A simple method to optimize HMC performance

A. Bussone[*a], **M. Della Morte**[a], **V. Drach**[b], **M. Hansen**[a], **A. Hietanen**[a], **J. Rantaharju**[a], **C. Pica**[a]

[a]*CP³-Origins, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*
[b]*CERN, Physics Department, 1211 Geneva 23, Switzerland*
 E-mail: bussone@cp3-origins.net

We present a practical strategy to optimize a set of Hybrid Monte Carlo parameters in simulations of QCD and QCD-like theories. We specialize to the case of mass-preconditioning, with multiple time-step Omelyan integrators. Starting from properties of the shadow Hamiltonian we show how the optimal setup for the integrator can be chosen once the forces and their variances are measured, assuming that those only depend on the mass-preconditioning parameter.

*Speaker.

1. Introduction and basic definitions

Modern HMC algorithms require tuning of multiple parameters for an efficient generation of configurations. The optimization of the parameters is a complicated task but has been achieved in multiple ways in QCD [1]. Here we present a general strategy, based on the existence of a shadow Hamiltonian, that can also be extended to strongly interacting BSM (Beyond Standard Model) theories.

In the following we specialize ourself to the case of the Omelyan integrator [2] with $\alpha = 1/6$. We consider an **SU**(2) gauge group with a doublet of unimproved Wilson fermions in the fundamental representation. For completeness the bare parameters of the simulation read $\beta = 2.2$ and $m_0 = -0.72$. For comparison the critical mass is estimated to be $m_{\rm cr} \simeq -0.77(1)$[1].

1.1 Shadow Hamiltonian

To each symplectic integrator corresponds an exactly conserved shadow Hamiltonian. In order to introduce it we begin by writing the evolution operator as

$$\exp\left(\tau \frac{\mathrm{d}}{\mathrm{d}t}\right) \star = \exp(\{H, \star\}) \equiv \exp(\tau D_H) \star.$$

where τ is a fictitious time parameter, $\{\cdot, \cdot\}$ the Poisson bracket, $H = T + S$ is the Hamiltonian of the system, T the kinetic part for the conjugate momenta, S the action we want to simulate, and $D_H \equiv D_T + D_S$ the associated operators. The Omelyan integrator is then given by the following evolution operator

$$\exp\left[\tau \alpha D_S\right] \exp\left[\frac{\tau}{2} D_T\right] \exp\left[\tau(1-\alpha) D_S\right] \exp\left[\frac{\tau}{2} D_T\right] \exp\left[\tau \alpha D_S\right],$$

with α being a free parameter. By using the Baker-Campbell-Haussdorff (BCH) formula one obtains that the conserved shadow Hamiltonian $\widetilde{H}$ is related to the target one H to be simulated, by

$$\widetilde{H} = H + \tau^2 \left\{ \frac{6\alpha^2 - 6\alpha + 1}{12} D_S[D_S(T)] + \frac{1 - 6\alpha}{24} D_T[D_S(T)] \right\} + \mathrm{O}(\tau^4).$$

By setting $\alpha = 1/6$ the second term vanishes and what remains is dependent on $D_S[D_S(T)] = \{S, \{S, T\}\}$, which as we will see in the following is directly related to the forces entering in the molecular dynamics simulations.

1.2 Mass preconditioning & multi time-scale

A way to reduce the fluctuations of the force is to employ *mass preconditioning* of the quark determinant [3]. The definitions for the massive and hermitian Dirac operators are the following

$$D_m = D + m, \quad Q = \gamma_5 D_m.$$

In the presence of mass preconditioning the probability distribution for the generation of configurations splits in three parts

$$P_S \propto \int \mathscr{D}[\phi_i, \phi_1^\dagger, \phi_2, \phi_2^\dagger] \exp\left(- \underbrace{S[U]}_{\text{Gauge}} - \underbrace{\phi_1^\dagger (DD^\dagger + \mu^2)^{-1} \phi_1}_{\text{HMC}} - \underbrace{\phi_2^\dagger \left(\frac{Q^2}{DD^\dagger + \mu^2} \right)^{-1} \phi_2}_{\text{Hasenbusch}} \right).$$

---

[1] We recall that theory is breaking chiral symmetry spontaneously and the current investigation show the theory is QCD-like.

