Field-Level Inference with Microcanonical Langevin Monte Carlo

Adrian E. Bayer [1 2 3 4 *] **Uroš Seljak** [3 4 *] **Chirag Modi** [2 5]

Abstract

Field-level inference provides a means to optimally extract information from upcoming cosmological surveys, but requires efficient sampling of a high-dimensional parameter space. This work applies Microcanonical Langevin Monte Carlo (MCLMC) to sample the initial conditions of the Universe, as well as the cosmological parameters σ_8 and Ω_m, from simulations of cosmic structure. MCLMC is shown to be over an order of magnitude more efficient than traditional Hamiltonian Monte Carlo (HMC) for a $\sim 2.6 \times 10^5$ dimensional problem. Moreover, the efficiency of MCLMC compared to HMC greatly increases as the dimensionality increases, suggesting gains of many orders of magnitude for the dimensionalities required by upcoming cosmological surveys.

1. Introduction

Modern cosmology has many open problems, ranging from measuring the neutrino mass to uncovering the nature of dark energy. To answer these questions, upcoming surveys of cosmic structure, such as DESI (DESI Collaboration, 2016), PFS (Takada et al., 2014), Rubin Observatory LSST (LSST Science Collaboration, 2009), Euclid (Laureijs et al., 2011), SPHEREx (Doré et al., 2014), SKA (Dewdney et al., 2009), and Roman Space Telescope (Spergel et al., 2013) will provide unprecedented amounts of high-resolution data. These rich datasets promise much information from small scales, however, the nonlinear nature of these scales presents a challenge in optimally extracting the information. One promising method to extract all of the information is field-level inference, which performs Bayesian inference of the initial conditions for every voxel in the cosmic field, as well as the cosmological parameters. This involves sampling a high dimensional parameter space: for modern survey volumes of $10\,(\mathrm{Gpc}/h)^3$ with resolutions of order $4\,\mathrm{Mpc}/h$ this leads to 10^8 dimensions or more, which can be prohibitively slow to converge with standard samplers.

The traditionally used method for sampling in field-level inference is (Metropolis adjusted) Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Neal et al., 2011; Betancourt, 2017). This method excels over pure Metropolis-Hastings (MH) by employing gradient information. However, in order to ensure detailed balance each step must undergo an accept-reject criterion, which in turn requires a small step size and reduces the efficiency. Recently, alternative sampling methods called Microcanonical Hamiltonian/Langevin Monte Carlo (MCHMC/MCLMC) (Robnik et al., 2022; Robnik & Seljak, 2023) have been introduced. These samplers move fast in regions of low density and slow in the regions of high density, in contrast to HMC. They also do not require an accept-reject criterion, allowing a larger tolerance of energy fluctuations, in turn enabling a larger step size and larger effective sample size (ESS). Moreover, their ESS scales with dimensionality more favorably than HMC.

In this work we perform field-level inference to jointly sample the initial modes, and the cosmological parameters Ω_m and σ_8. The parameter Ω_m corresponds to the amount of matter in the Universe and σ_8 corresponds to the amplitude of density fluctuations. There is an often reported tension between measurements of σ_8 made by different cosmological experiments (Di Valentino, 2021), thus using field-level inference to give an optimal measurement would aid in resolving this tension.

The structure of the paper is as follows. In §2 we review field-level inference and outline the HMC, MCHMC, and MCLMC methods. In §3 we discuss related work. In §4 we apply the methods and analyse the ESS as a function of nonlinearity and dimensionality. We finally conclude in §5.

2. Background

We now outline field-level inference in §2.1, HMC in §2.2, MCHMC in §2.3, and MCLMC in §2.4.

---

[*]Equal contribution [1]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA [2]Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA [3]Department of Physics, University of California, Berkeley, CA, USA [4]Lawrence Berkeley National Laboratory, Berkeley, CA, USA [5]Center for Computational Mathematics, Flatiron Institute, New York, NY, USA. Correspondence to: Adrian E. Bayer <abayer@princeton.edu>.

