A simple method to optimize HMC performance

Andrea Bussone

27 July 2016

LATTICE 2016

CP3 Origins

Collaborators: V. Drach, M. Della Morte, C. Pica, A. Hietnanen, J. Rantaharju, M. R. L. Hansen

Plan of the talk

Introduction

Motivations & Set-up

HMC algorithms have a large number of parameters

- Integrators: Omelyan $\alpha, \ldots$
- Mass preconditioning: Hasenbusch mass $\mu, \ldots$
- Multi-time scale integration: $n, m, k, \ldots$

Optimize them it is not an easy task!

[Urbach et al. Comput. Phys. Commun. 174 (2006) 87]

Remark: In BSM strongly interacting theories, changing fermion representation, forces hierarchy may be different from QCD

Model under investigation:

- **SU**(2) gauge group
- unimproved Wilson fermion in the fundamental representation
- $\beta = 2.2,\ m_0 = -0.72,\ m_{\rm cr} \simeq 0.75(1)$ **(no mass dependence study)**

Simulation cost and integrator choice

We define the Cost of a simulation

$$\text{Cost} = \frac{\#\text{MVM}}{P_{\text{acc}}}$$

#MVM is machine independent

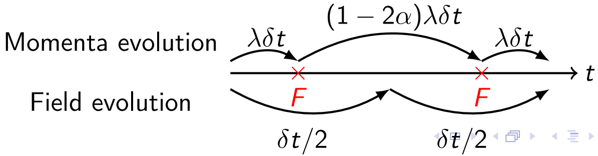
$P_{\text{acc}}(\underline{x}) \gtrsim 60 - 70\%$ Creutz formula
(neglect autocorrelation)

The goal is to minimize the cost! $\longrightarrow$ OPTIMIZATION PROBLEM

Omelyan integrator

$$\exp(\alpha\delta\tau S)\overbrace{\exp\left(\frac{\delta\tau}{2}T\right)}^{\text{field evolution}}\exp((1-2\alpha)\delta\tau S)\exp\left(\frac{\delta\tau}{2}T\right)\overbrace{\exp(\alpha\delta\tau S)}^{\text{momenta evolution}}$$

[Omelyan et al. Comput. Phys. Commun. 151 (2003) 272–314]



Basic definitions

Shadow Hamiltonian

For any symplectic integrator it exists an exactly conserved Hamiltonian

[Kennedy et al. Phys. Rev. D 87 (2013) no.3]

Shadow Hamiltonian is found through BCH asymptotic expansion in $\delta\tau$

$$\widetilde{H} = H + \delta\tau^2 \left(\frac{6\alpha^2 - 6\alpha + 1}{12} \underbrace{[S,[S,T]]}_{\text{Poisson bracket}} + \frac{6\alpha - 1}{24} [T,[T,S]] \right)$$

By setting $\alpha = 1/6$ and $T = \pi^2/2$ we have

$$\delta H \equiv \frac{\delta\tau^2}{72} [S,[S,T]] \equiv \frac{\delta\tau^2}{72} |\mathcal{F}|^2 = \frac{\delta\tau^2}{72} \sum_{x,\mu,a} T_R (F_\mu^a(x))^2$$

Poisson Bracket measurement

Integrator set-up

Mass preconditioning definitions

$$D = \text{massless Dirac operator},\ D_m = D + m,\ Q = \gamma_5 D_m$$

We split the fermion determinant for a doublet as

$$|\det Q|^2 = \det(DD^\dagger + \mu^2)\det\left(\frac{Q^2}{DD^\dagger + \mu^2}\right)$$

We introduce two kind of pseudofermions with action

$$S[\phi_1, \phi_2; U] = \underbrace{\phi_1^\dagger (DD^\dagger + \mu^2)^{-1}\phi_1}_{\text{HMC}} + \underbrace{\phi_2^\dagger \left(\frac{Q^2}{DD^\dagger + \mu^2}\right)^{-1}\phi_2}_{\text{Hasenbusch}}$$

