Lattice QCD gauge configuration generation at near physical point

孙玮

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

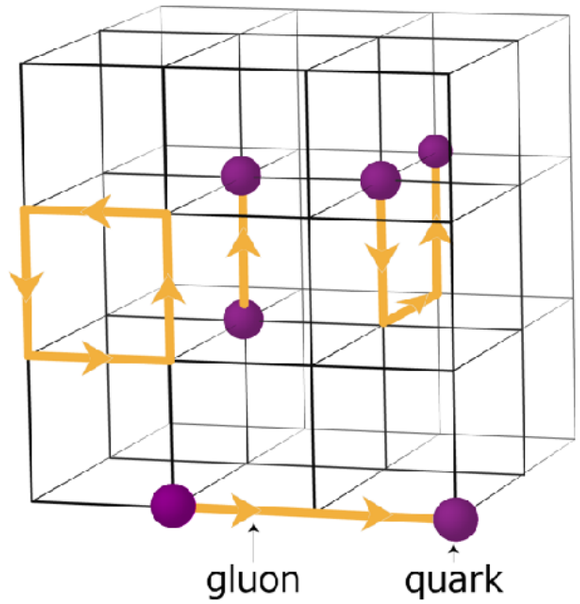
中国格点QCD第一届年会
Oct.30 - Nov. 2, 2021, online

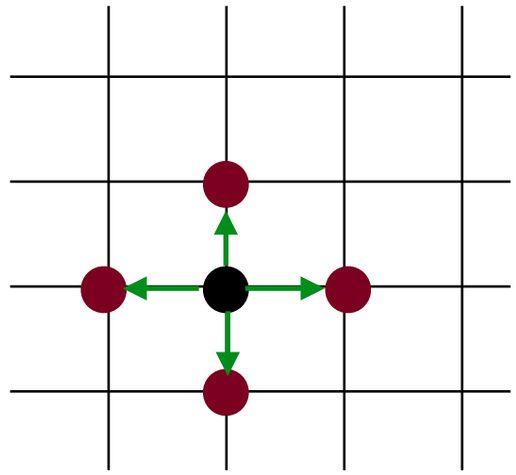
Outline

- Computational Perspective of Lattice QCD
- Brief Introduction to Hybrid Monte Carlo
- Brief Introduction to Multigrid Method
- Hasenbusch Preconditioning Method
- Summary

Computational Perspective of Lattice QCD

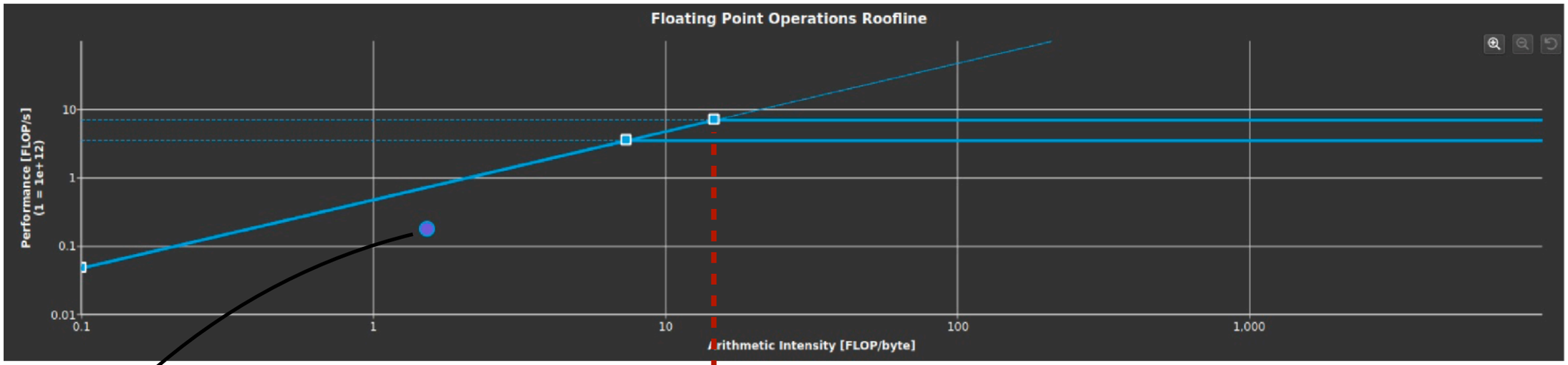
- Lagrangian formulation
  - Discretized spacetime, finite volume: finite *d.o.f*
  - Computer simulation: Markov Chain Monte Carlo, Importance Sampling.
- LQCD need supercomputer / cluster to simulate
  - $D_{x,y} = \sum_{\mu} (1-\gamma_\mu) U_\mu(x) \delta_{x+\hat{\mu},y} + (1+\gamma_\mu) U_\mu(x-\hat{\mu})^\dagger \delta_{x-\hat{\mu},y}$
  - Four dimensional 8(9) point stencil operation
- HPC(High Performance Computing)
  +
  HTC(High Throughput Computing)





Computational Perspective of Lattice QCD

- Roofline model (by Berkeley Lab)
- A performance analysis and optimization model



Memory bound | **Compute bound**

- Lattice QCD Wilson dslash AI (Arithmetic Intensity) ~ 1 Flops/Byte
- Memory bound
- Optimization strategy: increase the AI

Brief Introduction to Hybrid Monte Carlo

- Euclidean path integral on lattice

$$\langle O \rangle = \frac{1}{Z} \int \mathscr{D}U \mathscr{D}\bar{\psi} \mathscr{D}\psi O[U, \bar{\psi}, \psi] e^{-S_g - S_f}$$

- S_g gauge action, $S_f = \bar{\psi} D \psi$ fermion action, D Dirac operator
- Integrated over Grassmann variable $\bar{\psi}$, ψ

$$\langle O \rangle = \frac{1}{Z} \int \mathscr{D}U O'[U] \det D \; e^{-S_g}$$

- $\gamma_5 D \gamma_5 = D^\dagger$-> $\det D^\dagger = \det D$, $\det D$ is real
- Two flavor degenerate $N_f = 2$, $\det D_u \det D_d = \det D^2 = \det D^\dagger D \geq 0$
- Markov Chain Monte Carlo: $\frac{1}{Z} \det D^\dagger D \; e^{-S_g} \geq 0$ as probability distribution
- Note: for odd number of flavors
  e.g. $N_f = 1$, $\det D = \det \sqrt{D^\dagger D}$, use RHMC algorithm
- Focus on $N_f = 2$ algorithm in this talk

Brief Introduction to Hybrid Monte Carlo

- $\det D$ is nonlocal and nearly impossible to evaluate directly
- Rewrite $N_f = 2$ fermion determinant $\det D^\dagger D$ as bosonic pseudo-fermion field

$$\langle O \rangle = \frac{1}{Z}\int \mathscr{D}U O'[U] \det D^\dagger D \; e^{-S_g} = \frac{1}{Z}\int \mathscr{D}U \mathscr{D}\bar{\phi}\mathscr{D}\phi O'[U] e^{-S_g - S_{pf}}$$

- With pseudo-fermion action $S_{pf} = \phi^\dagger (D^\dagger D)^{-1}\phi$
- Hybrid Monte Carlo(HMC): Molecular Dynamics(MD) + Metropolis Accept/Reject
  - Introduce conjugate momentum $P \in su(3)$ of gauge link U
  - Construct Hamiltonian $H = \frac{1}{2}P^2 + S = \frac{1}{2}P^2 + S_g + S_{pf}$
  - Solve Hamiltonian equation numerically

$$\dot{P} = -\frac{\delta H}{\delta U} = -\frac{\delta S_g}{\delta U} - \frac{\delta S_{pf}}{\delta U} = F_g + F_f$$

$$\dot{U} = \frac{\delta H}{\delta P} = P$$

gauge force
easy to evaluate

fermionic force
time consuming $(D^\dagger D)^{-1}$

  - Metropolis Accept/Reject with probability $P_{acc} = \min(1,\ e^{-\Delta H})$

Basic HMC Algorithm

1. Start from some initial/random gauge field U

2. Generate momentum field $P \sim e^{-\frac{1}{2}P^2}$

3. Generate pseudo-fermion field $\phi \sim e^{-\phi^\dagger (D^\dagger D)^{-1} \phi}$

   (1)Generate gaussian random variable $\eta \sim e^{-\eta^\dagger \eta}$, then $\phi = D^\dagger \eta$

4. Compute Hamiltonian/Energy H with U

5. MD trajectory evolution with e.g. leapfrog integrator

   (1)Trajectory length τ_0, integration step N, step size $\epsilon = \frac{\tau_0}{N}$

   (2)Leapfrog integration $P(0) \rightarrow P(\epsilon/2) \rightarrow P(\epsilon)$, $U(0) \longrightarrow U(\epsilon)$

   every momentum P update step need to solve $(D^\dagger D)^{-1}$!

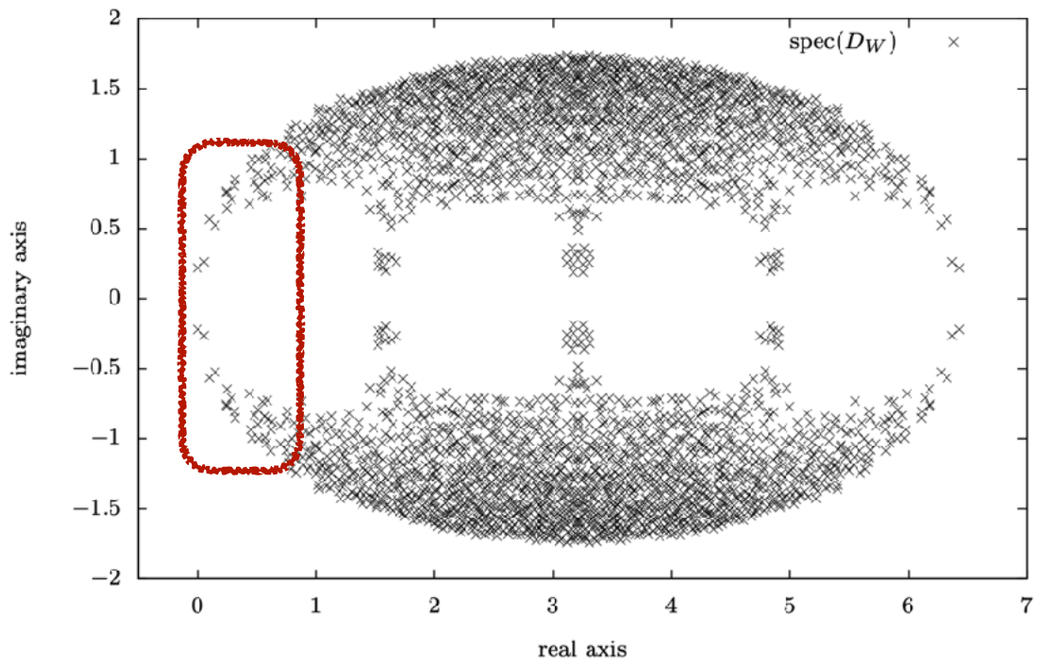
6. Compute Hamiltonian/Energy H′ with new gauge field U′

7. Accept/Reject U′ with $P_{acc} = \min(1, e^{-\Delta H})$, $\Delta H = H' - H$

8. if(Accept): next trajectory from U′; if(Reject): start from old U

Iterative Krylov Subspace Solvers

- Most time consuming part in LQCD: propagator D^{-1} / HMC $(D^\dagger D)^{-1}$



[J. Brannick *et.al.* **Numer.Math. 132 (2016) no.3, 463-490**]

- Wilson like fermion break chiral symmetry
- Near zero mode of Dirac operator even for massive quark
- Problems of simulating light quarks
  - Regular single-grid Krylov subspace solvers -> critical slowing down: multigrid method 😊
  - Fermionic force F_f blow up -> MD integration unstable: need new algorithm 😭

Even-Odd Preconditioning

- Group lattice sites into even-odd ($\sum x_i$ even or odd)

$$D = \begin{pmatrix} A_{ee} & D_{eo} \\ D_{oe} & A_{oo} \end{pmatrix}$$

- Asymmetric even-odd precondition

$$D = \begin{pmatrix} 1 & 0 \\ D_{oe}A_{ee}^{-1} & 1 \end{pmatrix} \begin{pmatrix} A_{ee} & 0 \\ 0 & A_{oo} - D_{oe}A_{ee}^{-1}D_{eo} \end{pmatrix} \begin{pmatrix} 1 & A_{ee}^{-1}D_{eo} \\ 0 & 1 \end{pmatrix}$$

$$\det D = \det A_{ee} \det \hat{D}_a = \det A_{ee} \det (A_{oo} - D_{oe}A_{ee}^{-1}D_{eo})$$

