Direct Computation of Viscosity from Differentiable Atomistic Simulations

Rayan Chatterjee
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi, India 110016
rayanchatterjee2@gmail.com

Srikanth Sastry
Theoretical Sciences Unit,
Jawaharlal Nehru Centre for Advanced Scientific Research,
Jakkur Campus, Bengaluru-560064, India
sastry@jncasr.ac.in

N M Anoop Krishnan*
Department of Civil Engineering
and School of Artificial Intelligence,
Indian Institute of Technology Delhi,
Hauz Khas, New Delhi 110016, India
krishnan@iitd.ac.in

Abstract

Shear viscosity calculation from molecular dynamics simulations demands long equilibration times and extensive statistical averaging to achieve convergence. We address this challenge by presenting a direct automatic-differentiation method to compute shear viscosity from differentiable molecular dynamics simulations. Our approach differentiates microscopic shear stress with respect to applied shear rate, but crucially identifies a characteristic timescale τ_α that defines a stable window for reliable gradient computation. We demonstrate that τ_α marks the onset of chaotic divergence in stress dynamics and corresponds to the timescale where stress autocorrelations decay by $> 90\%$, providing both theoretical justification and physical insight. Through systematic validation on Weeks-Chandler-Andersen systems across multiple realizations, our method yields viscosity estimates (1.92 ± 0.38) that agree with Green-Kubo predictions (2.24 ± 0.24) within statistical uncertainty, while circumventing the noise accumulation inherent in long-time correlation approaches. The identified stability window concept establishes a general framework for extracting transport properties from differentiable simulations of chaotic systems before gradient explosion occurs, with promising applications to thermal conductivity and diffusion coefficients. This work provides a principled solution to gradient instability in differentiable physics, enabling reliable parameter optimization and property prediction in complex molecular systems.

1 Introduction

Differentiable simulation has emerged as a transformative paradigm in scientific machine learning, enabling end-to-end gradient-based optimization of physical parameters Innes et al. [2019], Rackauckas

*Corresponding author

et al. [2020], discovery of governing equations Ingraham et al. [2018], Cranmer [2023], and design of control policies for complex systems Toussaint et al. [2018], Sanchez-Gonzalez et al. [2020]. Modern differentiable physics engines such as JAX-MD Schoenholz and Cubuk [2020], Brax Freeman and et al. [2021], and DiffTaichi Hu et al. [2019] have demonstrated remarkable success in learning interatomic potentials Unke et al. [2021], Gangan et al. [2025], Doerr et al. [2021], Thölke and De Fabritiis [2022], Thölke et al. [2024], Batzner et al. [2022] and optimizing material properties through end-to-end differentiable workflows.

However, a fundamental challenge arises when applying automatic differentiation to chaotic physical systems. Backpropagation through chaotic dynamical systems is well known to suffer from gradient explosion or vanishing due to exponential sensitivity to initial conditions Lea et al. [2000], Wolf et al. [1985]. Many systems of scientific interest—including potential fine-tuning for atomistic simulation Gangan et al. [2025], non-equilibrium molecular dynamics J Evans and P Morriss [2007], and long-time material behavior Allen and Tildesley [2017]—exhibit this extreme sensitivity characterized by positive Lyapunov exponents. Recent work has highlighted how this chaotic sensitivity causes gradients to explode or vanish exponentially when backpropagating through extended simulation trajectories Lea et al. [2000], Metz et al. [2021], fundamentally limiting the applicability of gradient-based methods to short-time dynamics.

Existing mitigation strategies for gradient instability in chaotic systems include gradient clipping Pascanu et al. [2013], truncated backpropagation, and least-squares shadowing techniques Wang et al. [2014], Ni and Wang [2017], Blonigan and Wang [2018], but these approaches often sacrifice either physical accuracy or computational efficiency. The challenge becomes particularly acute for transport property estimation, where physically meaningful quantities like viscosity, thermal conductivity, and diffusion coefficients emerge only from long-time correlations Green [1954], Kubo [1957].

Classical approaches for computing transport coefficients rely on well-established methods: the Green–Kubo formalism applied to equilibrium MD simulations Green [1954] computes properties from autocorrelation functions of microscopic fluxes, while non-equilibrium molecular dynamics Evans and Morriss [1984] directly applies perturbations and measures the response. Both methods require careful statistical analysis and selection of integration windows to avoid noise accumulation. While differentiable simulations have successfully leveraged automatic differentiation for force field optimization Gangan et al. [2025], direct gradient-based estimation of transport properties via differentiable MD remains largely unexplored due to chaos-induced instability, despite its potential for enabling rapid property screening and inverse design.