Hence we are now dealing with three forces: Gauge, HMC, Hasenbusch. A further acceleration can be achieved by considering multiple time-step integrators [1], which consists of taking different integration step sizes for the different forces. We assume that in the outermost level there is the evolution for S_1 with time step $\delta\tau = \tau_f/n$, in the middle the integrator for S_2 with m steps and the innermost is for S_3 with k steps.

The shadow Hamiltonian associated to the Omelyan integrator with three time-scales and mass preconditioning is a quite lengthy expression but by setting the parameter $\alpha = 1/6$ it is given by

$$\widetilde{H} = H + \frac{\delta\tau^2}{72}\left[\{S_1,\{S_1,T\}\} + \frac{1}{4m^2}\{S_2,\{S_2,T\}\} + \frac{1}{16m^2k^2}\{S_3,\{S_3,T\}\}\right] + \mathrm{O}(\delta\tau^4).$$

We use the conventions adopted in [4, 5] and the above formula reduces to

$$\widetilde{H} = H + \frac{\delta\tau^2}{72}\sum_{x,\mu,a}\left[T_{R,1}\left(F_1^{a\mu}(x)\right)^2 + T_{R,2}\frac{\left(F_2^{a\mu}(x)\right)^2}{4m^2} + T_{R,3}\frac{\left(F_3^{a\mu}(x)\right)^2}{16m^2k^2}\right] + \mathrm{O}(\delta\tau^4) \qquad (1.1)$$

$$\equiv H + \frac{\delta\tau^2}{72}\sum_{x,\mu,a}\left(|\mathscr{F}_1|^2 + \frac{|\mathscr{F}_2|^2}{4m^2} + \frac{|\mathscr{F}_3|^2}{16m^2k^2}\right) + \mathrm{O}(\delta\tau^4) \equiv H + \delta H + \mathrm{O}(\delta\tau^4),$$

that is a function only of the forces used during the simulations. We immediately see that the shadow Hamiltonian is related to the different parts of the force weighted by the corresponding normalization for the generators. Already at this point one can see what drove our assignments for the different levels. We want to suppress the contribution of the *bigger* force (the gauge one) and hence that will go in the innermost level, followed by the HMC force and finally the Hasenbusch one at the outermost level.

2. Benchmarks in small volumes

In order to test the measurements of the Poisson brackets we check the scaling with $\delta\tau$ of $|\Delta H| \propto \delta\tau^2$ and $|\Delta(\delta H) + \Delta H| \propto \delta\tau^4$, indeed, since $\widetilde{H}$ is conserved along the trajectory, we have

$$\Delta\widetilde{H} = 0 = \Delta H + \Delta(\delta H) + \mathrm{O}(\delta\tau^4) \Longrightarrow \Delta H = -\Delta(\delta H) + \mathrm{O}(\delta\tau^4). \qquad (2.1)$$

Another test is to measure directly ΔH along the trajectory and compare it with the one built from the knowledge of the forces, eq. (2.1). We run one trajectory from a thermalized configuration with the following set-up for the levels of integration
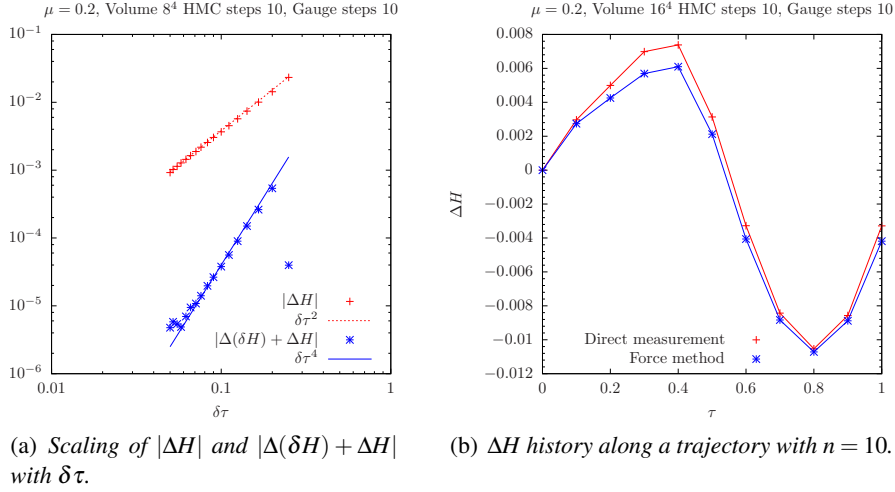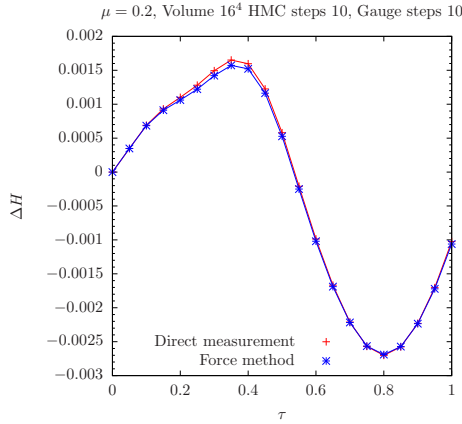
- level 0: Hasenbusch, $n = 4, 5, \ldots, 20$,
- level 1: HMC, $m = 10$,
- level 2: Gauge, $k = 10$.