- Symmetric even-odd precondition

$$D = \begin{pmatrix} A_{ee} & 0 \\ 0 & A_{oo} \end{pmatrix} \begin{pmatrix} 1 & A_{ee}^{-1}D_{eo} \\ A_{oo}^{-1}D_{oe} & 1 \end{pmatrix}$$

$$\det D = \det A_{ee} \det A_{oo} \det \hat{D}_s = \det A_{ee} \det A_{oo} \det (1 - A_{oo}^{-1}D_{oe}A_{ee}^{-1}D_{eo})$$

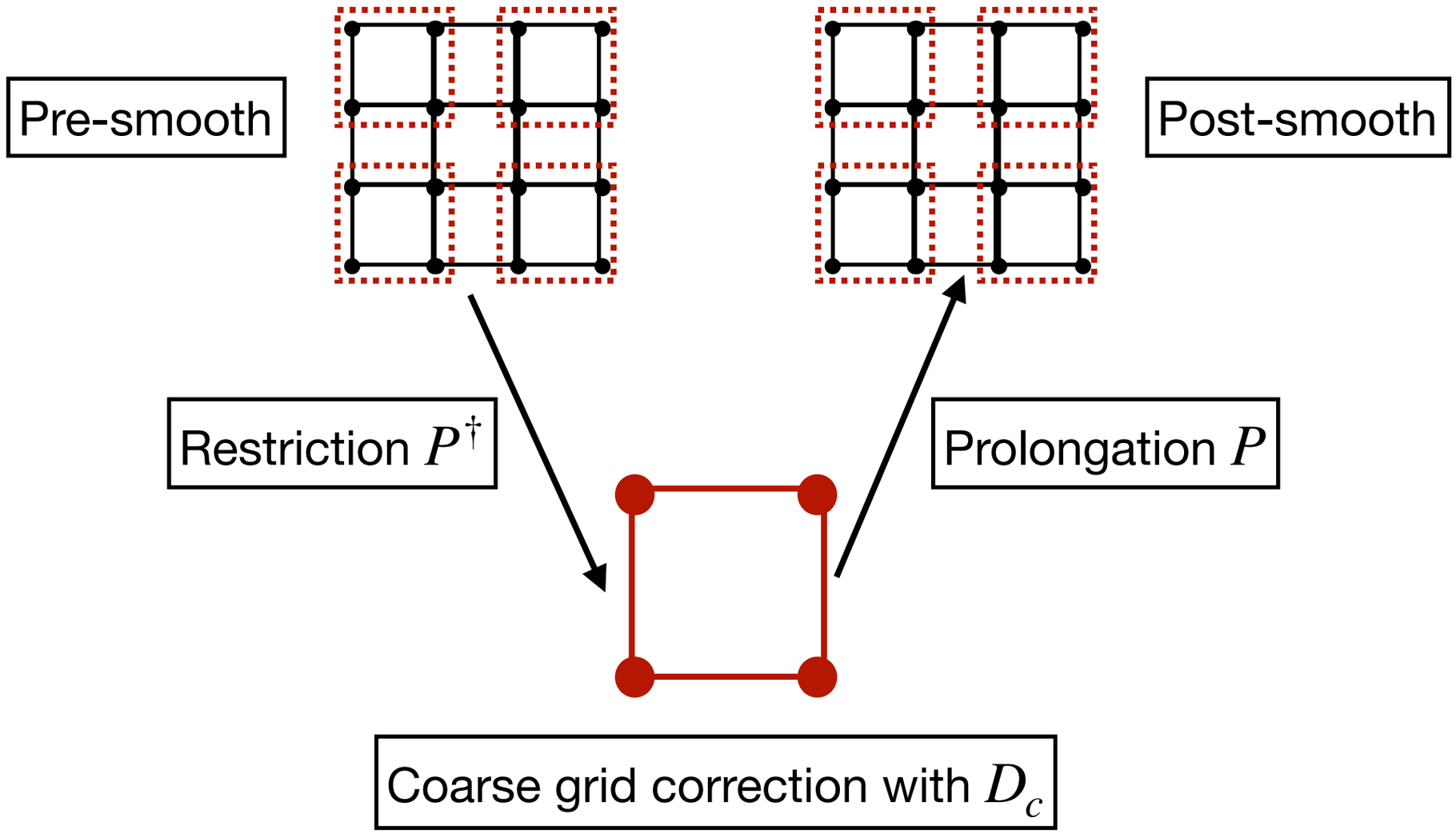
- With preconditioned operator $\hat{D}_a$ and $\hat{D}_s$
  - Better conditioned linear system -> speed up in inversion D^{-1} (e.g. propagator)
  - Factorized pseudo-fermion action in HMC -> more efficient simulation

Brief Introduction to Multigrid Method

- Motivated from multigrid method in solving partial differential equation
- Deal with high and low frequency mode on fine and coarse grid
- Stochastic nature of gauge field -> adaptive aggregation based algebraic multigrid [R. Babich, **Phys.Rev.Lett.105:201602,2010**]
- Capture low frequency mode (slow to convergence) with near null vectors
- Multgrid setup
  - Generate near null vector with $Dv_i = 0$
  - Construct prolongator P and restrictor R from v_i
  - Create coarse operator with Galerkin projection $D_c = RDP$, usually choose $R = P^\dagger$

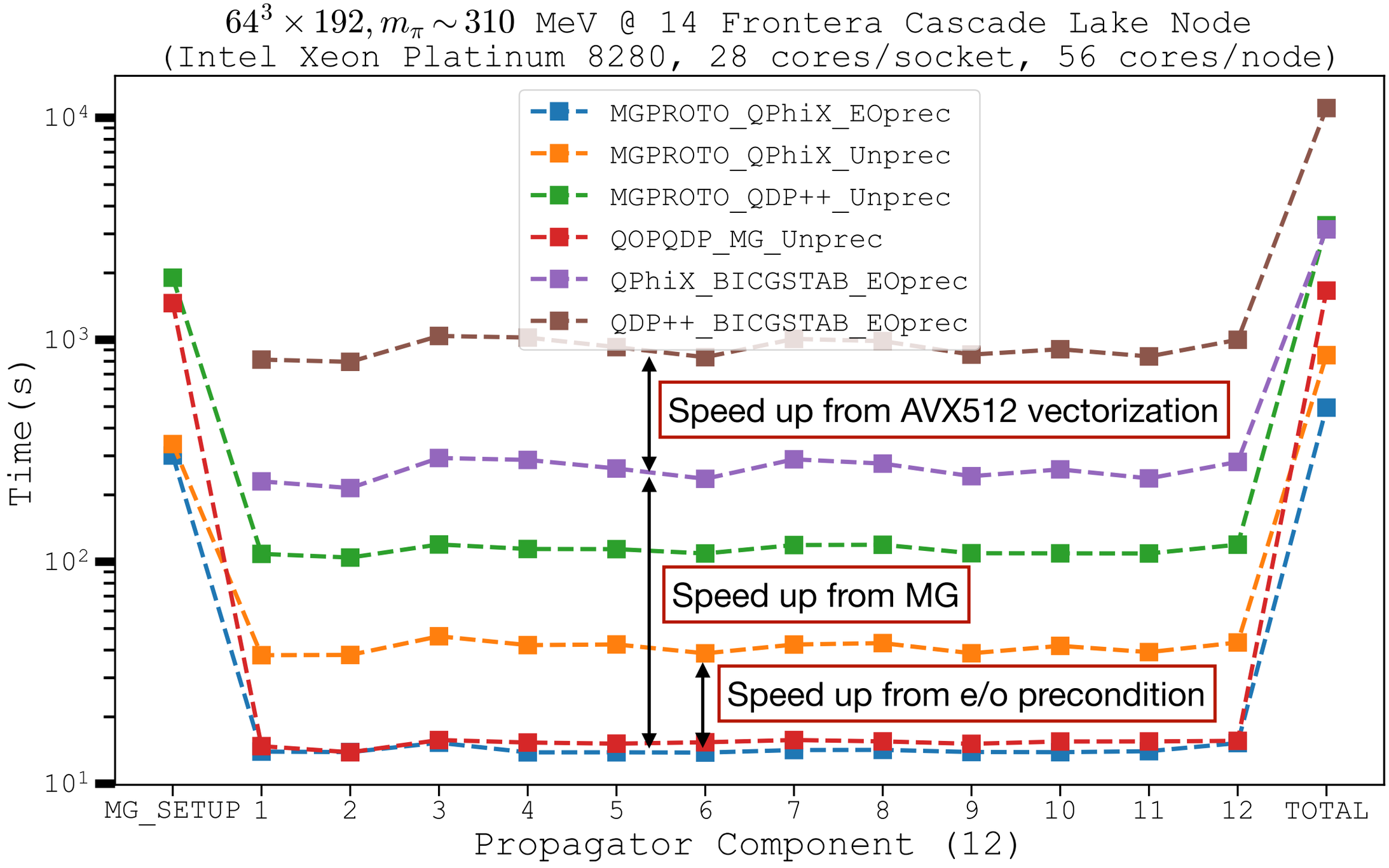
Brief Introduction to Multigrid Method

- Multigrid V-cycle (can be used as solver itself)



- More efficient choice: use MG as preconditioner of Krylov solver (e.g. GCR, FGMRES) and solve the coarse grid recursively (K-cycle)

MG Use Case: Quark Propagator



MG Use Case: Distillation Method

- Quark field smearing scheme
  - Factorize operators construction and propagator computation
  - Large operator set can be used to do the GEVP analysis
  - Compute disconnected diagram easily
- Large amount of $Dx = b$
  - On every configuration $N_t \times N_{spin} \times N_{vec}$
  - Very suitable for MG especially at light quark (reuse MG setup)

- step 1： Laplace eigenvectors V
- step 2： perambulator

$$\tau_{\alpha\beta}(t',t) = V^{\dagger}(t')\, D^{-1}_{\alpha\beta}(t',t)\, V(t)$$

- step 3： elemental

$$\Phi_{\alpha\beta}(t) = V^{\dagger}(t)\left[\Gamma(t)\right]_{\alpha\beta} V(t)$$

- step 4： correlation functions

$$C^{(2)}_{M}(t',t) = \left\langle \bar{d}(t')\,\Gamma^{B}(t')\,u(t')\cdot\bar{u}(t)\Gamma^{A}(t)d(t)\right\rangle$$
$$= \mathrm{Tr}\left[\Phi^{B}(t')\,\tau(t',t)\,\Phi^{A}(t)\tau(t,t')\right]$$
$$+\mathrm{Tr}\left[\Phi^{A}(t)\tau(t,t)\right]\mathrm{Tr}\left[\Phi^{B}(t')\,\tau^{\dagger}(t',t')\right]$$

- $N_f = 2,\ 16^3 \times 128,\ m_\pi = 350\ \mathrm{MeV}, N_{conf} = 7000,\ N_{vev} = 70$
  - 2 month * 100 V100 GPU on IHEP cluster, ~140 TB storage
  - On Huawei Kunpeng 920 (Arm), MG is 3.6x vs. BiCGStab [Wei sun, Yujiang Bi, ACAT 2021]

MG Use Case: HMC

- Most time consuming part in HMC $(D^\dagger D)^{-1}$
- Problems of near zero mode of Dirac operator
  - Slow to convergence -> solved by MG
  - Large fluctuation in fermionic force term

- Hasenbusch mass preconditioning: [M. Hasenbusch, Phys.Lett. B519 (2001) 177-182]
  - factorize fermion determinant

$$\det D = \frac{\det D(m_0)}{\det D(m_1)} \frac{\det D(m_1)}{\det D(m_2)} \ldots \det D(m_n)$$

force term can be fine tuned

easy to solve

- HMC + MG
  - Reuse / refresh MG setup after several MD integration
  - Multiple time scale MD integration scheme

MG Use Case: HMC

- Based on these algorithms, we have generated several ensembles
- $N_f = 2$ anisotropic clover (by 宫明)

Ensemble	$L^3 \times T$	β	a_s(fm)	ξ	$N_{\rm cfg}$	$m_{J/\psi}$(MeV)
I	$16^3 \times 128$	2.8	0.1026	5	~ 7000	2743(1)
II	$16^3 \times 128$	2.8	0.1026	5	~ 6000	3068(1)

Ensemble	$L^3 \times T$	β	a_s(fm)	ξ	$N_{\rm cfg}$	m_π(MeV)
I	$16^3 \times 128$	2.0	0.1517	5	~ 7000	350

