathcal{T}\,\langle\varepsilon^2(t)\rangle$



$$\langle\varepsilon^2(t)\rangle \approx \frac{a(t)}{\mathcal{N}\mathcal{T}n^*}$$

For sufficiently small n, $n^* \approx n$
Otherwise, $n^* \ll n$

- GPA of $\left(v_1(t), v_2(t), \cdots, v_N(t)\right)$ → Almost $1/n$ scaling
  - Failure of multi-particle GPA

Formulation

- Quantity A_i obtained from particle i

$$\text{Particle average (of size } n\text{): } X_n = \frac{1}{n}\sum_{i=1}^{n} A_i$$

- Since the particles are identical, we have

$$\langle A_i \rangle = \mu$$

$$Var[A_i] = \sigma^2$$

$$Cov[A_i, A_j] = \zeta \quad (i \neq j)$$

❖ In fact, $\{A_i\}$ are exchangeable random variables.

Results

$$\mathrm{Var}[X_n] = \frac{\sigma^2}{n} + \frac{n-1}{n}\zeta = \frac{\sigma^2}{n^*}$$

- If $(n-1)\zeta \ll \sigma^2$, $\mathrm{Var}[X_n] \approx \frac{\sigma^2}{n}$

$$n^* = \frac{n}{1 + (n-1)\frac{\zeta}{\sigma^2}}$$

- $\zeta \to 0$ (uncorrelated case): $n^* \to n$
- $\zeta \to \sigma^2$ (completely correlated case):$n^* \to 1$

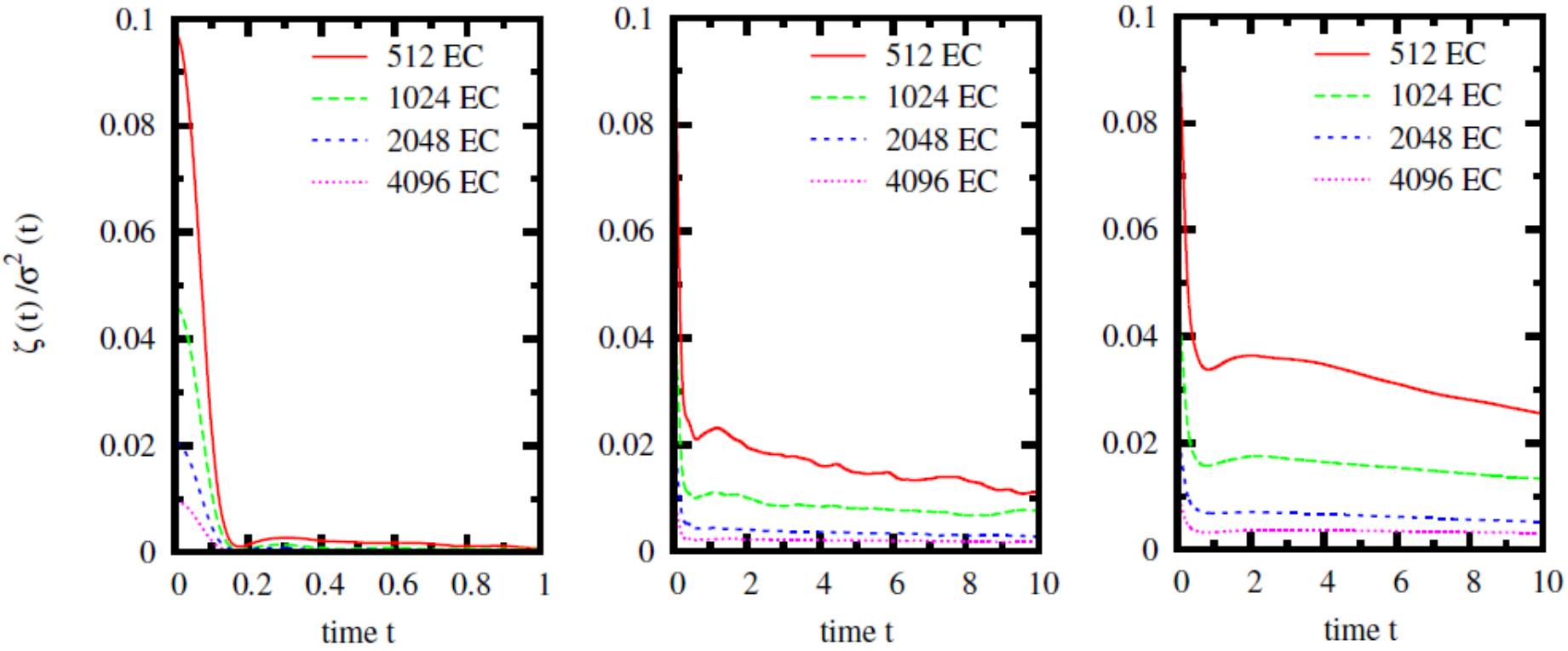
For the investigation of the reduction of statistical errors due to particle averaging, we need to investigate the corvariance ζ (or the correlation coefficient ζ/σ^2).