In this work, we address this challenge by developing a principled framework for transport property estimation that exploits the finite-time stability of chaotic systems. Specifically, we focus on shear viscosity computation and demonstrate that by identifying a characteristic Lyapunov-like timescale τ_α, one can extract accurate viscosity estimates before gradient explosion undermines the computation. This approach bridges the gap between classical molecular simulation methods and modern differentiable programming paradigms, providing both theoretical insight into the stability limits of differentiable chaotic systems and a practical method for reliable transport property estimation.

Our contributions are as follows:

- We develop a direct gradient-based method for viscosity estimation using automatic differentiation of atomistic simulations that circumvents traditional correlation-based approaches.
- We establish the theoretical foundation linking gradient explosion to chaotic sensitivity in stress dynamics and identify a characteristic stability timescale τ_α that provides a principled cutoff for reliable computation.
- We demonstrate that viscosity estimates computed within this stability window match Green–Kubo predictions while avoiding long-time noise accumulation.
- We provide a general framework for identifying optimal time windows in differentiable simulations of chaotic systems, with broad implications for other transport coefficients and scientific machine learning applications involving long-time dynamics.

2 Methods

2.1 Differentiable MD Framework

We implement differentiable molecular dynamics using SLLOD equations Lees and Edwards [1972] with Nosé–Hoover thermostat [Nosé, 1984, Hoover, 1985] to model homogeneous shear flow. The system consists of N particles interacting via the Weeks–Chandler–Andersen (WCA) potential Weeks et al. [1971]: $U(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6] + \epsilon$ for $r \leq 2^{1/6}\sigma$, zero otherwise.

The governing equations are:

$$\dot{\mathbf{r}}_i = \mathbf{p}_i/m_i + \dot{\gamma} y_i \hat{\mathbf{x}}, \quad \dot{\mathbf{p}}_i = \mathbf{F}_i - \dot{\gamma} p_{y,i}\hat{\mathbf{x}} - \xi \mathbf{p}_i, \quad \dot{\xi} = \frac{1}{Q}\left[\sum_i \frac{\mathbf{p}_i^2}{m_i} - 3Nk_BT\right] \tag{1}$$

where $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U$, ξ is the thermostat variable, and $Q = 1.0$ is the thermostat mass.

Simulation parameters: $N = 256$, $\rho = 0.8442$, $T = 0.722$ (reduced units), $\Delta t = 0.005$, periodic boundaries. Equilibration: 10,000 timesteps. For validation: minimal $N = 2$ system with analytical comparison.

2.2 Automatic Differentiation Approach

Viscosity is computed as $\eta = d\sigma_{xy}/d\dot{\gamma}$ where the microscopic stress is:

$$\sigma_{xy} = \frac{1}{V}\left[\sum_{i=1}^{N} \frac{p_{x,i}p_{y,i}}{m_i} + \sum_{i<j} r_{x,ij}F_{y,ij}\right] \tag{2}$$

We implement the full MD loop in JAX-MD Schoenholz and Cubuk [2020] and apply forward-mode automatic differentiation (`jax.jacfwd`) to compute $d\sigma_{xy}/d\dot{\gamma}$ exactly through all timesteps. This avoids finite-difference noise but remains sensitive to chaotic divergence.

2.3 Stability Window Identification

To identify the optimal extraction window, we simulate trajectory pairs at zero shear and infinitesimal shear $\dot{\gamma} = 10^{-8}$ with random initial configurations. The exponential divergence of stress differences reveals a characteristic Lyapunov-like timescale τ_α marking the stability boundary. We correlate this with stress autocorrelation decay to establish the physical significance of the stability window.

3 Results

3.1 Minimal System Validation

We validate our AD implementation using a two-particle system with analytical comparison (see Appendix). Specifically, we derive the functional form of the gradient analytically when unrolled through the simulation which is used to validate the automatic differentiation. Table 1 shows reasonable agreement between analytical and AD gradients for both kinetic and total stress (both kinetic and potential) contributions, confirming implementation correctness before applying to larger systems where analytical solutions are intractable.