The results for 8^4 and 16^4 volumes are showed in Fig. 1. It is worth to notice that the minimum of ΔH scales as predicted and when the minimum (or a maximum) is attained, then ΔH cannot grow due to the existence of the shadow Hamiltonian, and it is well understood in terms of the various underlying force contributions, see Figs. 1(b), 1(c).

3. Cost of a simulation and its minimization

Although the cost of a simulation is not unique we define it as

$$\mathrm{Cost} = \frac{\#\mathrm{MVM}}{P_{\mathrm{acc}}}.$$

$\mu = 0.2$, Volume 8^4 HMC steps 10, Gauge steps 10

(a) *Scaling of $|\Delta H|$ and $|\Delta(\delta H) + \Delta H|$ with $\delta\tau$.*

$\mu = 0.2$, Volume 16^4 HMC steps 10, Gauge steps 10

(b) *ΔH history along a trajectory with $n = 10$.*

$\mu = 0.2$, Volume 16^4 HMC steps 10, Gauge steps 10

(c) *ΔH history along a trajectory with $n = 20$.*

Figure 1: *Benchmarks for the Poisson brackets measurements.*

The number of Matrix-Vector-Multiplications (# MVM) is machine independent[2] . Furthermore we neglect the autocorrelation since that conceivably has a mild dependence on μ and therefore should be mostly contribute as an overall factor to the cost.

We link the acceptance P_{acc} to ΔH through the Creutz formula [6]

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

and the connection between the variances of ΔH and δH is given by [7]

$$\text{Var}(\Delta H) \simeq 2\text{Var}(\delta H).$$

We want to optimize the choice of parameters μ, n, m, k while keeping the integrator, the solver and the number of Hasenbusch splittings fixed. The variance of δH can be written by using eq. (1.1),

---

[2]The gauge part contributes for a maximum of $\sim 5\%$ of the cost, hence is negligible. We also took into account the gauge part as a check and it does not affect the results of this work.
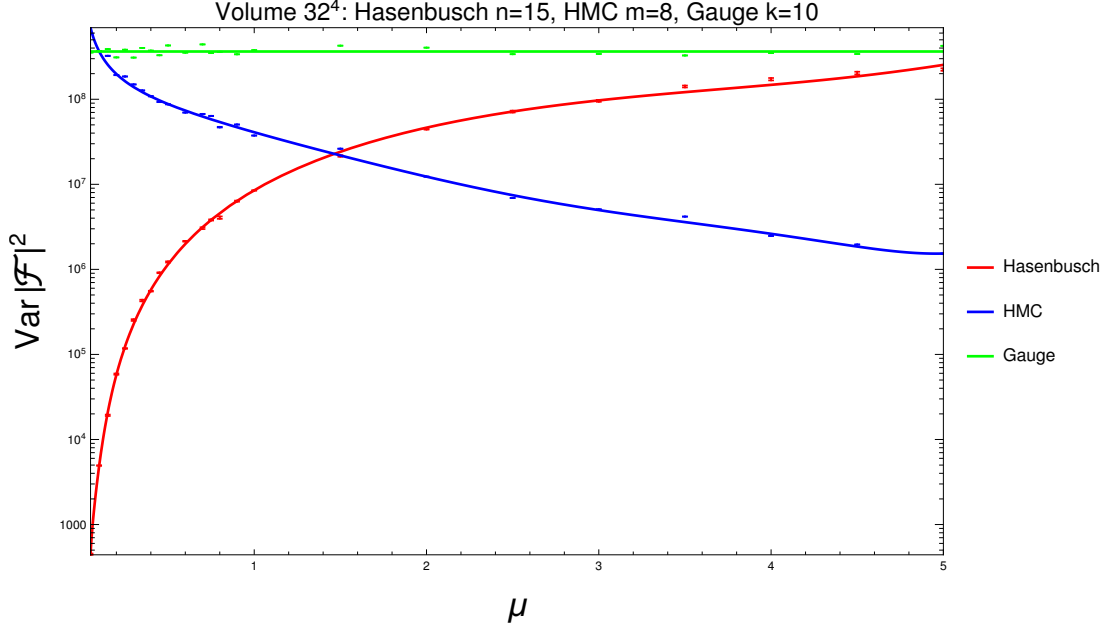
3