Correlation Coefficient $\zeta(t)/\sigma^2(t)$

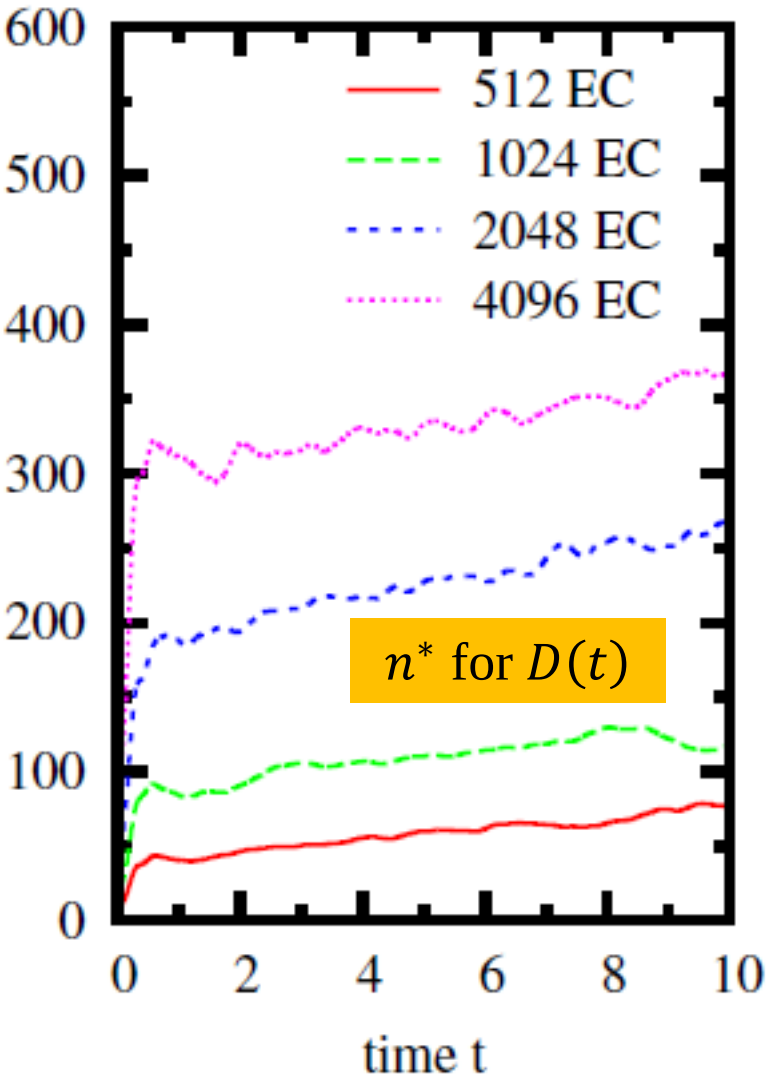
VACF

Time-dependent diffusion coefficient

MSD



Effective Number of Particle Averaging n^*



- Dependences on various averaging parameters are under investigation.

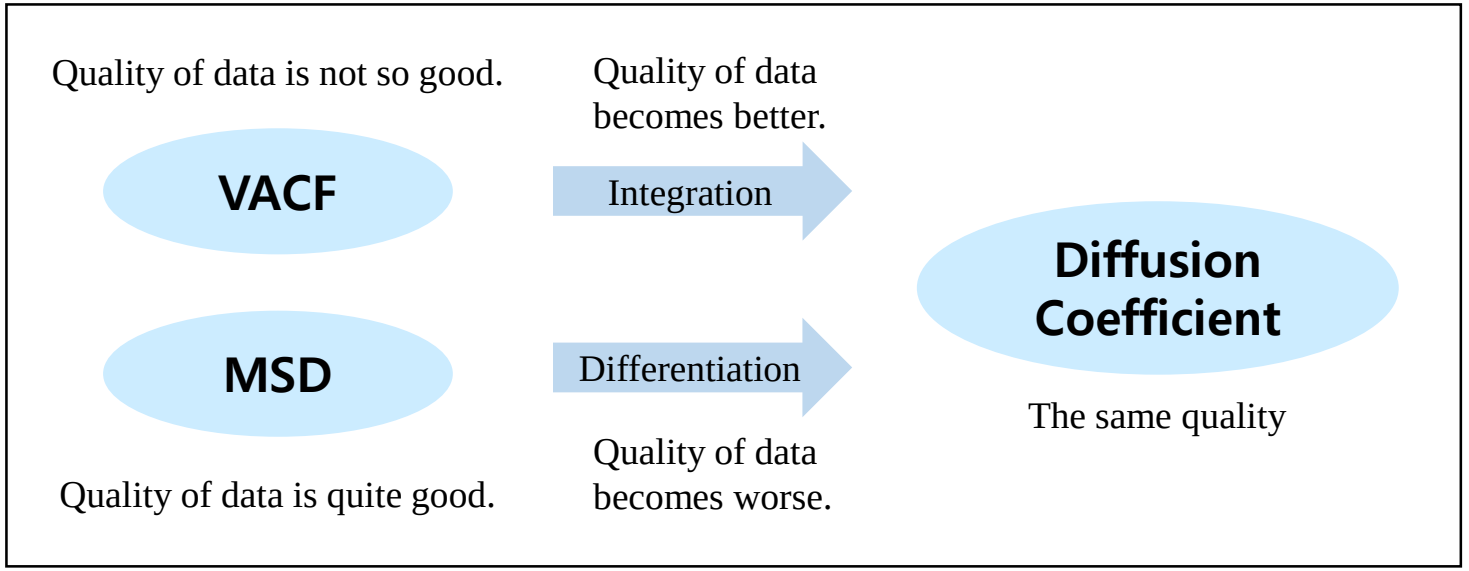
Practical implications

Full particle-averaging may be very expensive (especially, for on-the-fly calculation).

- If trajectory computation is expensive, use full particle-averaging.
- Otherwise, reduce the size of particle-averaging and calculate a longer trajectory.

Summary of Part 1

1. Equivalence of the VACF and MSD methods



2. Scaling behavior of statistical error

$$\langle \varepsilon^2(t) \rangle \approx \frac{a(t)}{n^* \mathcal{N} \mathcal{T}}$$