Step	Analytical (KE)	AD (KE)	Analytical (Total)	AD (Total)
0	0	0	0	0
1	4.028×10^{-6}	4.029×10^{-6}	0.00915	0.00846
2	0.00890	0.00846	0.0167	0.0171

Table 1: Analytical vs AD gradients for kinetic-energy (KE) and total stress contributions.

3.2 Chaotic Stress Dynamics and Stability Timescale

Figure 1(a) demonstrates chaotic sensitivity in a 256-particle system by comparing stress evolution under zero shear (blue) and infinitesimal shear $\dot{\gamma} = 10^{-8}$ (red). The inset shows exponential

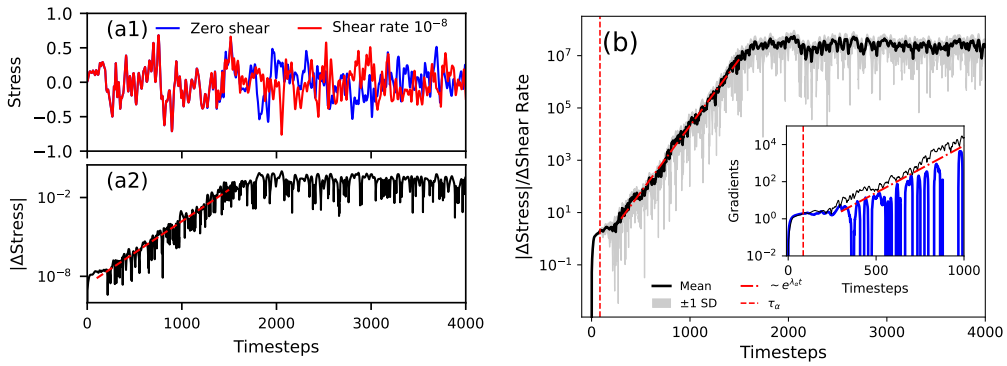

Figure 1: (a) Stress trajectories for zero (blue) and infinitesimal (red) shear rates. Inset: exponential stress divergence. (b) Ensemble-averaged divergence $\pm$ 1 SD (grey) with exponential fit (red) yielding τ_α (vertical line). Inset: viscosity from AD (blue) and divergence scaling (black) show similar patterns.

divergence of stress differences (black line), confirming chaotic dynamics with linear growth on semi-log scale.

To quantify the characteristic timescale, we simulate ensembles of 10 trajectories each at zero and infinitesimal shear rates. Figure 1(b) shows mean absolute divergence growing exponentially over six orders of magnitude before saturation. Exponential fitting yields λ_α and the characteristic timescale $\tau_\alpha = 1/\lambda_\alpha = 85.47 \pm 3.11 \times 10^{-5}$ timesteps. The inset confirms that AD-computed viscosity exhibits similar growth-plateau-explosion behavior as trajectory divergence, validating our gradient computation approach.

3.3 Viscosity Convergence and Physical Interpretation

Figure 2 demonstrates that τ_α provides an optimal extraction window where AD viscosity estimates agree with Green-Kubo (GK) predictions while avoiding noise accumulation.

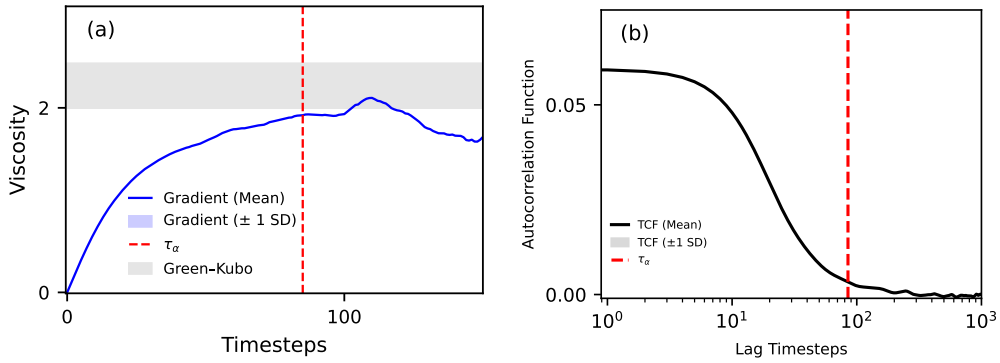

Figure 2: (a) AD viscosity (blue band) vs GK estimate (grey band) at τ_α (red line). (b) Stress autocorrelation function decay with τ_α marking $> 90\%$ decorrelation.