Figure 2: $\mathrm{Var}(|\mathscr{F}_i|^2)$ *as a function of* μ *for the different forces.*

and neglecting the covariances[3]

$$\mathrm{Var}(\delta H) \simeq \frac{\delta\tau^4}{(72^2)}\left[\mathrm{Var}(|\mathscr{F}_1|^2)(\mu) + \frac{\mathrm{Var}(|\mathscr{F}_2|^2)(\mu)}{(4m^2)^2} + \frac{\mathrm{Var}(|\mathscr{F}_3|^2)(\mu)}{(16m^2k^2)^2}\right]. \tag{3.1}$$

The total average number of MVM is given in terms of the averages a each level by

$$\#\mathrm{MVM} = (2n+1)\#\mathrm{MVM}_1(\mu) + 2n(2m+1)\#\mathrm{MVM}_2(\mu). \tag{3.2}$$

The idea is to assume that $\mathrm{Var}(\delta H)$ and # MVM depend explicitly upon n, m and k as in eqs. (3.1, 3.2) and the dependence on μ of $\mathrm{Var}(|\mathscr{F}_i|^2)$ at fixed m_0 is the only quantity to be modeled, see Figs. 2, 3.

In Fig. 2 we show the variances for the different forces and their resulting fits. We can identify two different regions: a *strong* dependence for small μ and a *weak* dependence for large μ. In the weak dependence region we have an inverted hierarchy with respect to what was our choice. In Figs. 3(a), 3(b) we show the number of MVMs per step and sub-step and their fits.

We can now build the cost as a function of n, m, k and μ. For simplicity we fix $k = 10$ and in order to find the minimum in the other parameters we require $P_{\mathrm{acc}} \gtrsim 70\%$[4]. With this set-up we found the minimum, $\mathrm{Cost}_{\min}$, to be at $(n, m, \mu) \simeq (5, 3, 0.3)$. In Fig. 4 we show the cost, normalized to the minimum, and the acceptance in the plane (μ, n) and (μ, m), for definiteness $\beta = 2.2$, $m_0 = -0.72$, $V = 32^4$. In Fig. 4(a) we fix $m = 3$, and one can see that the minimum is close to the boundary $P_{\mathrm{acc}} \sim 70\%$. By taking the cost of a simulation to be $1 < \mathrm{Cost}/\mathrm{Cost}_{\min} < 1.25$ we see that the minimum is quite broad $0.2 \lesssim \mu \lesssim 0.6$ and $4 \lesssim n \lesssim 8$. Same conclusions can be drawn for the Fig. 4(b).

---

[3]We checked that those are indeed of negligible size.