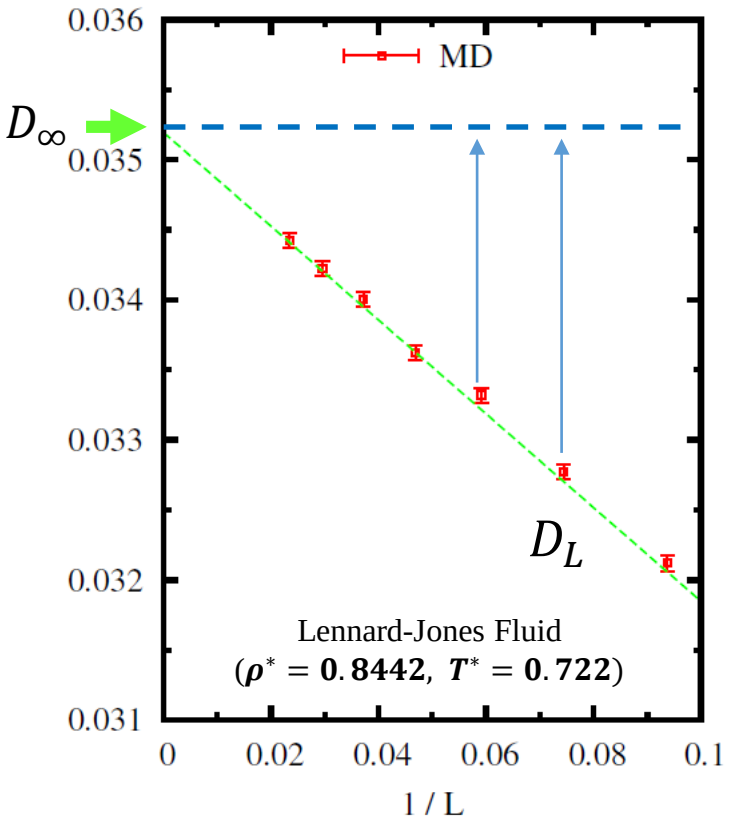
- Under the GPA, $a(t)$ can be expressed in terms of ordinary (i.e., two-time) correlation functions.
- The GPA works very well for various systems.
- $n^* \ll n$ for full particle-averaging

Part 2.

Finite-System-Size Correction on Diffusion Coefficient

1. Correction formula
2. Microscopic interpretation
3. Multi-species systems

Finite-System-Size Correction



$$D_\infty = D_L + \frac{2.837 k_B T}{6\pi\eta L}$$

- Under periodic boundary conditions
- L = simulation box size
- η = shear viscosity

❖ Derived from continuum theory

Hasimoto, *J. Fluid. Mech*. **5**, 317 (1959).

Dünweg and Kremer, *J. Chem. Phys*. **99**, 6983 (1993).

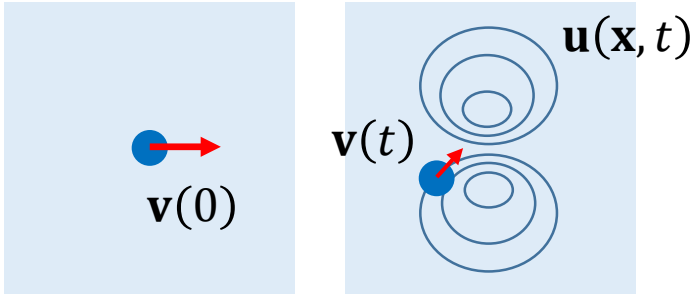
Yeh and Hummer, *J. Phys. Chem*. B **108**, 15873 (2004).

Connection to the Tail of VACF

$$D = \frac{1}{3}\int_0^\infty \langle \mathbf{v}(0)\cdot\mathbf{v}(t)\rangle dt$$

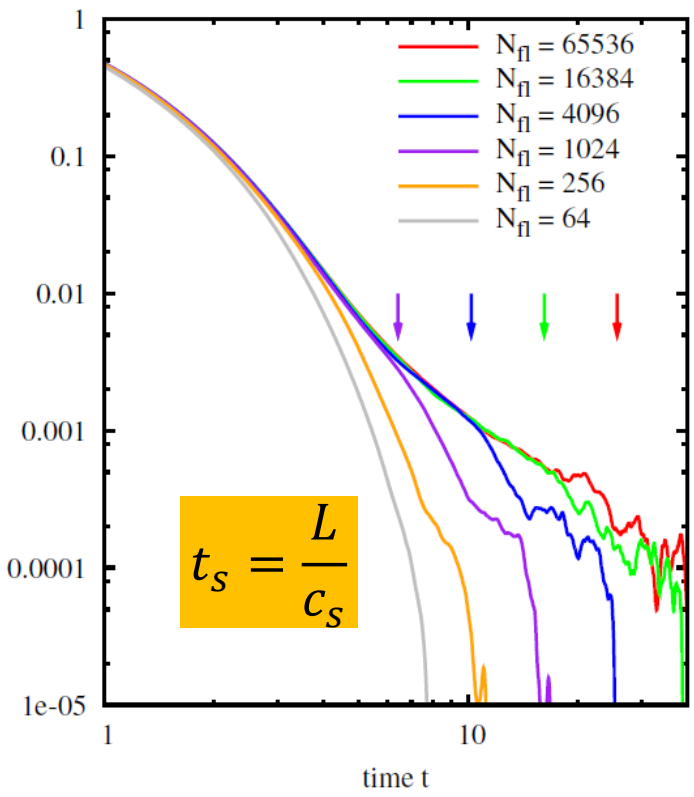
Algebraic tail of the VACF (also derived from continuum theory)

$$\langle \mathbf{v}(0)\cdot\mathbf{v}(t)\rangle = \frac{2k_BT}{3m\rho}\frac{1}{[4\pi(D+\nu)]^{3/2}}t^{-3/2} \quad (t\to\infty)$$