2.1. Field-Level Inference

The goal of field-level inference is to infer the initial modes of the Universe s and/or the cosmological parameters λ. Given a forward model $f(s, \lambda)$, which in this work computes the 3d dark matter overdensity field, the posterior is given by

$$-2 \log P(s, \lambda | d) = \sum_{\vec{k}} \left[\frac{|f_{\vec{k}}(s, \lambda) - d_{\vec{k}}|^2}{N_{\vec{k}}} + |s_{\vec{k}}|^2 \right], \quad (1)$$

where the first term is the likelihood, with data d and noise N, the second term is the prior, and the sum is over all voxels in k-space. Note that s is defined to have unit variance.

Inferring s corresponds to the task of reconstructing the initial conditions of the Universe, however, often s are treated as nuisance parameters and one marginalizes over s to yield only constraints on the cosmological parameters λ.

2.2. Hamiltonian Monte Carlo

The traditional approach for sampling in the context of field-level inference is HMC (Duane et al., 1987; Neal et al., 2011; Betancourt, 2017). Given a d-dimensional target distribution $p(z) \propto e^{-\mathcal{L}(z)}$, where $z \in \mathbb{R}^d$, HMC uses the gradient $\nabla \mathcal{L}(z)$ to improve the sampling efficiency compared no-gradient methods such as MH. It considers the Hamiltonian $H(z, \Pi)$, where Π is the canonical momentum, and samples the canonical ensemble in $2d$-dimension phase space, denoted by $p(z, \Pi) \propto e^{-H(z, \Pi)}$. The success of HMC relies on the tuning of the Hamiltonian such that the marginal of $p(z, \Pi)$ over Π converges to the target distribution,

$$p(z) \propto \int_{\mathbb{R}^d} d\Pi \, e^{-H(z, \Pi)}. \quad (2)$$

The most popular choice is the Hamiltonian of a particle in a potential, $H(z, \Pi) = \frac{1}{2} \Pi^2(z) + \mathcal{L}(z)$, for which the solution is the set of ODEs,

$$dz = u dt,$$
$$du = -\nabla \mathcal{L}(z) dt, \quad (3)$$

where t is time and u is velocity. Following Hamiltonian dynamics ensures the trajectory conserves the Hamiltonian, or energy, allowing efficient exploration at a fixed energy level. Different energy levels must be explored to obtain an accurate set of samples, which is achieved by resampling the momentum Π according to its marginal distribution (a normal distribution) and results in inefficiencies (Betancourt, 2017). Moreover, HMC additionally requires an MH accept-reject step, which necessitates a sufficiently small step size to ensure a frequent rate of acceptance.

2.3. Microcanonical Hamiltonian Monte Carlo

Unlike HMC which considers the marginal of the canonical distribution, the approach of MCHMC is to tune the Hamiltonian such that the microcanonical distribution marginalized over the momentum variables gives the target distribution, as follows

$$p(z) \propto \int_{\mathbb{R}^d} d\Pi \, \delta(H(z, \Pi) - E), \quad (4)$$

where $\delta(\cdot)$ denotes the delta function, and E is the energy. There are many solutions for the Hamiltonian, but the one adopted for MCHMC yields the equations of motion,

$$dz = u dt,$$
$$du = P(u) f(z) dt, \quad (5)$$

where we have introduced the projection $P(u) \equiv (I - uu^T)$ and force $f(z) \equiv -\nabla \mathcal{L}(z)/(d-1)$ (Ver Steeg & Galstyan, 2021; De Luca & Silverstein, 2022; Robnik et al., 2022). The key difference to HMC in Eqn. (3) is the projection. Unlike HMC, the MCHMC dynamics converges to the target distribution while maintaining a constant energy.

However, following Eqns. 5 exactly will not ensure ergodicity; for example, Nöther's Theorem implies that any symmetry will lead to a conserved quantity and in turn confine the dynamics to a subspace of the energy surface. Ergodicity can be achieved by energy-conserving momentum resampling (Chernov & Markarian, 2006; De Luca & Silverstein, 2022; Robnik & Seljak, 2023), in which billiard-like bounces to the momentum are introduced such that the direction of the trajectory is changed, but the magnitude of the momentum, and thus the energy, is the same. Introducing momentum decoherence in this manner is analogous to the resampling of the momentum in HMC, but with fixed energy. The frequency of momentum resampling is a hyperparameter of MCHMC.