Integrator set-up

Benchmark: Scaling behaviour

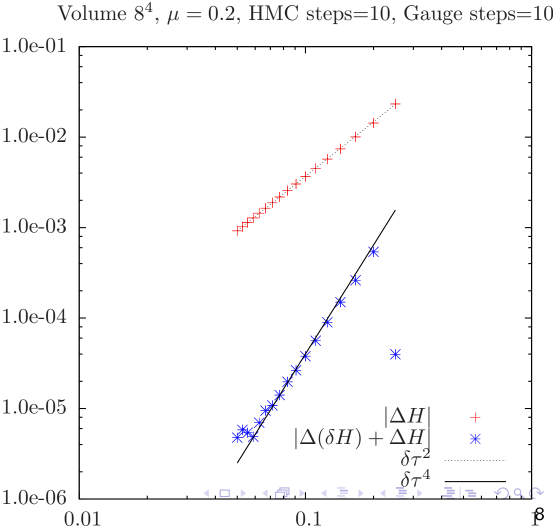
We choose 2nd order Omelyan integrator with $\alpha = 1/6$ and three-time scales

Levels

- 0: Hasenbusch $n = 4, 5, \ldots, 20$
- 1: HMC $m = 10$
- 2: Gauge $k = 10$

Remark: Hierarchy is related to magnitude and variances of the forces

We run one trajectory from a thermalized configuration



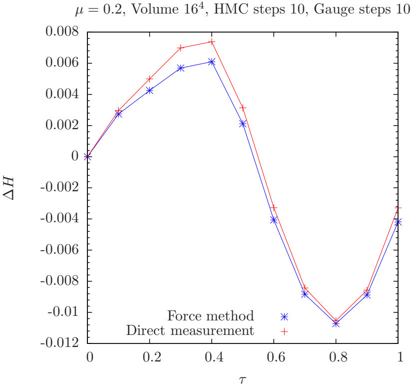
Integrator set-up

Benchmark: ΔH Approximation

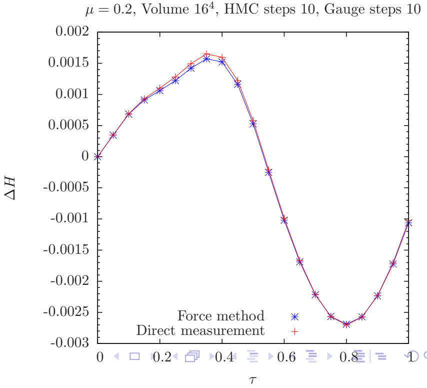
$$\widetilde{H} = H + \delta H \Longrightarrow \Delta H = -\Delta(\delta H) \propto |\text{Force}|^2$$

The minimum scales as predicted as the bound is saturated

$n = 10$



$n = 20$



Bulding the cost of a simulation

Cost of a simulation

Acceptance relation

[Gupta et al. Phys. Lett. B 242 (1990) 437]

$$P_{\mathrm{acc}}(\Delta H) = \mathrm{erfc}\left(\sqrt{\mathrm{Var}\left(\Delta H\right)/8}\right)$$

[Clark et al. PoS LATTICE 2008 (2008) 041]

$$\mathrm{Var}\left(\Delta H\right) \simeq 2\mathrm{Var}\left(\delta H\right)$$

$$P_{\mathrm{acc}}(\mathrm{Var}(\delta H)) = \mathrm{erfc}\left(\sqrt{\mathrm{Var}\left(\delta H\right)}/4\right)$$