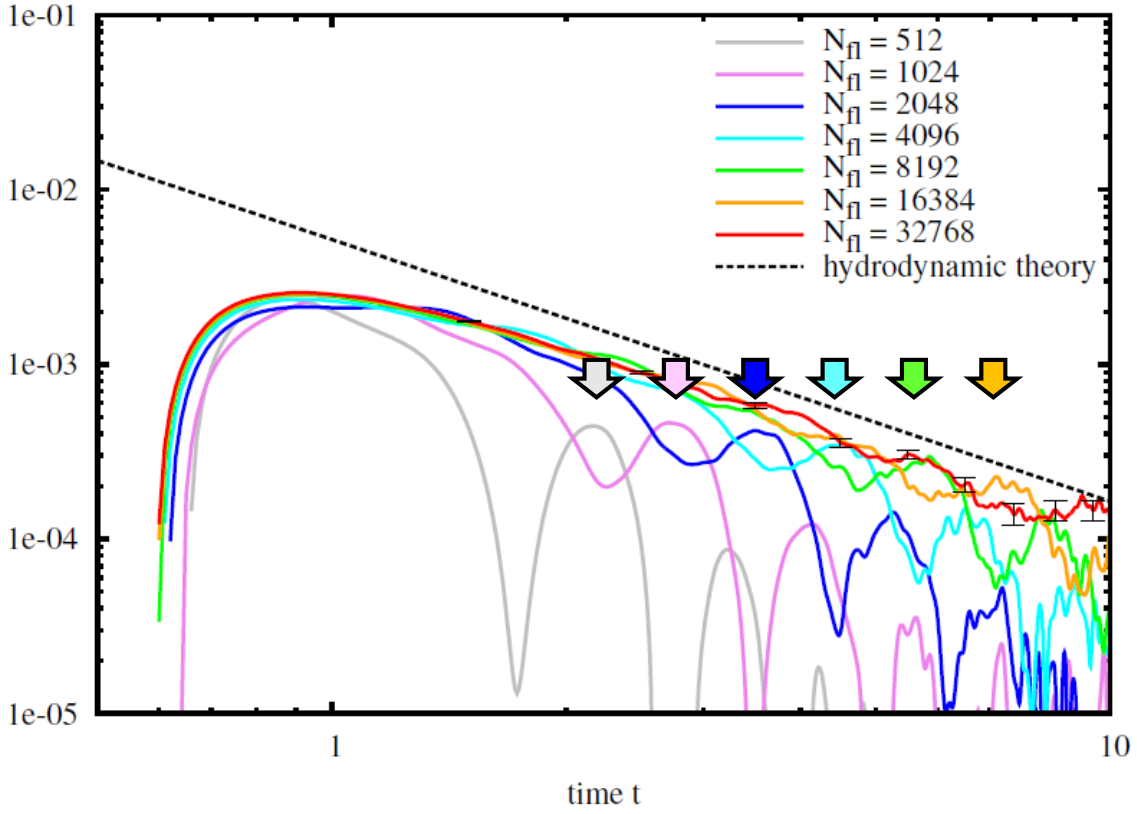


Long-time Tail of VACF

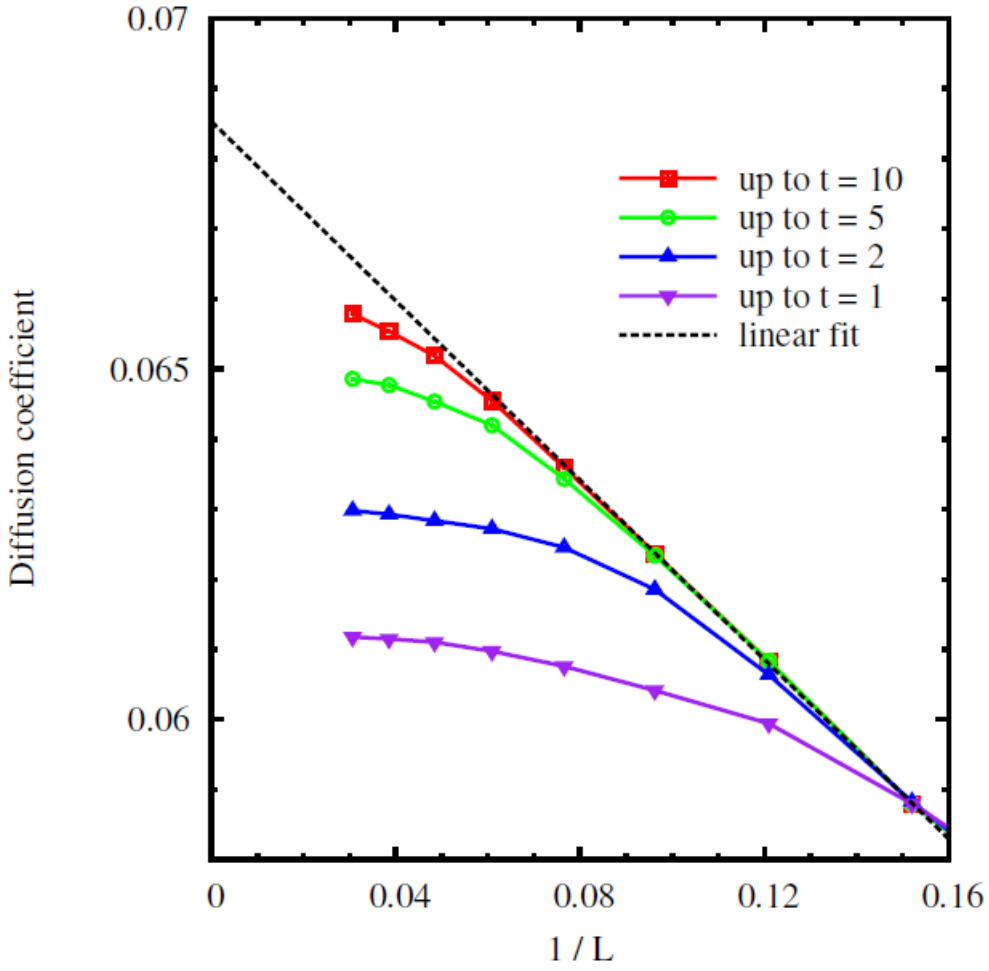
Lower density



Higher density



For a larger system, the diffusion coefficient should be estimated at a larger time.

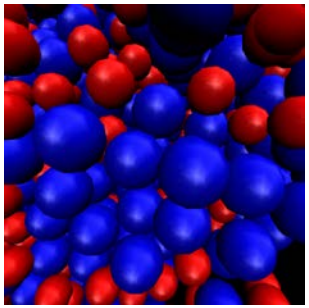


Corrections in Multi-Species Systems

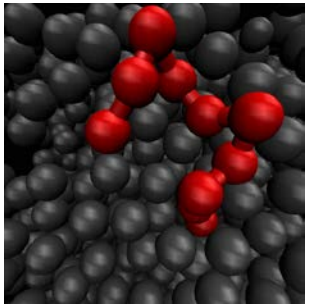
$$D_\infty = D_L + \left(\frac{2.837\, k_B T}{6\pi\eta}\right)\frac{1}{L}$$

Question

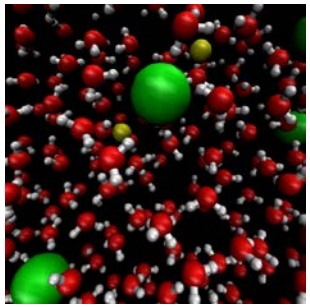
- Will **the same finite-system-size correction** be applied **to the self-diffusion coefficient of each species**?
- In other words, will the value of the slope (with respect to L^{-1}) be the same?