2.4. Microcanonical Langevin Monte Carlo

To further speed up reaching ergodicity, the ODEs can be modified by considering Langevin dynamics (Grenander & Miller, 1994; Girolami & Calderhead, 2011) such that,

$$dz = u dt, \quad (6)$$
$$du = P(u) [f(z) dt + \eta dW], \quad (7)$$

where η is a hyperparameter and W is a standard normal random vector. This additional term proportional to η can be understood physically as a diffusion term which enforces better exploration of the target, in turn boosting ergodicity.

MCLMC has two hyperparameters, the step size and the amount of noise η. Both of these parameters can be tuned during a burn-in stage by monitoring fluctuations in the energy and ensuring they are below a certain threshold.
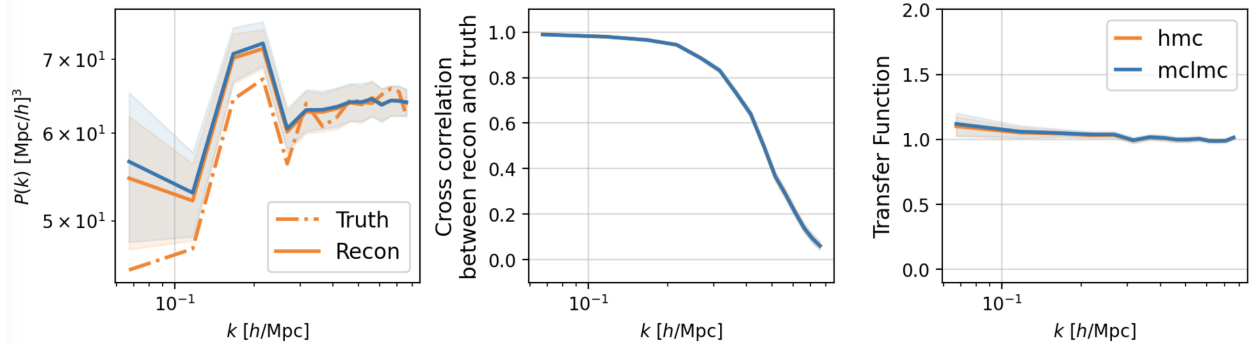
Figure 1. Samples of initial modes. Left: power spectrum of whitened modes, Center: cross-correlation, Right: transfer function.
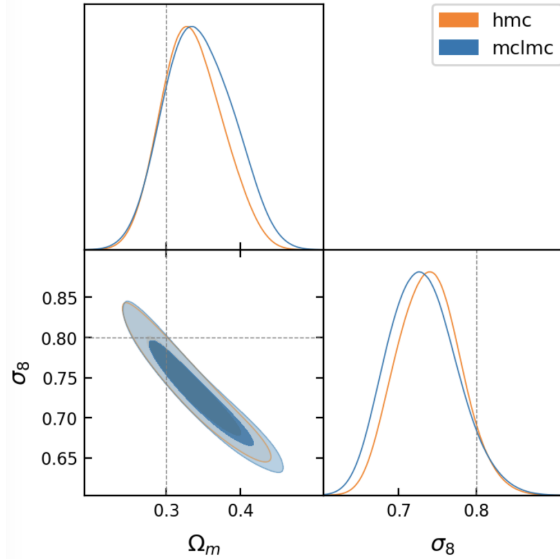

Figure 2. Samples of cosmological parameters. Dashed lines signify the truth.

Figure 3. Comparison of ESS per gradient evaluation for different levels of nonlinearity. Top: ESS of MCLMC (blue) and HMC (orange), for the modes (x) and cosmological parameters (+). Bottom: Ratio of ESS from MCLMC versus HMC.

3. Related Work

Current approaches for sampling in field-level inference field-level inference include approximating the posterior around the optimum (Seljak et al., 2017; Millea & Seljak, 2022) and HMC (Jasche & Wandelt, 2013). Hybrid HMC approaches based on variational inference have also been developed (Modi et al., 2023).