- $N_f = 2 + 1$ isotropic clover (by 孙鹏)

name	$V = L^3 \times T$	Lattice spacing a	β	m_π	m_{η_s}	L	$m_\pi L$	Trajectories
C11P29S	24^3x72	0.108fm	6.20	290MeV	640MeV	2.6fm	3.8	13000
C11P29M	32^3x64	0.108fm	6.20	290MeV	640MeV	3.5fm	5.0	10000
C11P22M	32^3x64	0.108fm	6.20	220MeV	640MeV	3.5fm	3.9	10000
C11P22L	48^3x96	0.105fm	6.20	220MeV	640MeV	5.4fm	5.6	1000
C11P12L	48^3x96	0.105fm	6.20	120MeV	700MeV	5.4fm	3.1	1000
C11P15L	48^3x96	0.105fm	6.20	145MeV	700MeV	5.4fm	3.7	400
C11P14L	48^3x96	0.105fm	6.20	135MeV	700MeV	5.4fm	3.4	producing
C08P30S	32^3x96	0.08fm	6.41	300MeV	650MeV	2.6fm	3.9	11000
C08P22M	48^3x96	0.08fm	6.41	210MeV	650MeV	3.8fm	4.1	1000
C06P36S	48^3x144	0.055fm	6.72	360MeV	670MeV	2.6fm	4.8	650
C06P30S	48^3x144	0.055fm	6.72	300MeV	650MeV	2.6fm	4.0	producing

Twisted Hasenbusch Preconditioning

[Wei Sun, Balint Joo, Kate Clark, ECP report]

- Problems of near zero mode of Dirac operator
  - Slow to convergence -> solved by multigrid
  - Large fluctuation in fermionic force term
- Mass ratio method: directly set the bare quark mass
  - Easy to implement, but still with near zero mode

$$\det D = \frac{\det D(m_0)}{\det D(m_1)} \frac{\det D(m_1)}{\det D(m_2)} \ldots \det D(m_n)$$

- Twist ratio method: add a twist μ to Wilson/Clover Dirac operator D
- $M(\mu) = \hat{D}_s + i\mu\gamma_5 A_{oo} = 1 - A_{oo}^{-1} D_{oe} A_{ee}^{-1} D_{eo} + i\mu\gamma_5 A_{oo}$, $M(\mu = 0) = D$
- $M^\dagger(\mu) M(\mu) = \hat{D}_s{}^\dagger \hat{D}_s + \mu^2 A_{oo}^2$, eigenvalue bounded from below

$$\det D = \frac{\det M(0)}{\det M(\mu_1)} \frac{\det M(\mu_1)}{\det M(\mu_2)} \ldots \det M(\mu_n)$$

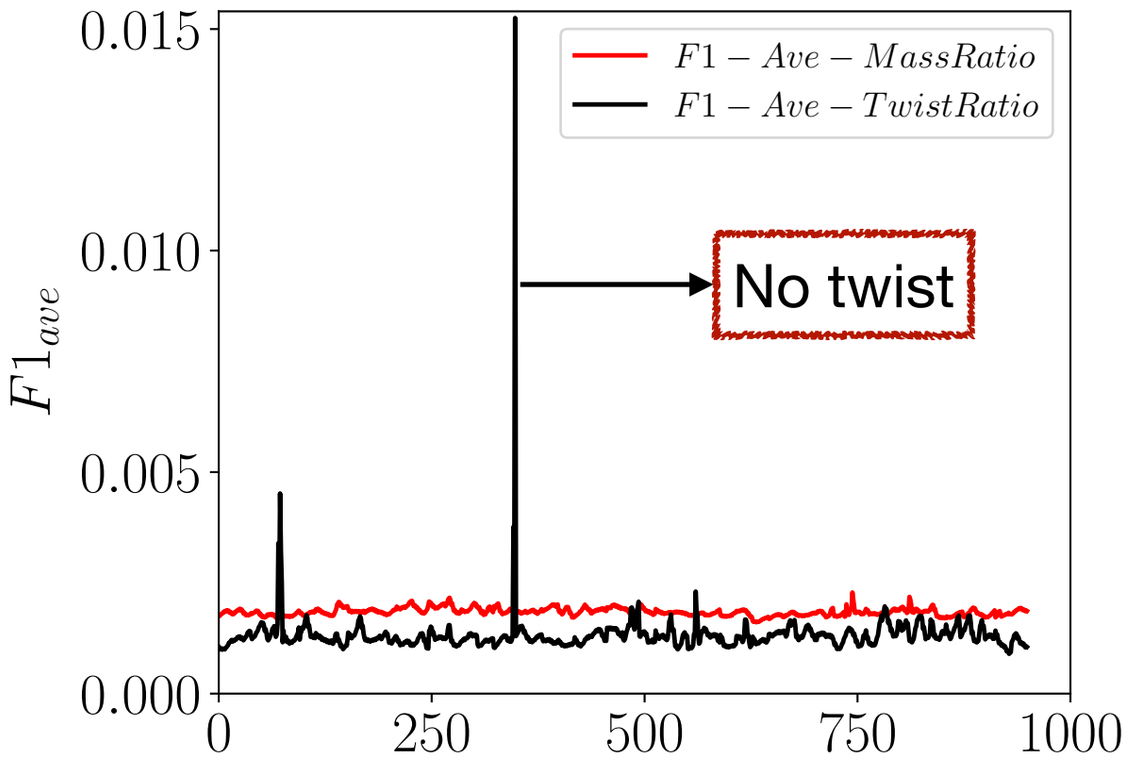
- Other approaches: twisted mass reweighting and exponential clover by M. Luscher [Comput.Phys.Commun. 184 (2013), Comput.Phys.Commun. 255 (2020)]

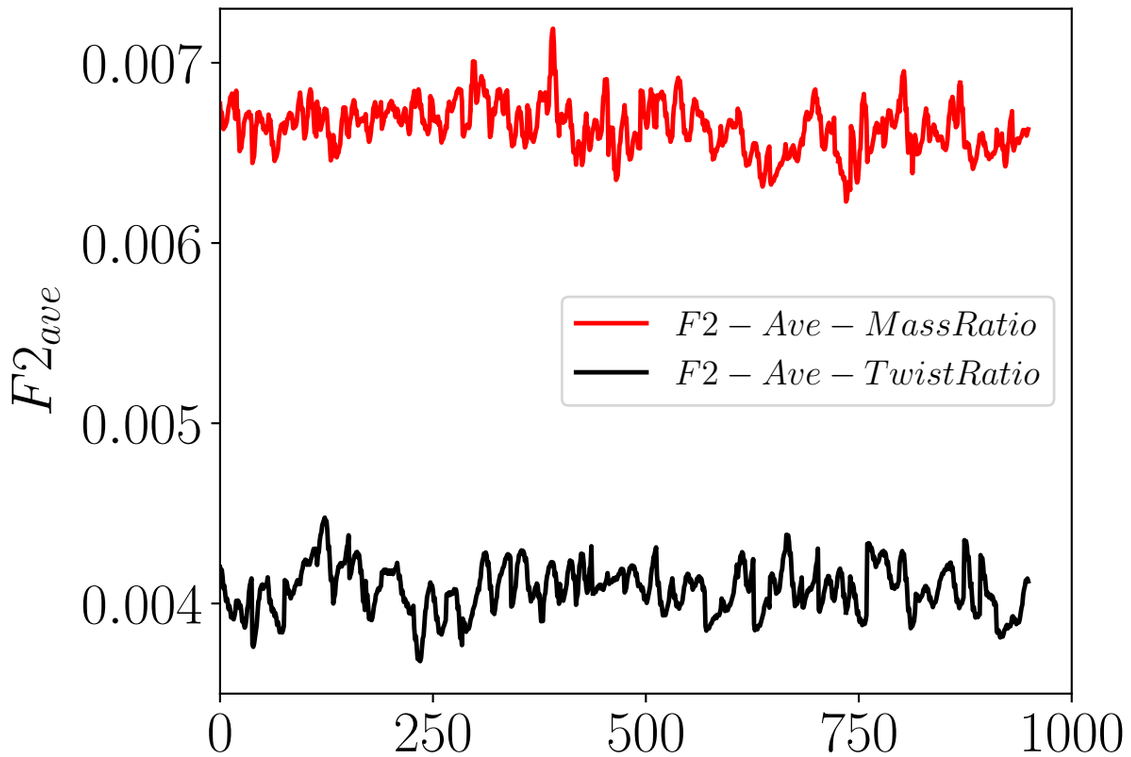
Twisted Hasenbusch Preconditioning: test

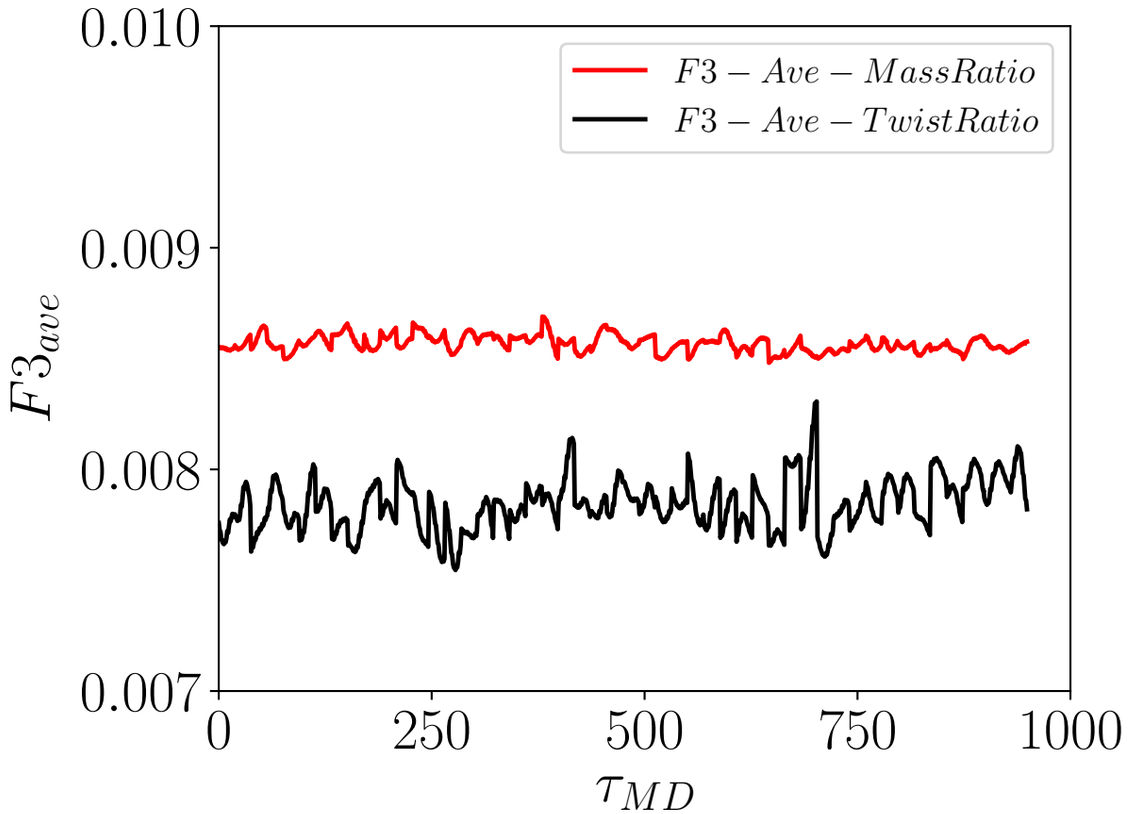
- Clover, $N_f = 2 + 1,\ 64^3 \times 128,\ m_\pi \approx 170\ MeV,\ \beta = 6.3$
- 512 Summit V100 GPUs
- 50 trajectories
- Multiple time scale force gradient MD integrator, $\tau_0 = 0.707$
- MG solver relative residue: MD (1e-11), Accept/Reject(1e-12)
- Mass-Ratio parameter:
  - $\frac{[-0.2416]}{[-0.2400]} \times \frac{[-0.2400]}{[-0.2320]} \times \frac{[-0.2320]}{[-0.2180]} \times \frac{[-0.2180]}{[-0.1870]} \times [-0.1870]$