Ar/Kr mixture

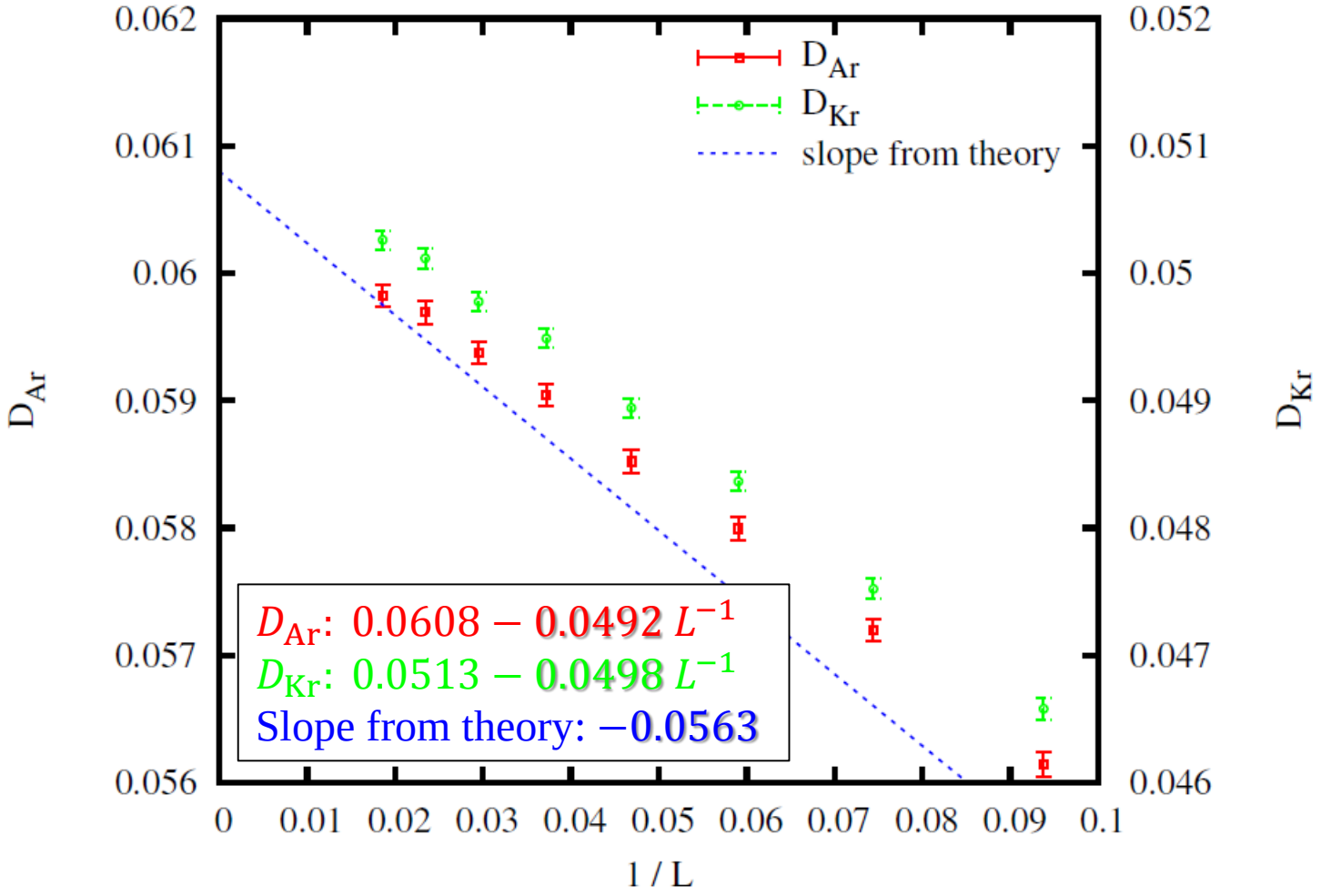


FENE chains
in a WCA fluid



H_2O + NaCl

Equimolar Ar/Kr Mixture
$(\rho^* = 0.7138,\ T^* = 0.965)$



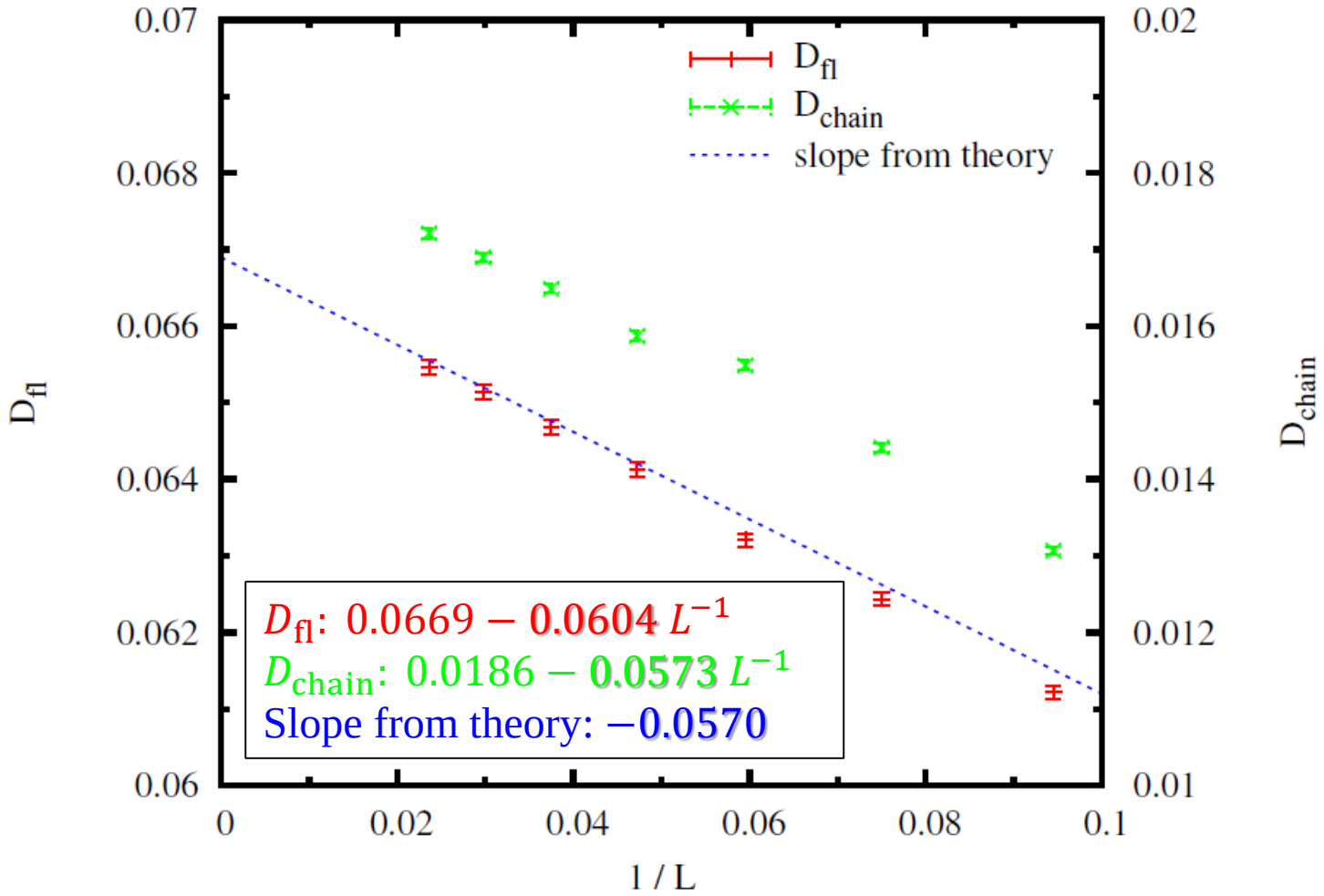
Although the slopes for the finite-system-size effects on D_{Ar} and D_{Kr} are different from the theoretical prediction, the values of the two slopes are very similar.