In terms of MCHMC, one type of microcanonical hamiltonian dynamics in the context of sampling was first explored by Ver Steeg & Galstyan (2021). The dynamics have also been applied in the context of optimization by De Luca & Silverstein (2022). A more general solution to sampling was provided by Robnik et al. (2022), and the extension to Langevin dynamics was made by Robnik & Seljak (2023) in the context of lattice field theory. Robnik et al. (2022); Robnik & Seljak (2023) showed MCHMC/MCLMC to be
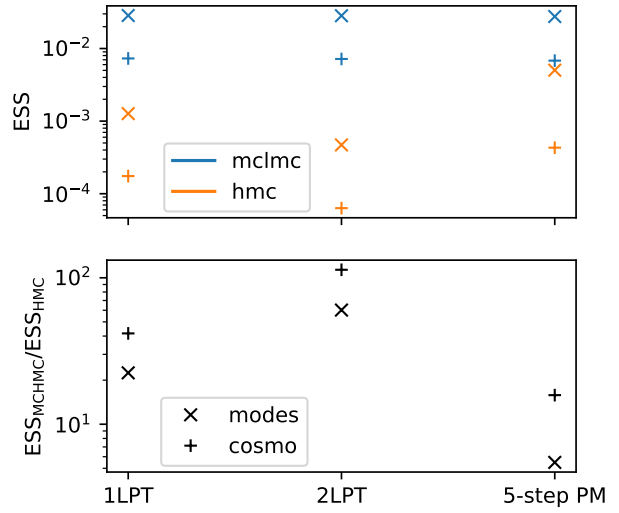
superior to other HMC-based methods, such as NUTS (Hoffman et al., 2014), which we will thus not consider here.

4. Experiments

We simulate the nonlinear dark matter field using pmwd (Li et al., 2022), a differentiable particle-mesh (PM) code written in JAX. It is capable of fitting 512^3 particles on a single 80GB NVIDIA A100 GPU, and can evaluate one PM step in ~ 0.1 seconds. This code takes the cosmological parameters and the 3d intial gaussian field as input, and outputs the final 3d nonlinear dark matter field. It can perform nonlinear modeling to various levels of accuracy; we consider 3 levels of nonlineaarity, the Zel'dovich Approximation (ZA, or 1LPT), 2nd order Lagrangian perturbaiton theory (2LPT), and 2LPT followed by 5 steps of PM simu-
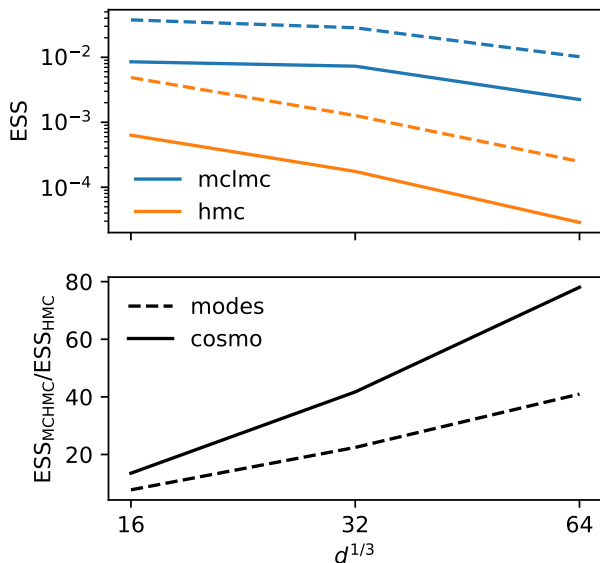
Figure 4. Scaling of ESS per gradient evaluation as a function of dimensionality d for 1LPT. Top: ESS for MCLMC (blue) and HMC (orange), for the modes (dashed) and cosmological parameters (solid). Bottom: Ratio of ESS from MCLMC versus HMC.

lation. We consider three dimensionalities (corresponding to the number of voxels): $16^3, 32^3$, and 64^3. In each case the voxel resolution is $4\,\mathrm{Mpc}/h$.