At τ_α, GK yields $\eta = 2.24 \pm 0.24$ (consistent with literature Hartkamp et al. [2013]) while AD gives $\eta = 1.92 \pm 0.38$, representing agreement within statistical uncertainty. This confirms that τ_α defines a stable window for viscosity extraction before gradient explosion.

Crucially, τ_α coincides with physical decorrelation timescales. Figure 2(b) shows the stress time correlation function (TCF) decays by $> 90\%$ before τ_α (red line), beyond which correlations saturate and noise dominates. This connection between chaotic divergence and correlation decay provides physical insight into why the stability window works: gradients remain meaningful while stress correlations carry transport information, but become unreliable once correlations are lost to noise.

4 Conclusions

We present a direct gradient-based method for viscosity estimation from differentiable MD that addresses gradient explosion in chaotic systems. By identifying a Lyapunov-like stability timescale τ_α, we extract viscosity estimates ($\eta = 1.92 \pm 0.38$) matching Green-Kubo predictions ($\eta = 2.24 \pm 0.24$) before chaos-induced instability. Crucially, τ_α coincides with $> 90\%$ stress decorrelation decay, providing physical insight into the stability window. This framework enables gradient-based computation of transport properties from chaotic systems, with broad implications for differentiable physics and materials design applications requiring end-to-end optimization of transport phenomena.

5 Limitations and Future Work

Our approach exhibits $\pm 20\%$ uncertainty due to chaotic sensitivity and 2-3× computational overhead from gradient storage. The method is limited to short-time dynamics within the stability window and currently validated only on WCA systems. Dependence on differentiable simulation infrastructure may restrict applicability to complex force fields. Future work will explore gradient stabilization techniques (least-squares shadowing, ensemble methods) to extend the stability window, validate on realistic materials (water, polymers, ionic systems), and demonstrate the framework for other transport coefficients such as thermal conductivity, and heat capacity. Theoretical development should establish rigorous connections between Lyapunov timescales and transport property convergence, ultimately enabling reliable differentiable transport property estimation for materials discovery.

6 Acknowledgement

The authors gratefully acknowledge the computational resources provided by the High Performance Computing (HPC) facility at the Indian Institute of Technology Delhi.

References

Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022. doi: 10.1038/s41467-022-29939-5.

Patrick J. Blonigan and Qiqi Wang. Multiple shooting shadowing for sensitivity analysis of chaotic dynamical systems. *Journal of Computational Physics*, 354:447–475, 2018. doi: 10.1016/j.jcp.2017.10.032.

Miles Cranmer. Interpretable machine learning for science with pysr and symbolicregression.jl, 2023. URL `https://arxiv.org/abs/2305.01582`.

S. Doerr et al. Torchmd: A deep learning framework for molecular simulations. *Journal of Chemical Theory and Computation*, 17(4):2355–2363, 2021.

D. J. Evans and G. P. Morriss. Nonlinear-response theory for steady planar couette flow. *Physical Review A*, 30(3):1528–1530, 1984.

Daniel Freeman and et al. Brax–a differentiable physics engine for large scale rigid body simulation. `https://github.com/google/brax`, 2021.

A. S. Gangan, E. D. Cubuk, S. S. Schoenholz, M. Bauchy, and N. M. Anoop Krishnan. Force-field optimization by end-to-end differentiable atomistic simulation. *Journal of Chemical Theory and Computation*, 21(12):5867–5879, 2025. doi: 10.1021/acs.jctc.4c01784. URL `https://doi.org/10.1021/acs.jctc.4c01784`. PMID: 40472008.

M. S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. ii. irreversible processes in fluids. *The Journal of Chemical Physics*, 22(3):398–413, 1954.

R. Hartkamp, B. D. Todd, and S. Luding. A constitutive framework for the non-newtonian pressure tensor of a simple fluid under planar flows. *J. Chem. Phys.*, 138(24):244508, 2013. doi: 10.1063/1.4810746. URL http://link.aip.org/link/?JCP/138/244508/1.

William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. Difftaichi: Differentiable programming for physical simulation. *arXiv preprint arXiv:1910.00935*, 2019.

John Ingraham, Adam J. Riesselman, Chris Sander, and Debora S. Marks. Learning protein structure with a differentiable simulator. In *International Conference on Learning Representations*, 2018. URL https://api.semanticscholar.org/CorpusID:108301299.