FENE Chains in a WCA Fluid

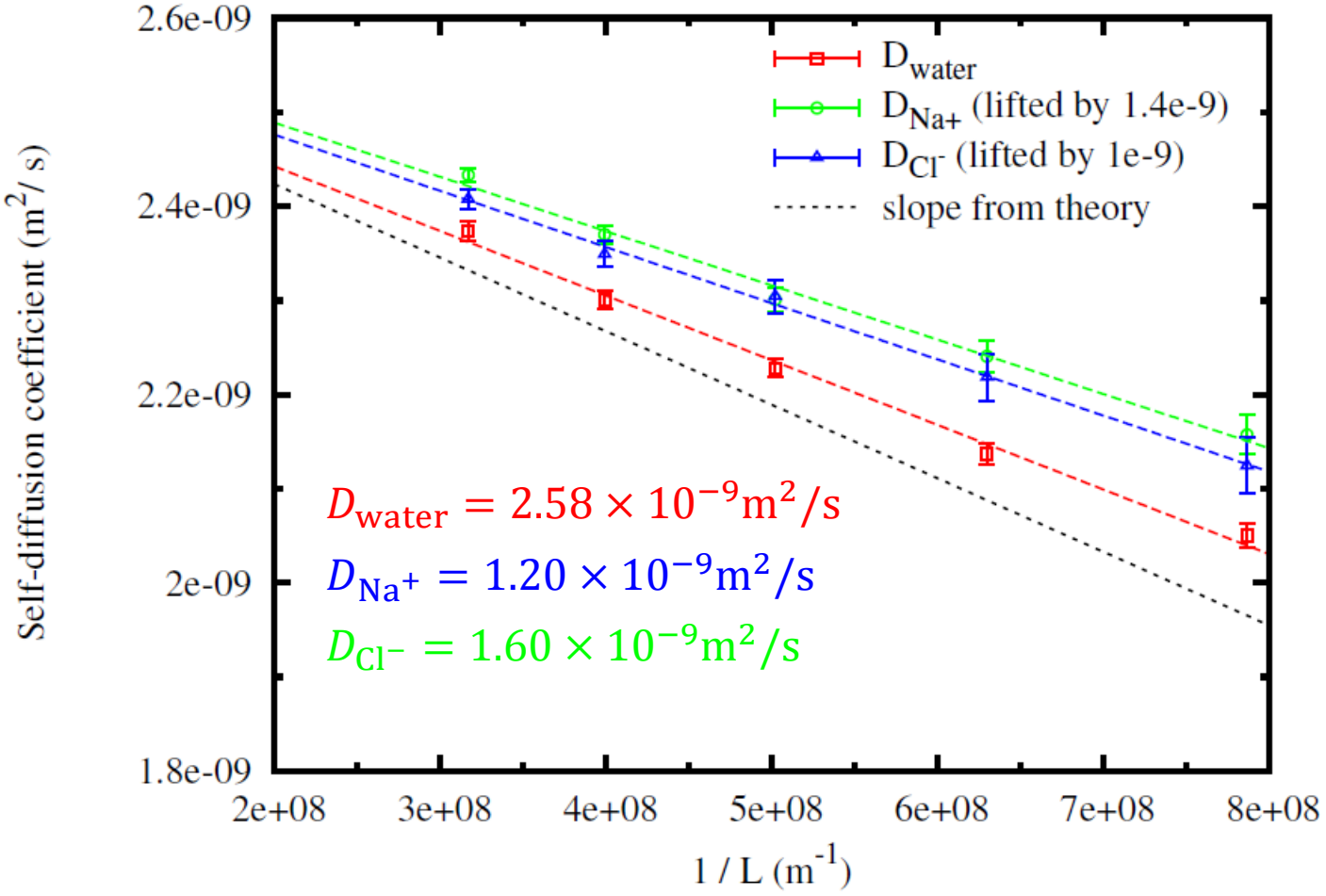
10 WCA Particles in a chain

$n_{\mathrm{fl}}: n_{\mathrm{chain}} = 100:1$

WCA Fluid with $\rho^* = 0.85$ and $T^* = 1$



H_2O + NaCl (0.87 M, 300K, 1 atm)



Despite quite different values of the three self-diffusion coefficients, their slopes for the finite-system-size correction are very similar.