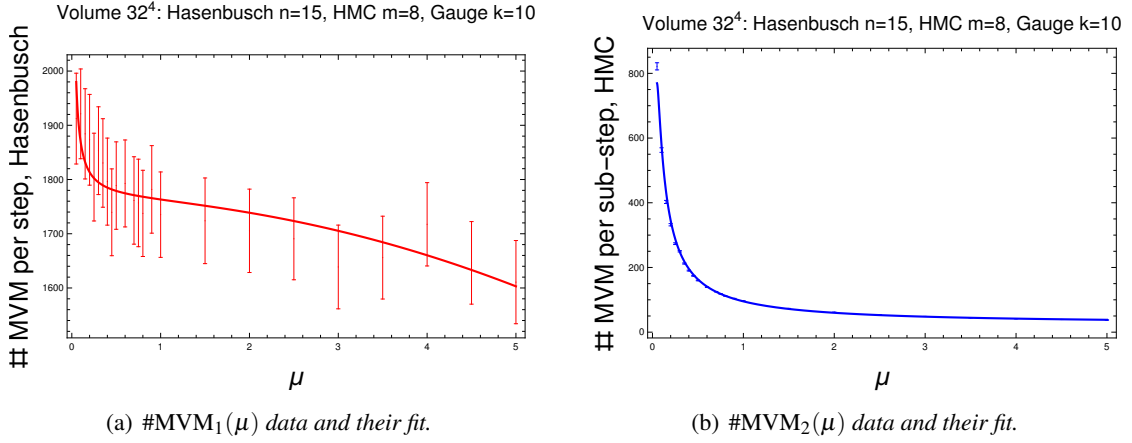[4]This requirement is needed for Creutz formula to hold true.

4

(a) #MVM$_1(\mu)$ *data and their fit.*

(b) #MVM$_2(\mu)$ *data and their fit.*

Figure 3: *Fitted quantities for the prediction of the acceptance and the cost function.*

(a) Cost/Cost$_{\min}$ *with* $m = 3$ *as function of* μ *and n.*

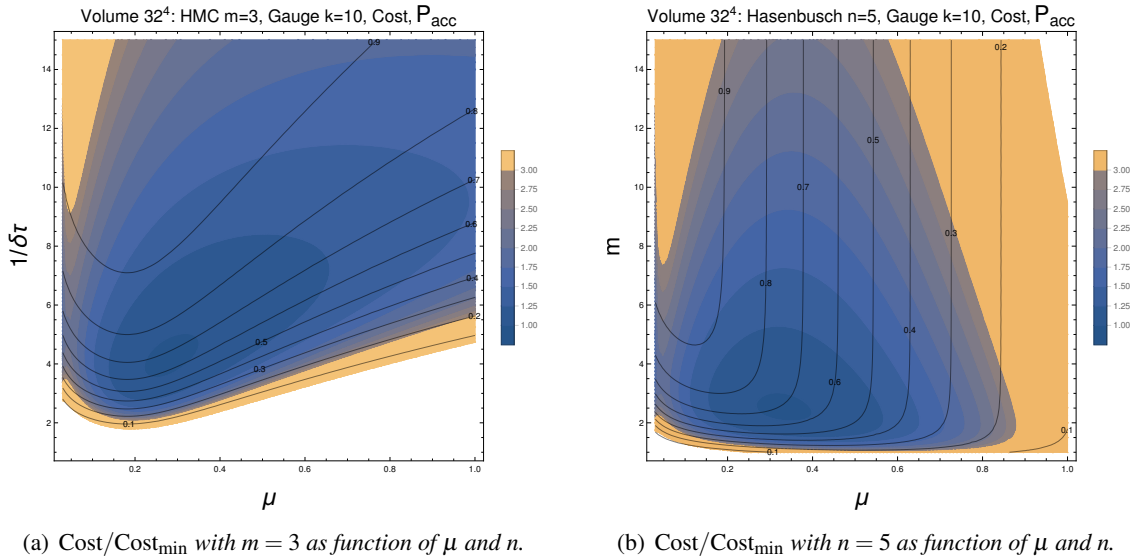(b) Cost/Cost$_{\min}$ *with* $n = 5$ *as function of* μ *and n.*

**Figure 4:** Cost/Cost$_{\min}$ *around the minimum* $(n, m, k) \simeq (5, 3, 0.3)$.

Comparison with simulation

We have run a simulation around the minimum to test our assumptions so far. The results are shown in Fig. 5, the blue line displays the prediction with the procedure described above, the circle points are the directly computed raw data (no fit in μ was performed) and the square point comes from the simulation. The results agree with both procedures within 10% that we consider satisfactory for the approach used and our goals.