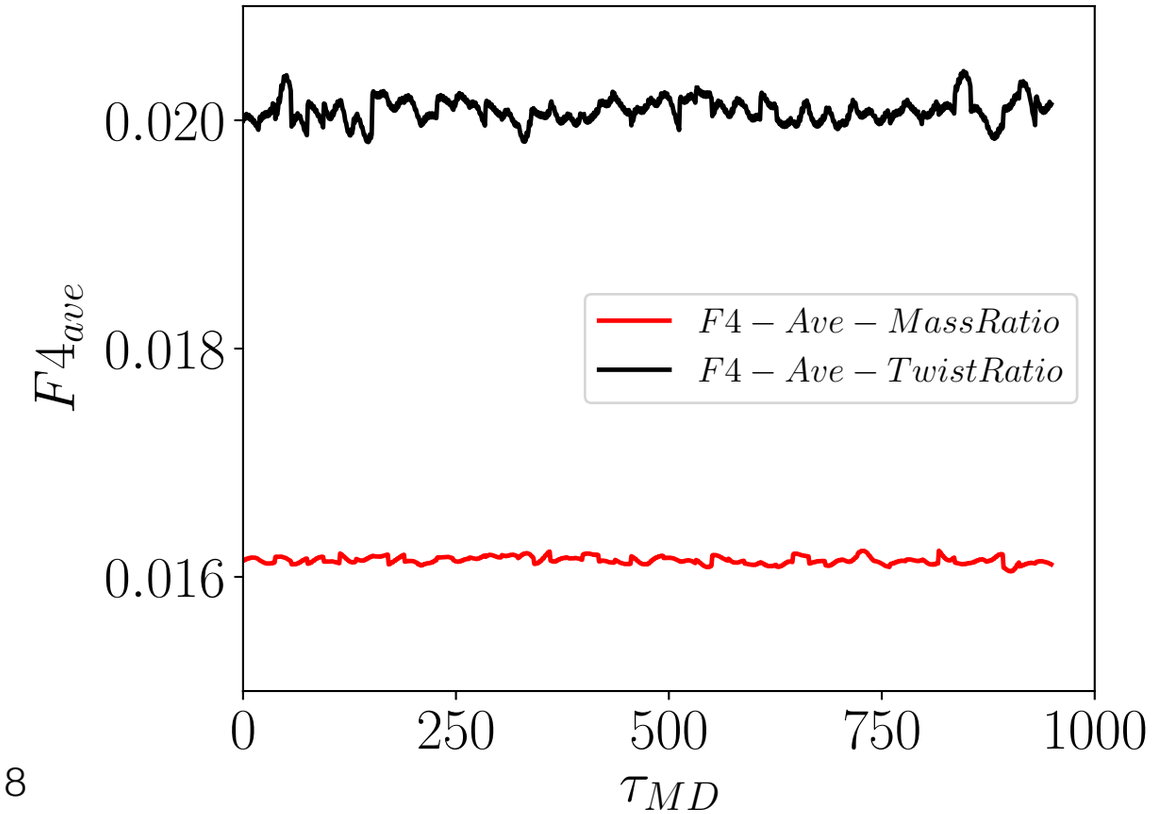
  $F1$ $F2$ $F3$ $F4$ $Fcancel$
- Twist-Ratio parameter:
  - $\frac{[0]}{[0.00035]} \times \frac{[0.00035]}{[0.0015]} \times \frac{[0.0015]}{[0.005]} \times \frac{[0.005]}{[0.0175]} \times [0.0175]$