Mike Innes, Alan Edelman, Keno Fischer, Chris Rackauckas, Elliot Saba, Viral B Shah, and Will Tebbutt. A differentiable programming system to bridge machine learning and scientific computing. *arXiv preprint arXiv:1907.07587*, 2019.

Denis J Evans and Gary P Morriss. *Statistical mechanics of nonequilbrium liquids*. ANU Press, 2007.

R. Kubo. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12 (6):570–586, 1957.

Daniel J Lea, Myles R Allen, and Thomas WN Haine. Sensitivity analysis of the climate of a chaotic system. *Tellus A: Dynamic Meteorology and Oceanography*, 52(5):523–532, 2000.

A. W. Lees and S. F. Edwards. The computer study of transport processes under extreme conditions. *Journal of Physics C: Solid State Physics*, 5:1921–1929, 1972.

Luke Metz, C Daniel Freeman, Samuel S Schoenholz, and Tal Kachman. Gradients are not all you need. *arXiv preprint arXiv:2111.05803*, 2021.

Angxiu Ni and Qiqi Wang. Sensitivity analysis on chaotic dynamical systems by non-intrusive least squares shadowing (nilss). *Journal of Computational Physics*, 347:56–77, 2017. doi: 10.1016/j.jcp.2017.06.033.

Shuichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28, pages 1310–1318, 2013. URL https://proceedings.mlr.press/v28/pascanu13.html.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International conference on machine learning*, pages 8459–8468. PMLR, 2020.

Samuel S Schoenholz and Ekin D Cubuk. Jax md: A framework for differentiable physics. *Advances in Neural Information Processing Systems*, 33:11428–11441, 2020.

P. Thölke and G. De Fabritiis. Torchmd-net: Equivariant transformers for neural network potentials. *arXiv preprint arXiv:2202.02541*, 2022.

P. Thölke et al. Torchmd-net 2.0: Advances in equivariant neural network potentials. *Journal of Chemical Theory and Computation*, 2024.

Marc A Toussaint, Kelsey Rebecca Allen, Kevin A Smith, and Joshua B Tenenbaum. Differentiable physics and stable modes for tool-use and manipulation planning. 2018.

Oliver T Unke, Stefan Chmiela, Huziel E Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T Schutt, Alexandre Tkatchenko, and Klaus-Robert Muller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021.

Q. Wang, R. Hu, and P. Blonigan. Least squares shadowing sensitivity analysis of chaotic limit cycle oscillations. *Journal of Computational Physics*, 267:210–224, 2014.

J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54:5237–5247, 1971.

Alan Wolf, Jack B Swift, Harry L Swinney, and John A Vastano. Determining lyapunov exponents from a time series. *Physica D: nonlinear phenomena*, 16(3):285–317, 1985.

Appendix A SLLOD Equations with Nosé–Hoover Thermostat

Appendix A.1 Equations of Motion

In this appendix, we describe the governing equations for simulating shear flow in molecular dynamics using the SLLOD formulation with a Nosé–Hoover thermostat, the numerical integration scheme employed, and the procedure for computing viscosity gradients via automatic differentiation.

Appendix A.2 Equations of Motion

The SLLOD equations extend Newtonian dynamics to model homogeneous shear flow while maintaining constant temperature through a Nosé-Hoover thermostat. For particle i:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \dot{\gamma}\, y_i\, \hat{\mathbf{x}}, \tag{3}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \dot{\gamma}\, p_{y,i}\, \hat{\mathbf{x}} - \xi\, \mathbf{p}_i, \tag{4}$$

$$\dot{\xi} = \frac{1}{Q}\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{m_i} - 3Nk_BT\right]. \tag{5}$$

Here:

- $\mathbf{r}_i = (x_i, y_i, z_i)$: particle position vector.
- $\mathbf{p}_i = (p_{x,i}, p_{y,i}, p_{z,i})$: particle momentum relative to streaming velocity.
- m_i: mass of particle i.
- $\dot{\gamma}$: imposed shear rate.
- y_i: y-coordinate of particle i.
- $\hat{\mathbf{x}}$: unit vector in x-direction.
- $\mathbf{F}_i$: interparticle force on particle i, $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U$.
- ξ: Nosé–Hoover thermostat variable controlling kinetic temperature.
- Q: thermostat mass parameter.
- N: total number of particles.
- k_B: Boltzmann constant.
- T: target (controlled) temperature.