4. Conclusions

We presented a strategy to optimize the parameters of the Omelyan integrator with $\alpha = 1/6$, Hasenbusch mass preconditioning and three time-scales. Our method relies on the existence of a shadow Hamiltonian.
The schematic recipe followed by this work is the following:
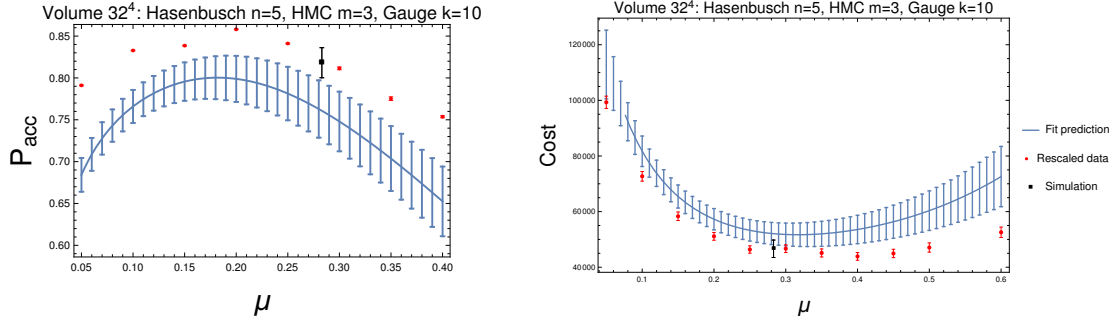
(a) *Comparison of the acceptance at the predicted minimum and a simulation.*

(b) *Comparison of the cost at the predicted minimum and a simulation.*

Figure 5: *Acceptance and cost at the minimum, comparison with simulation.*

- Start with a *reasonable* choice for the simulation of (n,m,k). In the present work we used conservative choices just to have a complete description to high values of μ.
- Measure the forces in each level, which we already computed for the evolution, and calculate $|\mathscr{F}_i|^2 = \sum_{x,\mu,a} T_{R,i} \left(F_i^{a\mu}(x)\right)^2$ and its variance $\mathrm{Var}(|\mathscr{F}_i|^2)$.
- Measure the number of MVMs in each level.
- By fitting the dependence in μ we are able to predict the acceptance and the cost dependence on (n,m,k,μ) with accuracy within 10%.

The minimization of the cost with this method is *cheap* since it employs the forces already calculated in the simulations. Generalizing it to a larger number of Hasenbusch levels on different quark determinant splitting [8] is rather straightforward, especially as long as covariances can be neglected. The results are encouraging and we plan to perform a study of the mass dependence as well as to consider different strongly interacting BSM models. Many aspects, not covered for space reason, will be discussed in a forthcoming publication.

Acknowledgments

References

[1] C. Urbach, K. Jansen, A. Shindler and U. Wenger, Comput. Phys. Commun. **174** (2006) 87 doi:10.1016/j.cpc.2005.08.006 [hep-lat/0506011].

[2] I. P. Omelyan, I. M. Mryglod, and R. Folk, Comput. Phys. Commun. **151** (2003) 272-314.

[3] M. Hasenbusch and K. Jansen, Nucl. Phys. B **659** (2003) 299 doi:10.1016/S0550-3213(03)00227-X [hep-lat/0211042].

[4] A. D. Kennedy, P. J. Silva and M. A. Clark, Phys. Rev. D **87** (2013) no.3, 034511 doi:10.1103/PhysRevD.87.034511 [arXiv:1210.6600 [hep-lat]].

[5] A. D. Kennedy and M. A. Clark, PoS LAT **2007** (2007) 038 [arXiv:0710.3611 [hep-lat]].

[6] S. Gupta, A. Irback, F. Karsch and B. Petersson, Phys. Lett. B **242** (1990) 437. doi:10.1016/0370-2693(90)91790-I

[7] M. A. Clark, A. D. Kennedy and P. J. Silva, PoS LATTICE **2008** (2008) 041 [arXiv:0810.1315 [hep-lat]].

[8] M. Luscher, Comput. Phys. Commun. **165** (2005) 199 doi:10.1016/j.cpc.2004.10.004 [hep-lat/0409106].