2nd order Omelyan integrator with $\alpha = 1/6$ and three-time scale

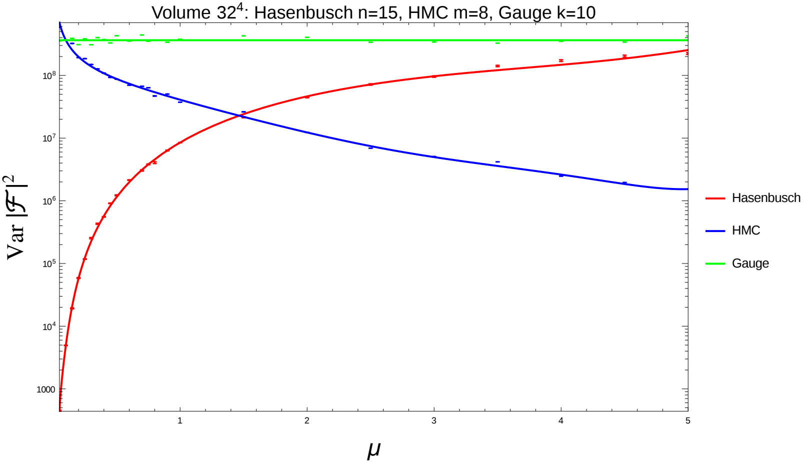
Neglecting covariances

$$\mathrm{Var}\left(\delta H\right) \simeq \frac{\delta\tau^4}{(72)^2}\left[\mathrm{Var}\left(|\mathcal{F}_1|^2\right)(\mu) + \frac{\mathrm{Var}\left(|\mathcal{F}_2|^2\right)(\mu)}{(4m^2)^2} + \frac{\mathrm{Var}\left(|\mathcal{F}_3|^2\right)(\mu)}{(16m^2k^2)^2}\right]$$

Idea: The n, m, k-dependence of the variances is known and they depend only on μ at a fixed mass m_0

Cost of a simulation

Variance fits



Strong dependence Flat dependence

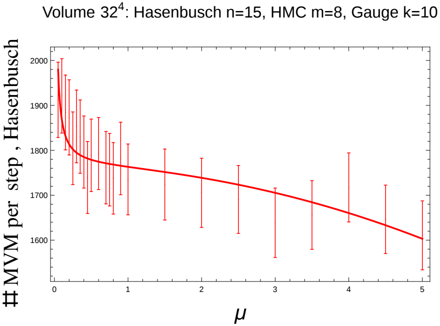
Remark: Hierarchy changes with the model you consider

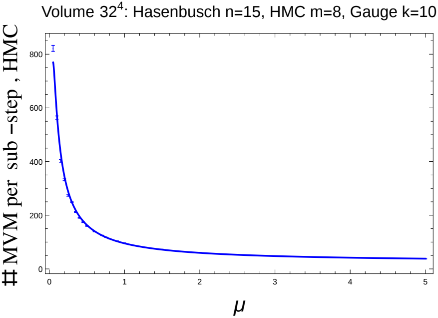
Cost of a simulation

Matrix Vector Multiplications fits

$$\#\ \mathrm{MVM} = (2n+1)\,\#\ \mathrm{MVM}_1(\mu) + 2n(2m+1)\,\#\ \mathrm{MVM}_2(\mu)$$

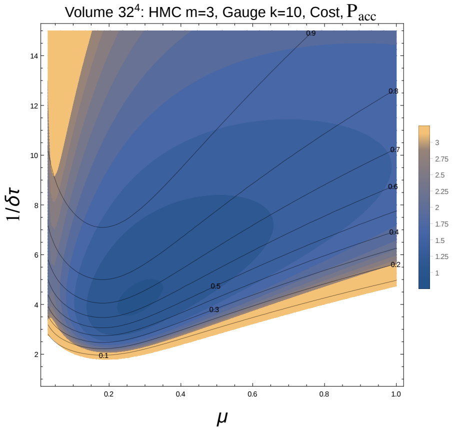
Idea: The n, m-dependence of the MVMs is known and they depend only on μ at a fixed mass m_0