Average Fermionic Force — Ratio Term F_{ave}

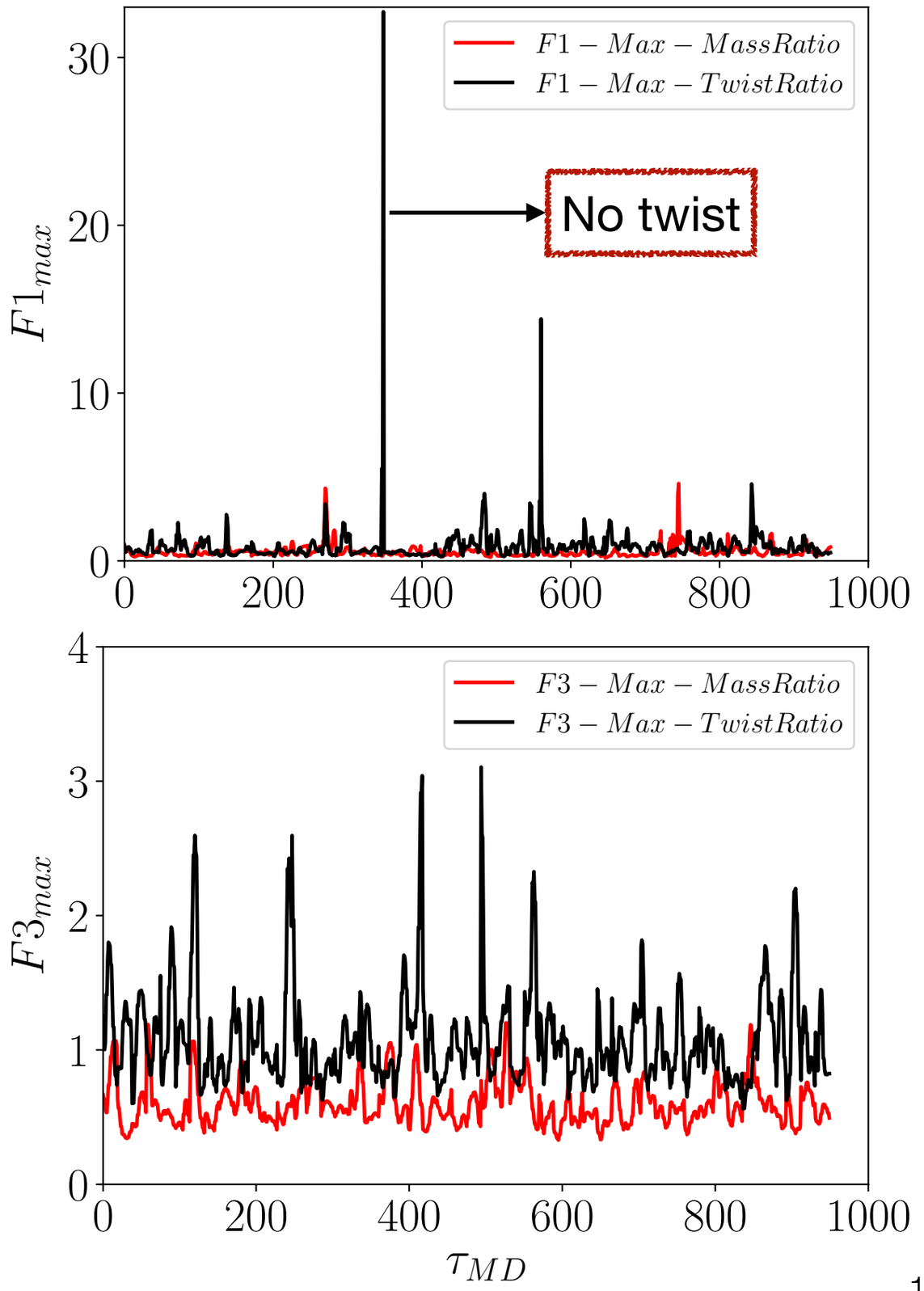


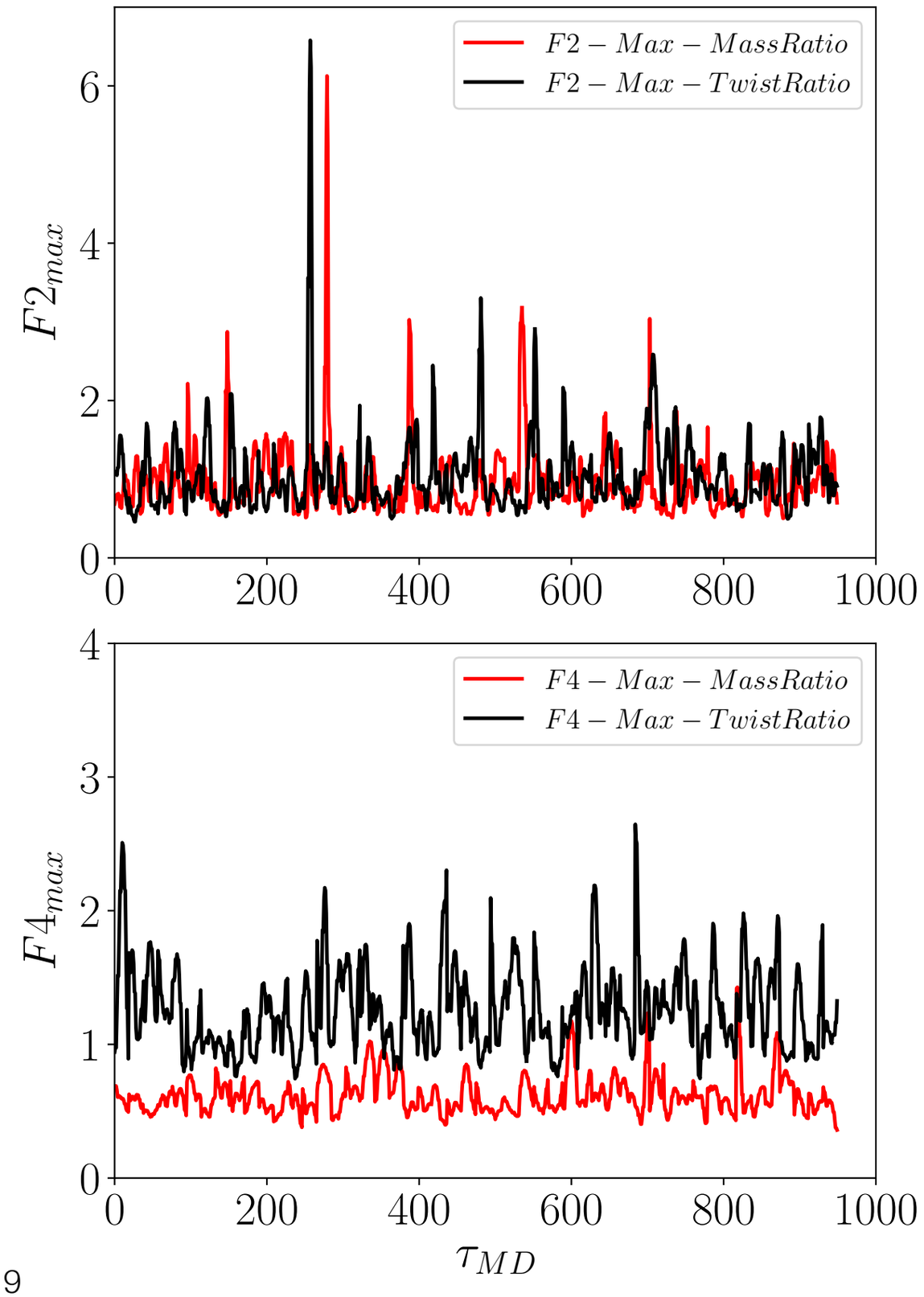




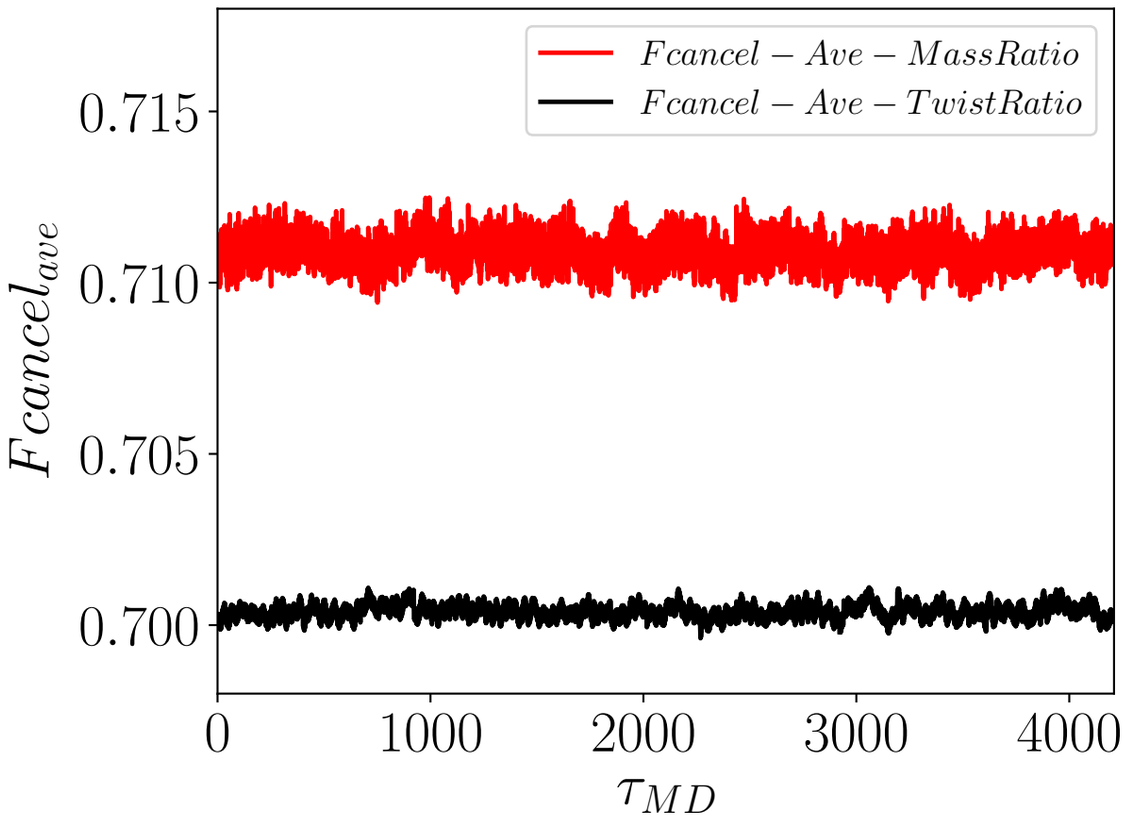


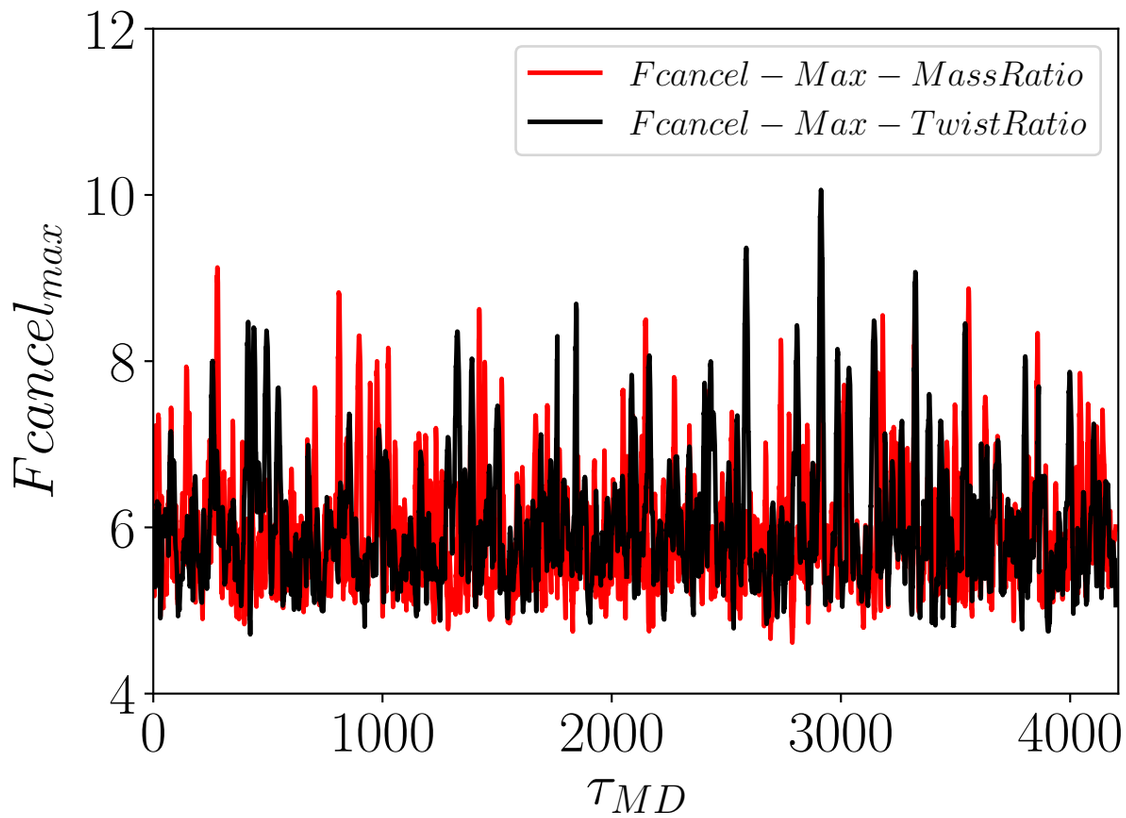
Maximum Fermionic Force — Ratio Term F_{max}