Part 3.

Estimation of Shear Viscosity

1. Equilibrium versus non-equilibrium MD estimation
2. Finite-system-size effects

Estimating Viscosity from Equilibrium MD

Green-Kubo formula

$$\eta = \frac{V}{k_B T}\int_0^\infty \langle p_{xy}(0)p_{xy}(t)\rangle dt$$

- Perturbation to the equilibrium system

$$\dot{\mathbf{x}}_i = \mathbf{v}_i + \gamma y_i \hat{\mathbf{e}}_x$$

$$m\dot{\mathbf{v}}_i = \mathbf{F}_i - \gamma m v_{y,i}\hat{\mathbf{e}}_x$$

- Linear response theory ($\gamma \to 0$)

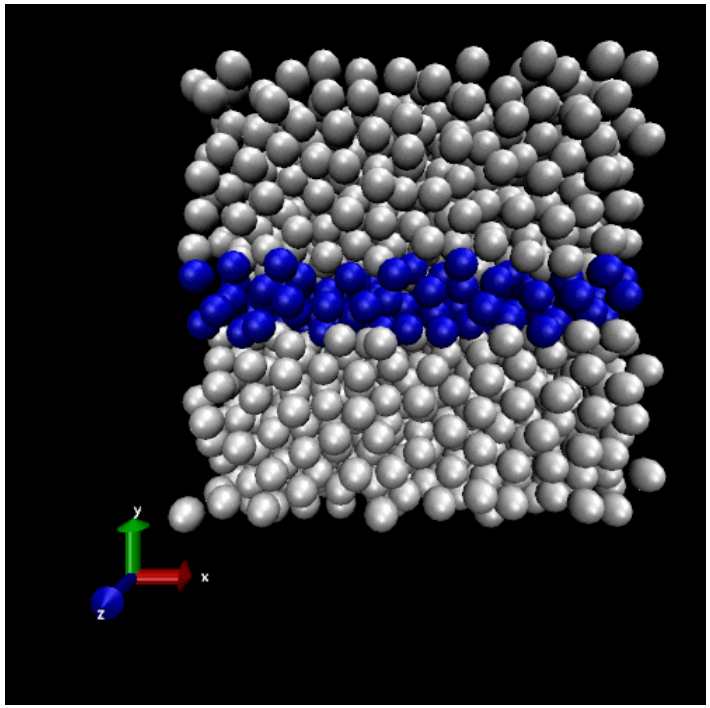
$$\langle p_{xy}(t)\rangle_{\text{shear}} = -\frac{\gamma V}{k_B T}\int_0^t \langle p_{xy}(0)p_{xy}(t')\rangle dt'$$

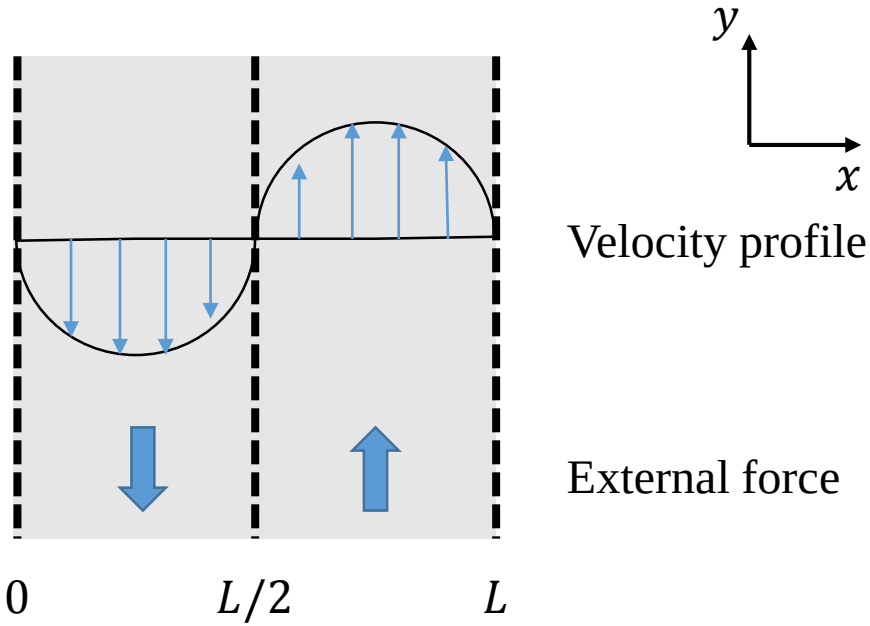
- Definition of shear viscosity

$$\eta = -\lim_{t\to\infty}\frac{\langle p_{xy}(t)\rangle_{\text{shear}}}{\gamma}$$

Non-Equilibrium MD Simulation (Poiseuille Flow)

- Position-dependent external force
- Steady-state velocity profile → shear viscosity
- Under periodic boundary conditions





Backer et al, J. Chem. Phys. 122, 154503 (2005).