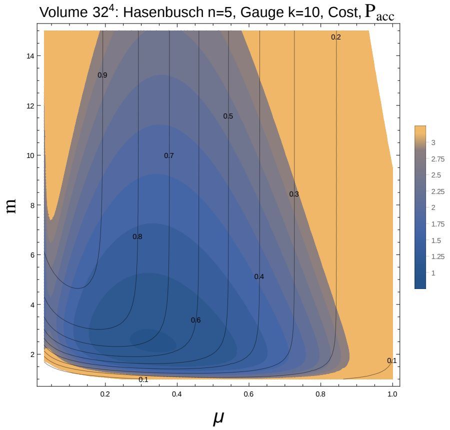




Cost of a simulation

Minimum cost





Normalized to the minimum that is found around

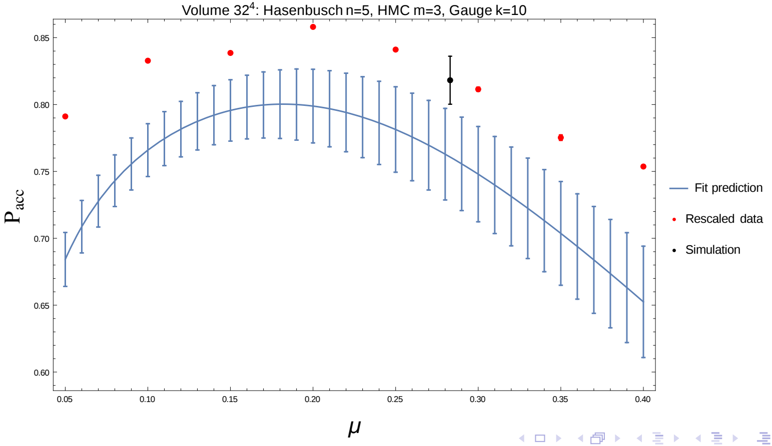
$(n, m, \mu) \simeq (5, 3, 0.3), \; k = 10$ fixed

Comparison with actual simulations

Cost of a simulation

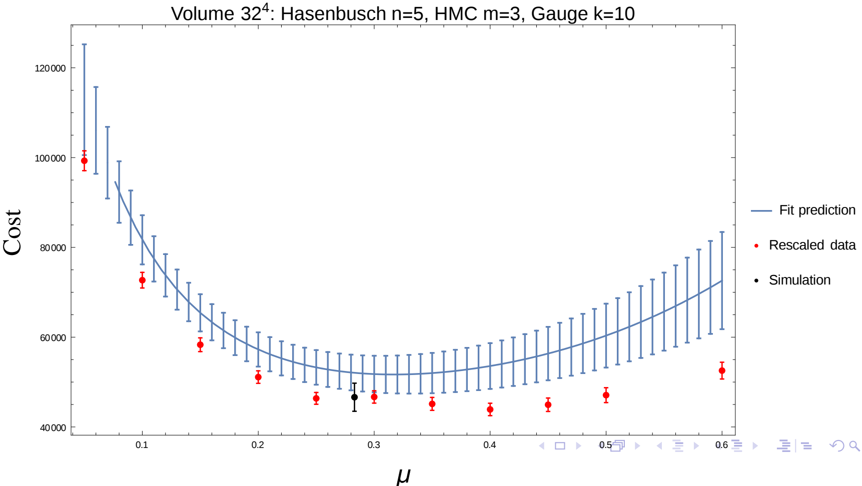
Comparison with actual simulations, $m_0 = -0.72$, 299 cnfgs

We consider the agreement within 10% as a check of all our assumptions



Cost of a simulation

Comparison with actual simulations, $m_0 = -0.72$, 299 cnfgs



Conclusions

Conclusions

Recipe

- Start with a *reasonable* choice for (n, m, k)
- Measure the *forces* in each level: $|\mathcal{F}_i|^2(\mu)$ and $\mathrm{Var}(|\mathcal{F}_i|^2)(\mu)$
- Measure the number of MVM in each level
- By fitting the dependence in μ we predict dependence of the *acceptance* and the *cost* upon (n, m, k, μ) within 10%

Future Goals

- Mass dependence study
- Apply this optimizing method to other strongly interacting models (BSM)
- Paper in preparation

Back-up slides

Number of configurations

$m_0 = -0.72$

μ	# confs
0.05	66
0.1	67
0.15	86
0.2	102
0.25	93
0.3	101
0.35	107
0.4	113
0.45	88
0.5	91

μ	# confs
0.6	88
0.7	103
0.75	106
0.8	108
0.9	111
1	1176
1.5	79
2	1479
2.5	87
3	91

Acceptance and cost of the simulation

$m_0 = -0.72$