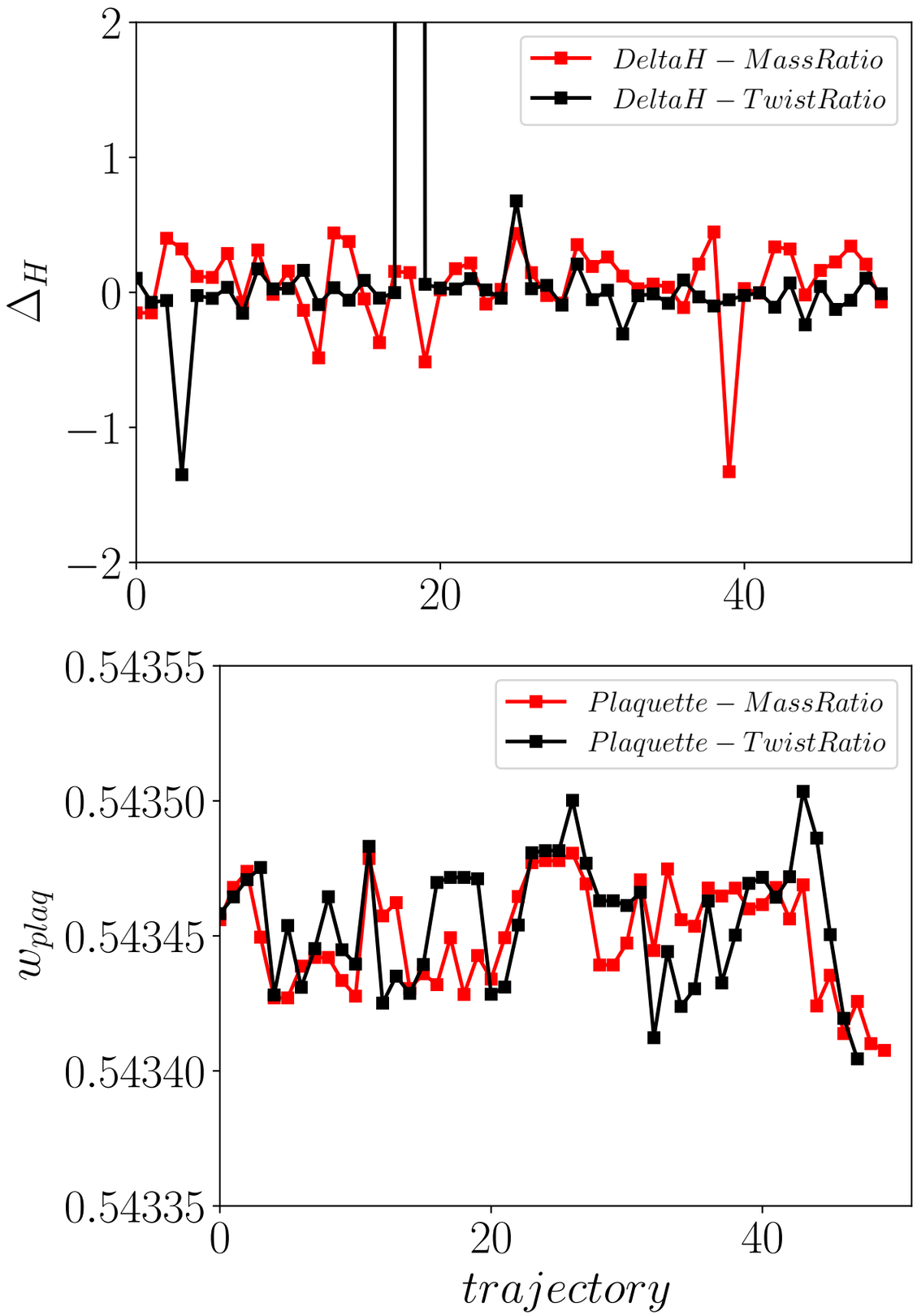


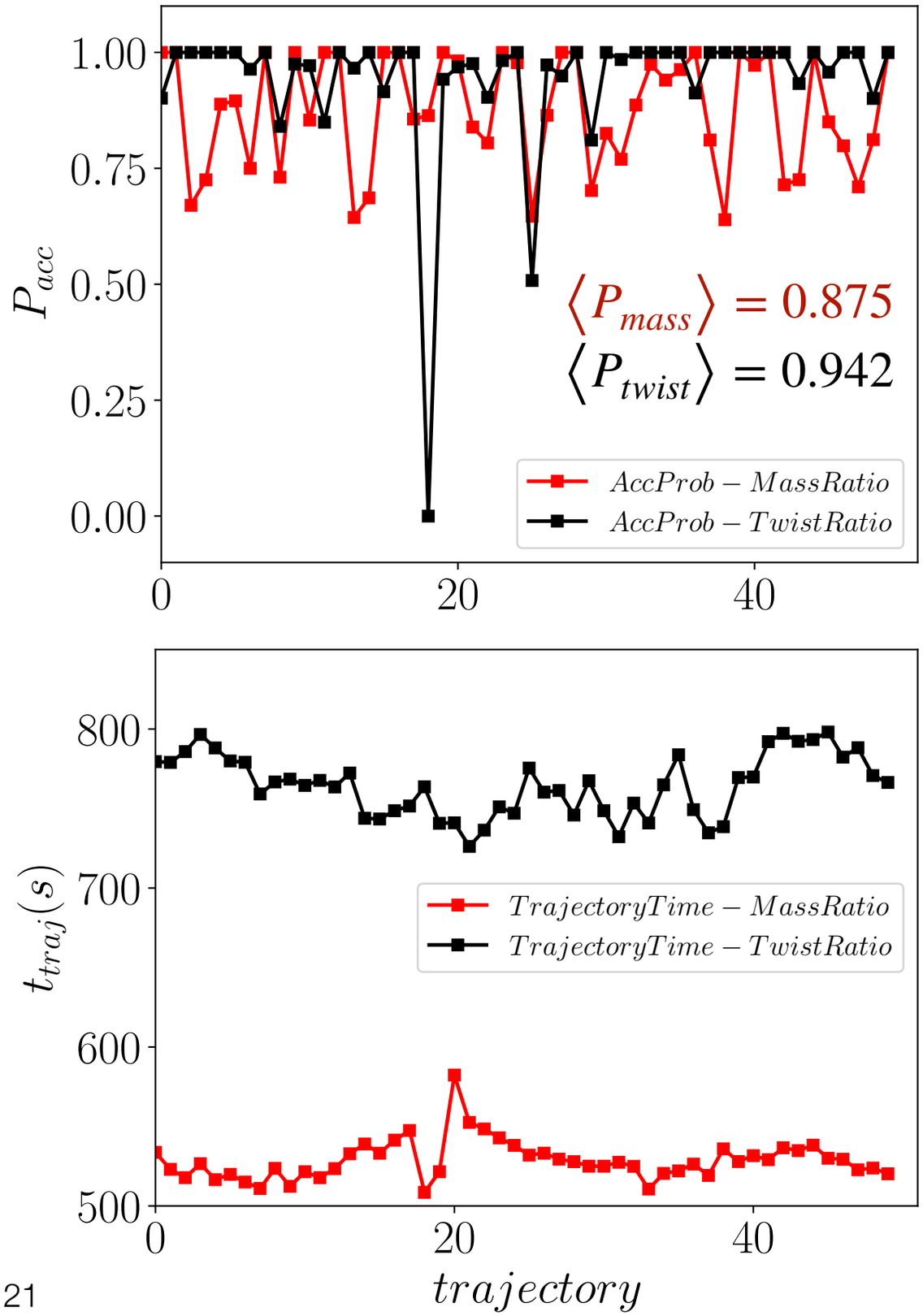
Fermionic Force — Cancel Term *Fcancel*





MD Integration Stability



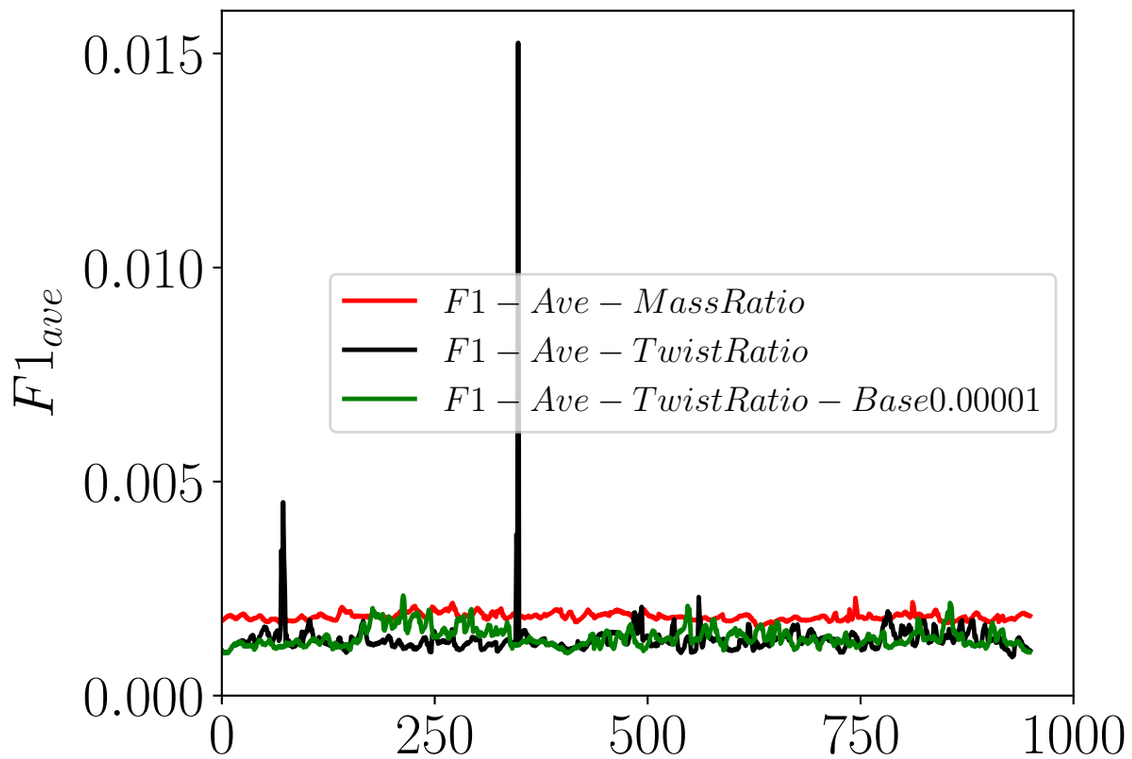


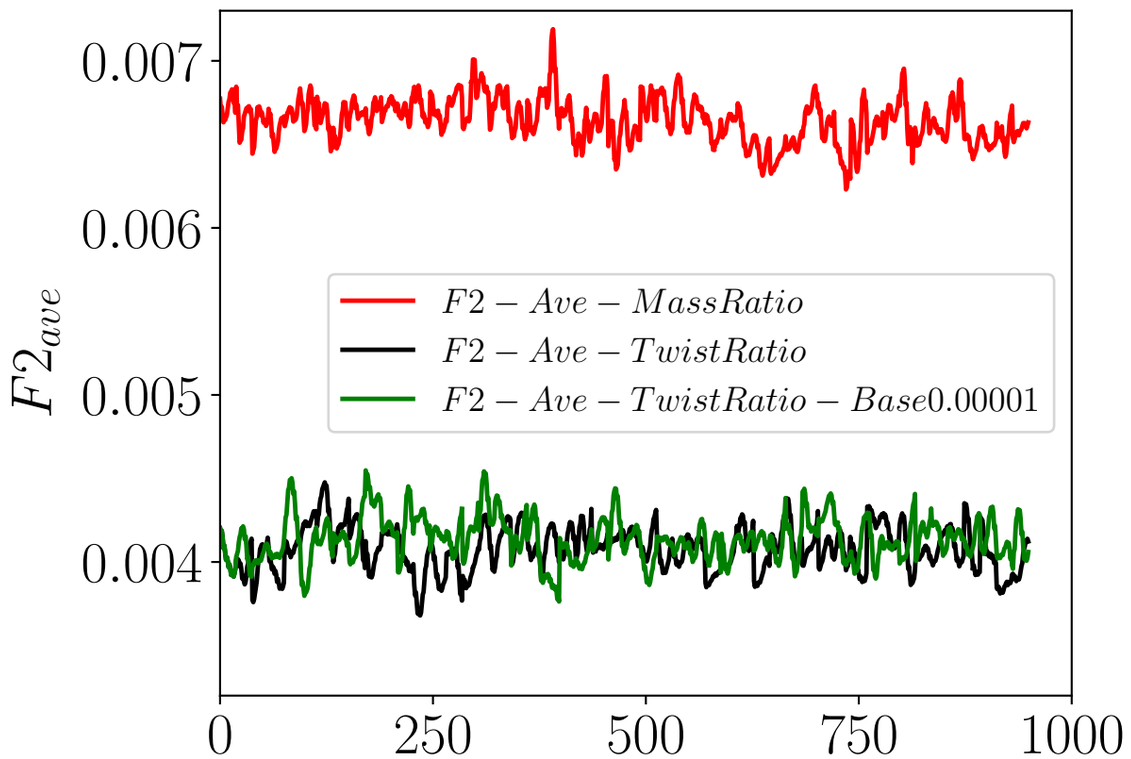
Problems & Improvement

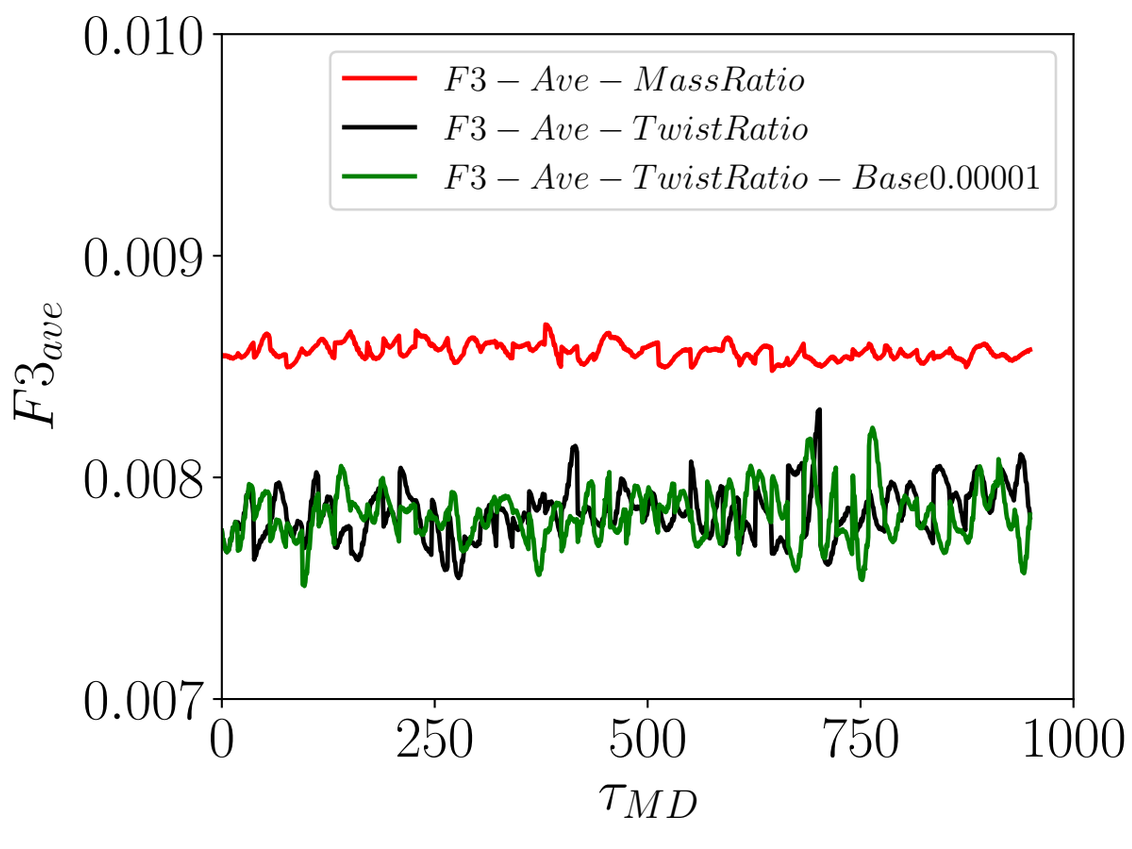
- No twist μ for the first ratio term ($F1$)
- Near zero mode -> large $F1$ -> MD integration unstable
- Add a small twist to the first ratio term in **MD integration**
  - $$\frac{[0]}{[0.00035]} \times \frac{[0.00035]}{[0.0015]} \times \frac{[0.0015]}{[0.005]} \times \frac{[0.005]}{[0.0175]} \times [0.0175]$$

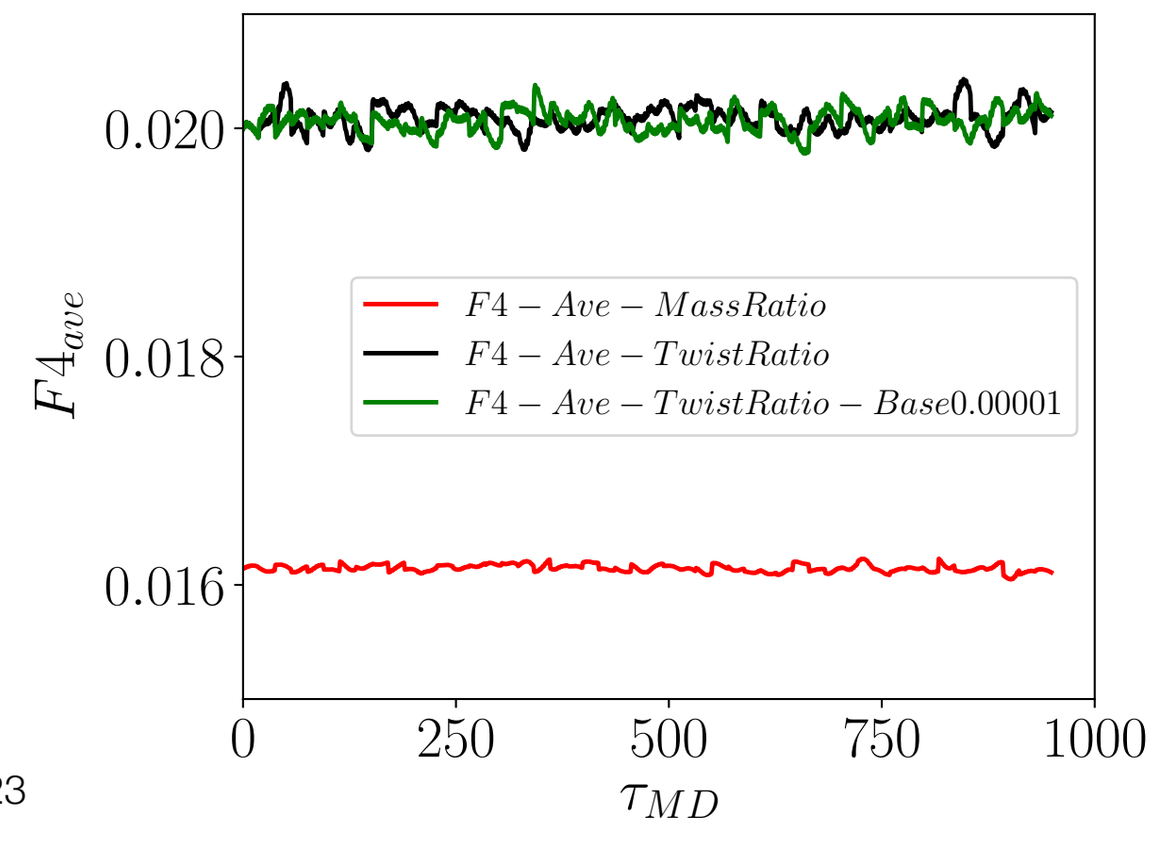
  - $$\frac{[0.00001]}{[0.00035]} \times \frac{[0.00035]}{[0.0015]} \times \frac{[0.0015]}{[0.005]} \times \frac{[0.005]}{[0.0175]} \times [0.0175]$$
- No twist ($\mu = 0$) in the Accept/Reject (energy calculation)
- No twist in Dirac operator in **action** -> correct distribution