The interaction potential U is given by the Weeks–Chandler–Andersen (WCA) potential which is a purely repulsive potential obtained by truncating and shifting the Lennard-Jones potential at its minimum, given by

$$U(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon, & r \le 2^{1/6}\sigma, \\ 0, & r > 2^{1/6}\sigma \end{cases}$$

where ϵ and σ are the energy and length scale parameters respectively.

Appendix A.3 Numerical Integration

We integrate Eqs. (3)-(5) using a reversible, symplectic velocity Verlet-like scheme adapted for SLLOD dynamics. This scheme alternates half-step updates for the thermostat and momenta with full-step position updates, followed by force recalculation. The time-reversibility and symplectic nature ensure energy and temperature control over long runs.

1. Half-step thermostat update.
2. Half-step momentum update.
3. Position update.
4. Force recomputation.
5. Second half-step momentum update.
6. Second half-step thermostat update.

Appendix A.4 Gradient Computation

The shear viscosity is defined as:

$$\eta = \frac{d\sigma_{xy}}{d\dot{\gamma}},$$

where the microscopic shear stress σ_{xy} is given by:

$$\sigma_{xy} = \frac{1}{V}\left[\sum_{i=1}^{N}\frac{p_{x,i}p_{y,i}}{m_i} + \sum_{i<j} r_{x,ij}F_{y,ij}\right] \tag{6}$$

We implement the full MD loop in a differentiable framework (JAX-MD) and apply forward-mode automatic differentiation (`jax.jacfwd`) with respect to $\dot{\gamma}$. This yields the exact derivative of σ_{xy} through all timesteps, enabling direct viscosity computation without finite-difference noise.

In a differentiable MD framework (e.g., JAX-MD), the full simulation loop is implemented as a differentiable function of $\dot{\gamma}$. We then apply forward-mode automatic differentiation (jax.jacfwd) to backpropagate through all integration steps, computing the total derivative $\frac{d\sigma_{xy}}{d\dot{\gamma}}$ exactly (up to floating-point precision).

By taking $m_i = 1$ in Eq. (6) the stress for a minimal two-particle system at the initial time is

$$\sigma^0_{xy} = \sum_{i=1}^{2} p^0_{xi}p^0_{yi} + 24\sum_{i=1}^{2}\sum_{j>i}(x^0_j - x^0_i)(2/r^{14} - 1/r^8)(y^0_j - y^0_i). \tag{7}$$

where all terms are constant. Thus the derivative with respect to $\dot{\gamma}$ is 0. After 1 step the derivative is

$$\begin{aligned}\frac{d\sigma^1_{xy}}{d\dot{\gamma}} = &\sum_{i=1}^{2} -p^0_{yi}p^1_{yi}dt + \sum_{i=1}^{2} 24\sum_{j>i}(y^0_j - y^0_i)dt\left[2/r^{14} - 1/r^8\right](y^1_j - y^1_i) + \\ &(x^1_j - x^1_i)\left(-28/r^{16} + 8/r^{10}\right)(x^1_j - x^1_i)(y^0_j - y^0_i)(y^1_j - y^1_i)dt\end{aligned} \tag{8}$$

where dt is the time step. The first term in Eq. (8) comes from the kinetic component of the stress, and the second term comes from the interaction component. The derivative after two time steps is

$$\begin{aligned}\frac{d\sigma^2_{xy}}{d\dot{\gamma}} = &\sum_{i=1}^{2}\left[-p^0_{yi}dt - p^1_{yi}dt + 24\sum_{j\neq i}(-28/r^{15} + 8/r^9)(x^1_j - x^1_i)^2(y^0_j - y^0_i)dt\right]p^2_{yi} \\ &+24\sum_{i=1}^{2}p^2_{yi}\sum_{j\neq i}(2/r^{14} - 1/r^8)(y^0_j - y^0_i)dt \\ &+24\sum_{i=1}^{2}\sum_{j>i}\left[(y^0_j - y^1_j - p^0_{yj}dt)dt - (y^0_i - y^1_i - p^0_{yi}dt)dt\right](2/r^{14} - 1/r^8)(y^2_j - y^2_i) \\ &+24\sum_{i=1}^{2}\sum_{j>i}\left[(y^0_j - y^1_j - p^0_{yj}dt)dt - (y^0_i - y^1_i - p^0_{yi}dt)dt\right](-28/r^{16} + 8/r^{10})(x^2_j - x^2_i)^2(y^2_j - y^2_i)\end{aligned} \tag{9}$$