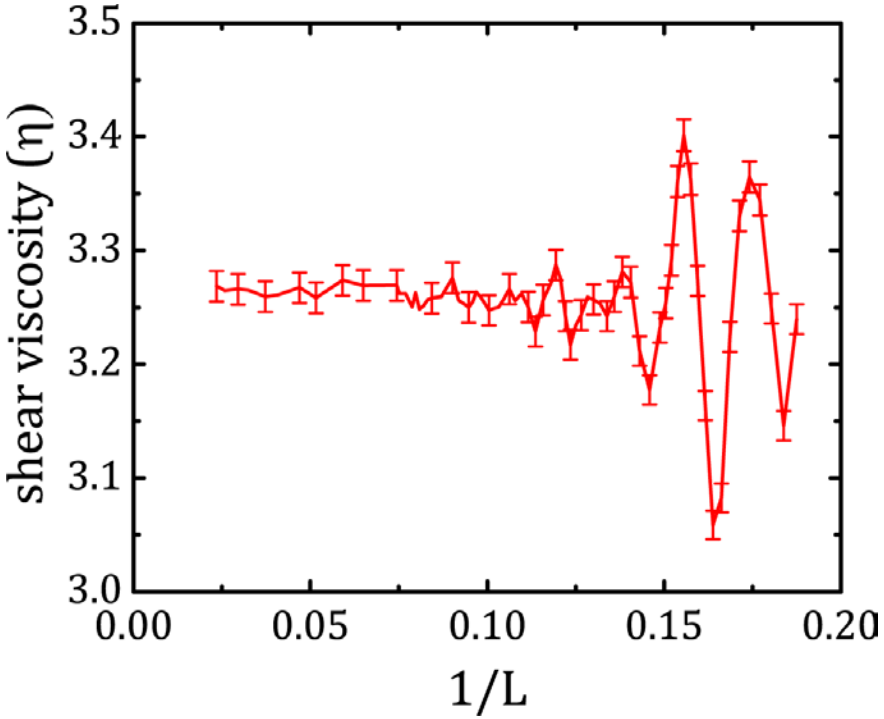
Comparison

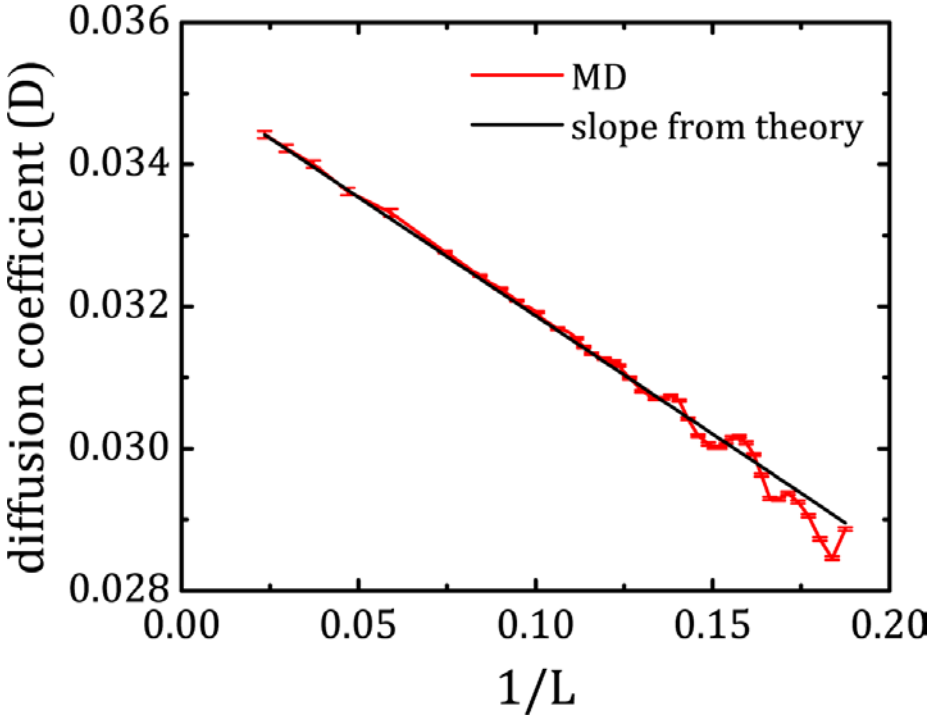
- Non-equilibrium MD
  - (+) Direct method
  - (+) Good signal to noise ratio
  - (–) Unrealistically large shear rate
  - (–) Use of thermostat
- Equilibrium MD
  - (+) theoretically sound
  - (–) large statistical error
    - The level of statistical error can be estimated by our approach.

Finite-System-Size Effects

Lennard-Jones Fluid
$(\boldsymbol{\rho^* = 0.8442},\ \boldsymbol{T^* = 0.722})$

- $N_{fl} = 128 \sim 65536$
- 16384 MD samples





Summary

- Equivalence of the VACF and MSD methods for the self-diffusion coefficient
- Scaling behavior of the statistical error

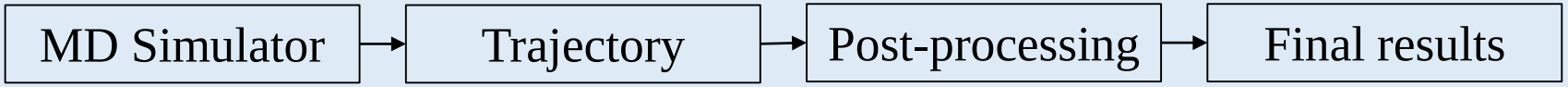
$$\langle \varepsilon^2(t) \rangle \approx \frac{a(t)}{\mathcal{N}\mathcal{T}n^*}$$

  - Theoretical error estimates from the GPA
  - $n^* \ll n$ for full particle-averaging

- $1/L$ correction for the finite-system-size effect on the diffusion coefficient
  - Microscopic interpretation by using the tail of VACF
  - In multi-species systems, the same correction applies to each species.
- Finite-system-size effects on the shear viscosity

LAMMPS Tutorial

Traditional MD simulation steps



- Dumping trajectory may not always be a good idea.
- Hard-coding may be required.

LAMMPS is a very powerful tool for particle-based simulation methods.

In the tutorial session, I will demonstrate:

How to calculate physical quantities (on the fly) from LAMMPS.

THANK YOU!!!