Average Fermionic Force — Ratio Term F_{ave}

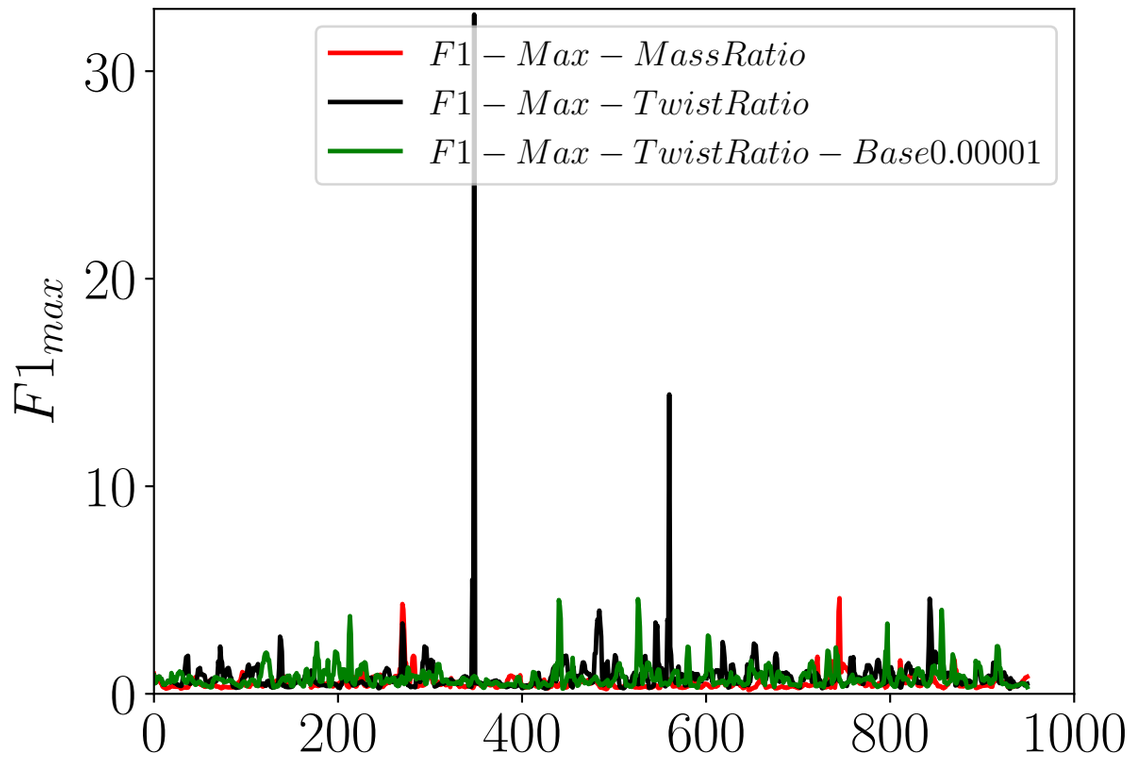


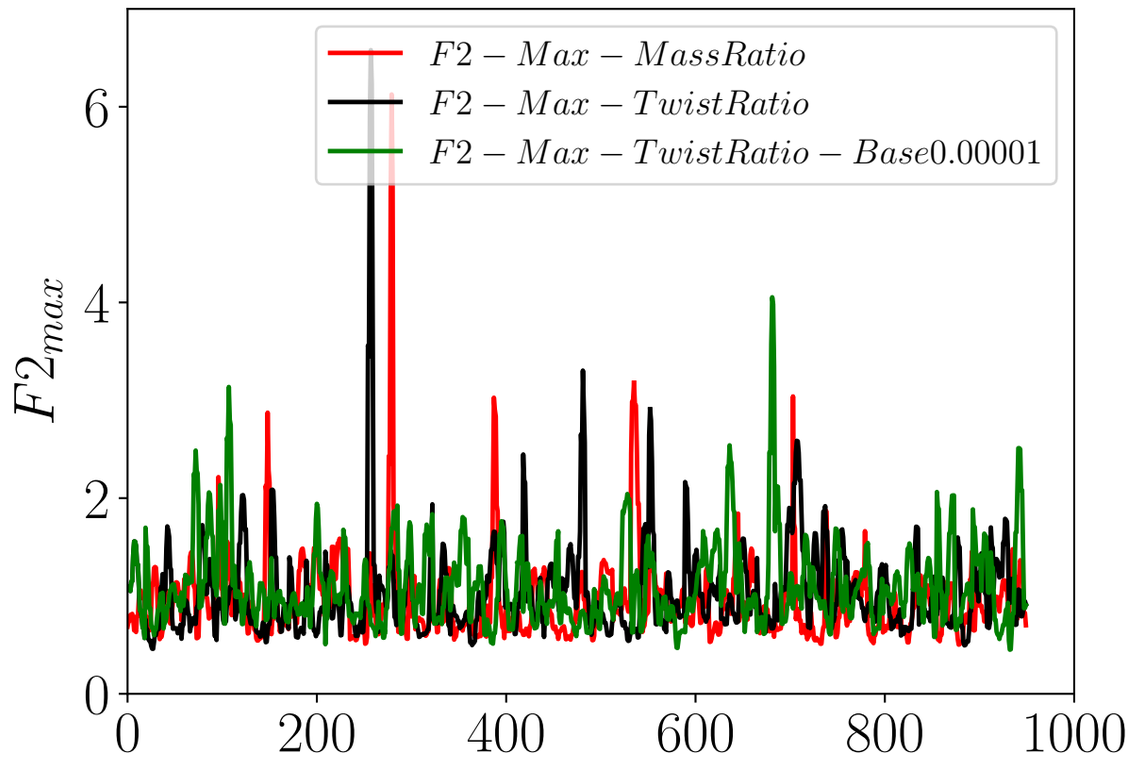


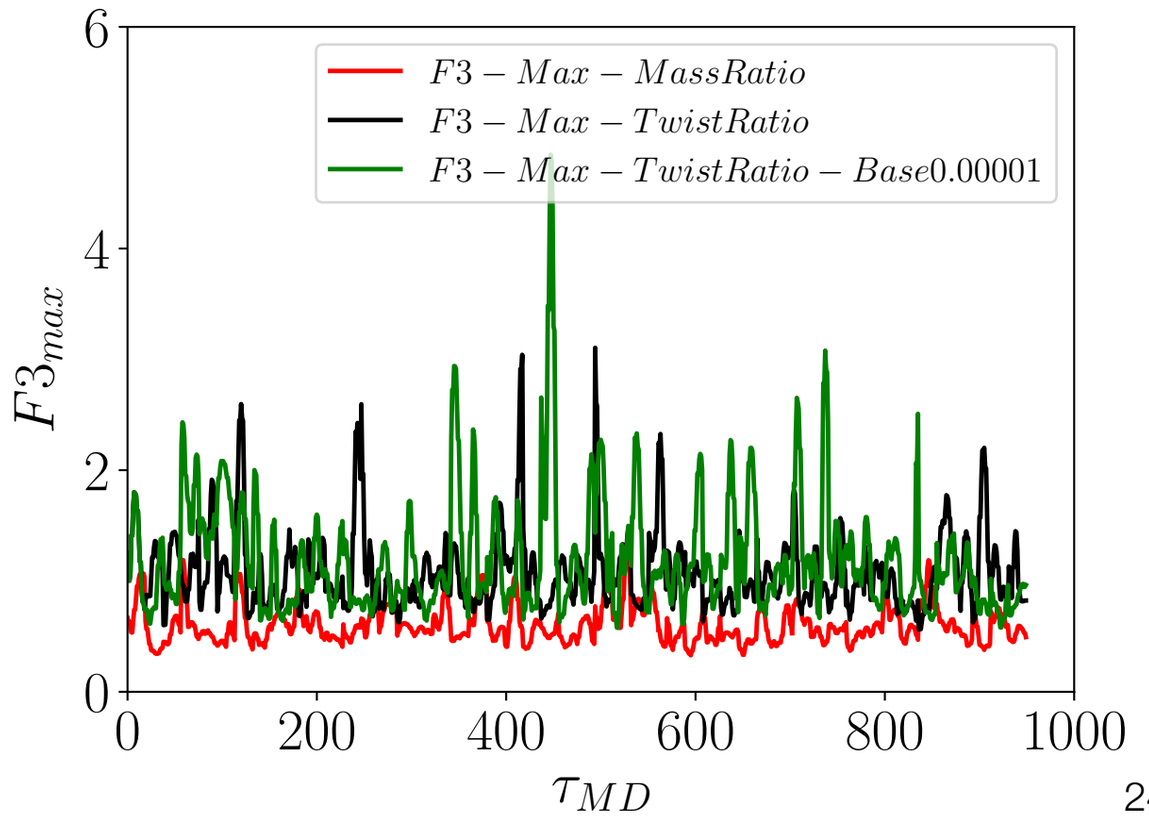


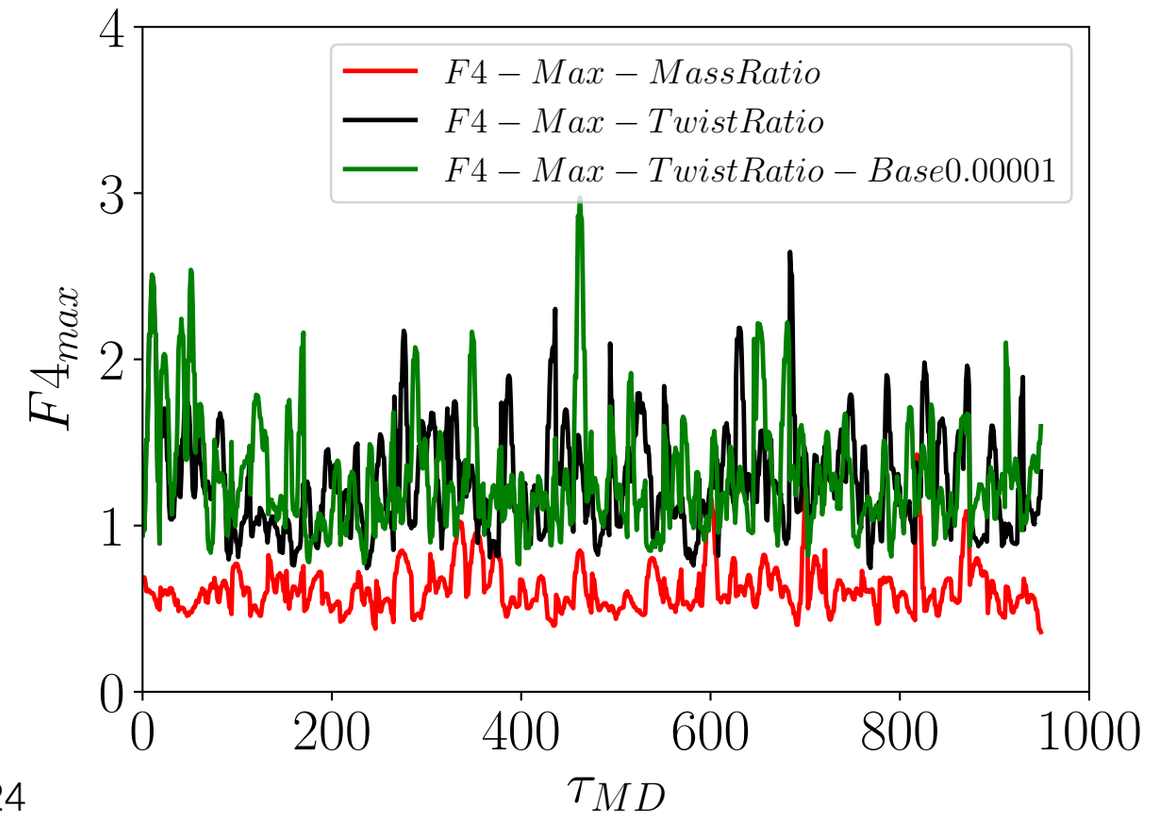


Maximum Fermionic Force — Ratio Term F_{max}

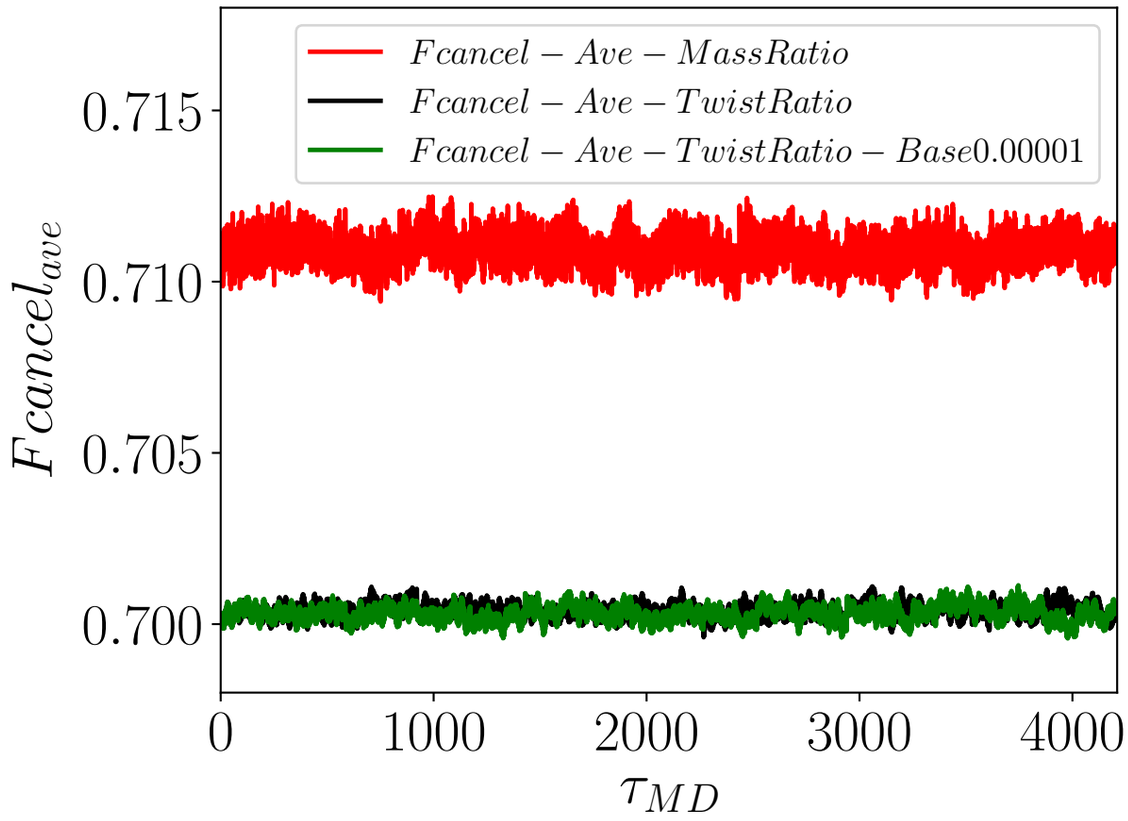


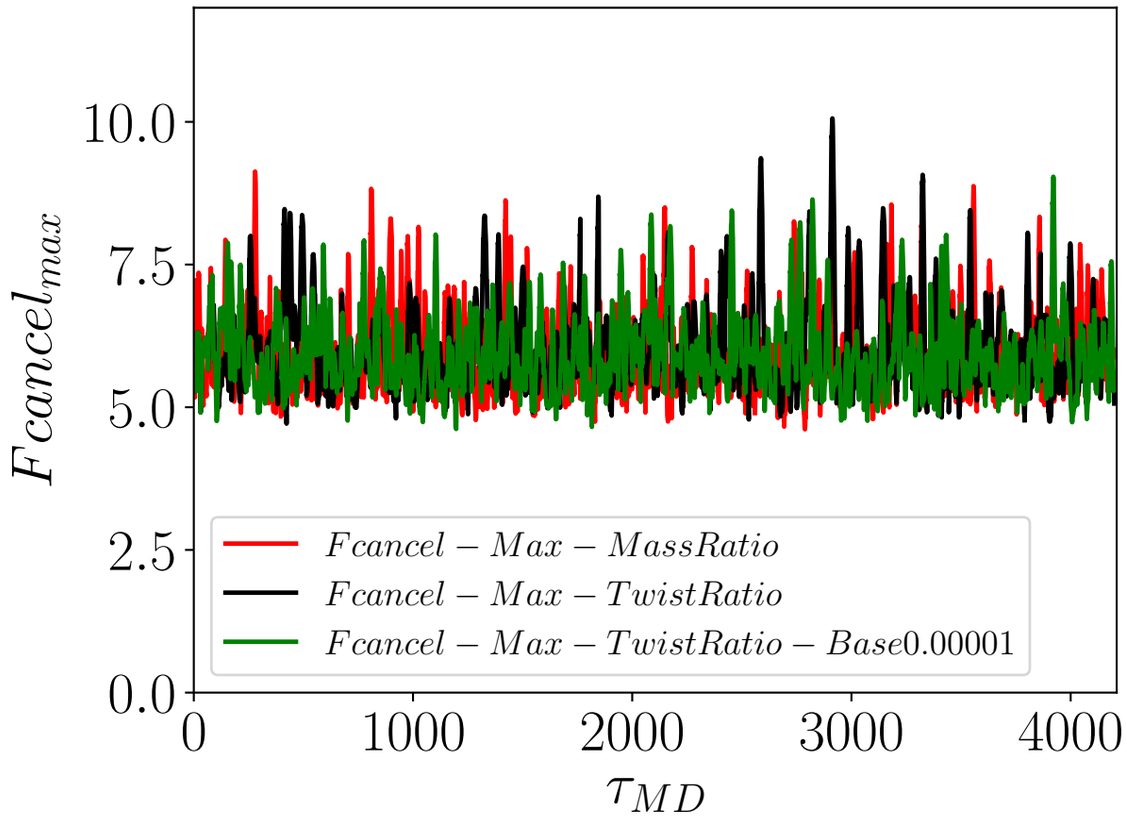




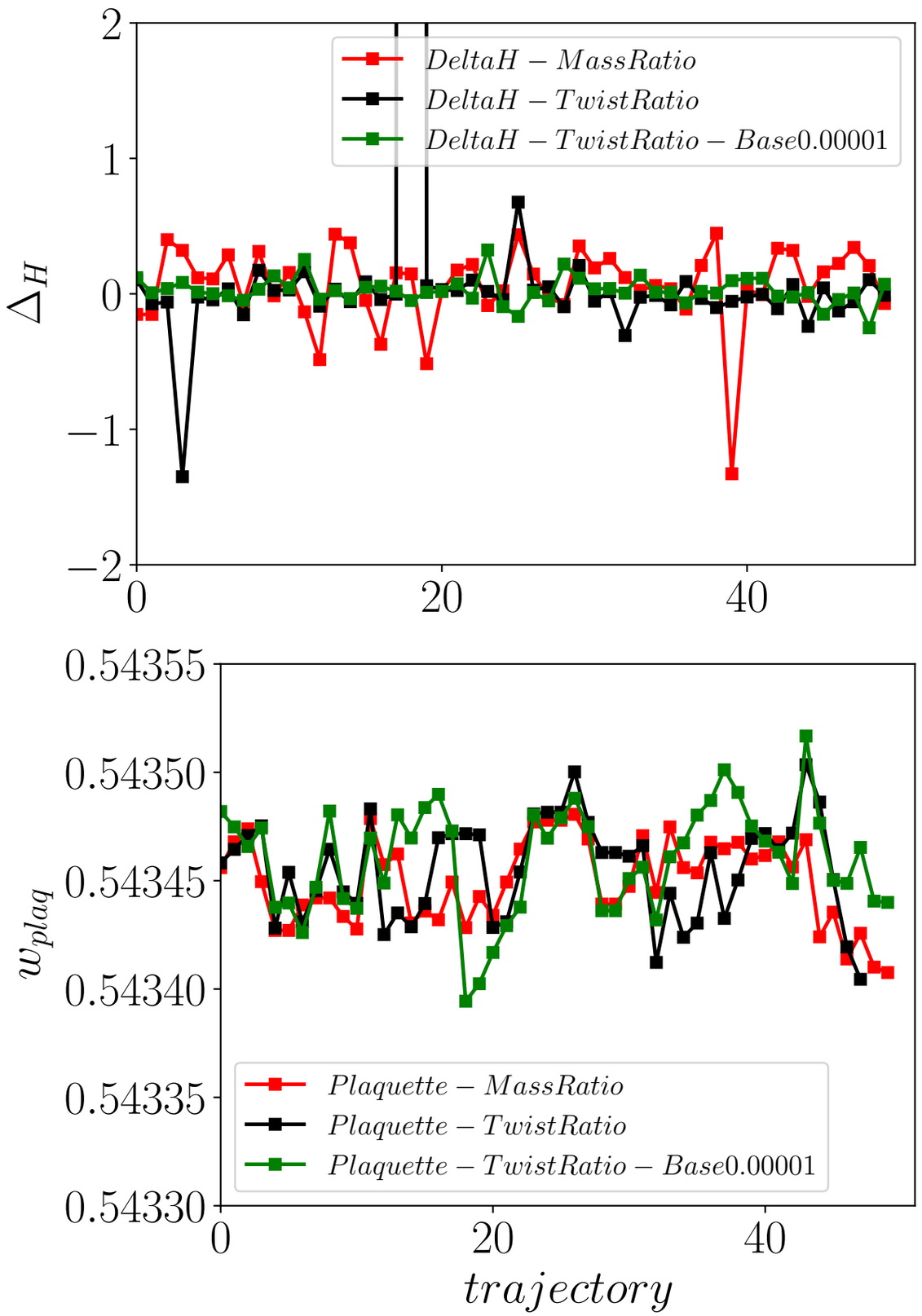


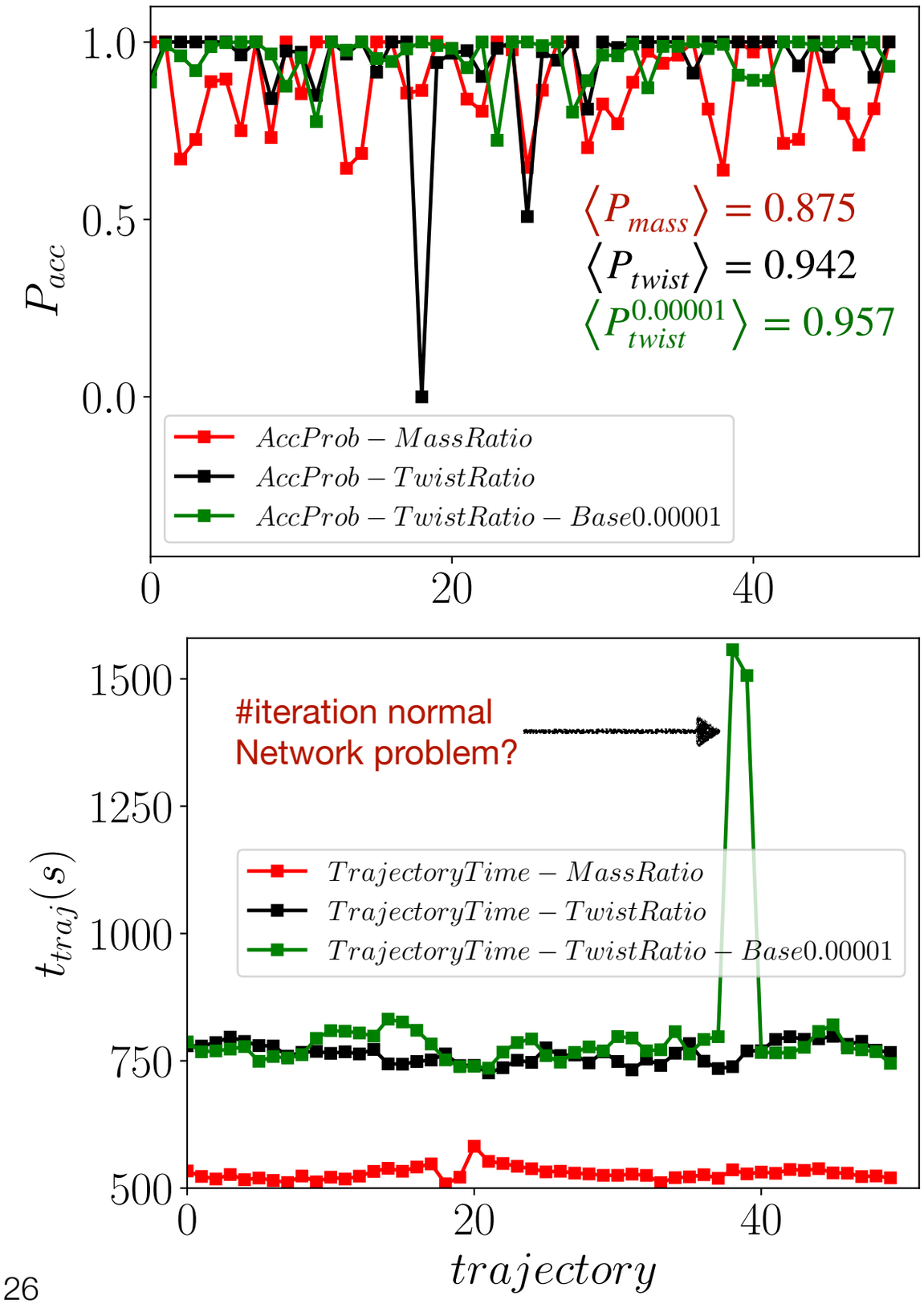
Fermionic Force — Cancel Term *Fcancel*





MD Integration Stability





Summary

- What we have
  - Asymmetric/Symmetric e/o preconditioning
  - Hasenbusch mass/twist ratio preconditioning
  - Multigrid for twisted operator (currently symmetric e/o)
  - Multiple time scale MD integrator
  - High order MD integrator (e.g. force gradient)
  - Adding twist in the MD integration can stabilize HMC
- Outlook
  - Strategies for parameters tunning
  - Optimize trajectory wall time
  - Coarse grid deflation
  - Multiple right-hand side solver