We perform inference over the initial modes in k-space (denoted s in §2.1), and the cosmological parameters Ω_m and σ_8. We precondition the modes by the theoretical posterior variance according to linear theory, i.e. by applying the multiplicative factor $(P_L(k) + N)/N$, where P_L is the linear power at the fiducial cosmology. This approximately ensures all voxels in k-space have unit variance, improving the sampler mixing. Similarly, we precondition the cosmological parameters to approximately have unit variance.

We run a chain of 10,000 samples with 2,000 steps of burn in (to be conservative). During HMC burn in, the step size is tuned to ensure a 65% acceptance rate. We use 40 leapfrog steps for HMC, which we found to be optimal. For MCLMC we tuned the step size and noise to ensure energy fluctuations per dimension are below a threshold of 10^{-4}. We apply the Minimum-Norm integrator to solve the MCLMC ODEs.

Fig. 1 shows the sampled modes for the 32^3 1LPT example. There is excellent agreement between HMC and MCLMC. Similarly, Fig. 2 shows the sampled cosmological parameters, again showing excellent agreement. This implies that MCLMC produces equally accurate samples to HMC.

Fig. 3 compares the ESS per gradient evaluation between 1LPT, 2LPT, and 5-step PM. We consider the ESS for the

modes and cosmological parameters separately. MCLMC outperforms HMC in all cases. The increase in ESS is greatest for 2LPT (two orders of magnitude), and lowest for 5-step PM. This is likely affected by the use of linear mode preconditioning, which we will improve upon in future work.

In terms of the efficiency of the algorithms, Fig. 4 shows the ESS per gradient evaluation as a function of dimensionality for 1LPT. It can be seen in the top panel that while the ESS of HMC steadily decreases, the ESS of MCLMC decreases more mildly. This is because HMC requires a sufficiently small step size to prevent a relative energy error greater than unity: the energy scales with dimensionality, thus the step size required to achieve a relative error of less than unity decreases with dimensionality. On the other hand, MCLMC has no such constraint, meaning that for a fixed condition number there is no dependence of the ESS on dimensionality. In this case there is a slight increase in condition number as we include larger scale modes, meaning there is a slight decrease in ESS for MCLMC when moving to higher dimensions. The lower panel of Fig. 4 shows how the ESS ratio of MCLMC versus HMC greatly increases with dimensionality. For 64^3 the improvement is a factor of 40 for the modes and 80 for the cosmological parameter. This shows much promise to speed up sampling by many orders of magnitude when moving to the higher dimensionalities required to analyze upcoming cosmological data.

5. Conclusions

We have shown that MCLMC has over an order of magnitude higher ESS than HMC in the context of field-level inference. Moreover, the scaling of the ESS with dimensionality is more favorable for MCLMC, showing great promise to improve sampling efficiency by many orders of magnitude when moving to the higher dimensions required for upcoming cosmological data. We have also shown this for various levels of nonlinearity (1LPT, 2LPT, and 5-step PM); the improved efficiency of MCLMC makes it feasible to sample even when performing highly nonlinear forward modeling.

Future work includes application to the galaxy field, which will require sampling over additional parameters describing the matter–galaxy connection. Moreover, one can improve constraints by perform joint inference with other cosmological tracers, such as peculiar velocities (Bayer et al., 2022). Additionally, the preconditioning of the posterior could be improved to achieve a greater ESS, for example by fitting a normalizing flow to the posterior before running MCLMC (Hoffman et al., 2019). Finally, the method will be applied to higher-dimensional maps which correspond to the data from upcoming cosmological surveys, including survey masks and systematic effects.

References

Bayer, A. E., Modi, C., and Ferraro, S. Joint velocity and density reconstruction of the Universe with nonlinear differentiable forward modeling. 10 2022. URL https://arxiv.org/abs/2210.15649.

Betancourt, M. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*, 2017.

Chernov, N. and Markarian, R. Chaotic billiards, volume 127 of mathematical surveys and monographs. *American Mathematical Society, Providence, RI*, 2006.

De Luca, G. B. and Silverstein, E. Born-infeld (BI) for AI: Energy-conserving descent (ECD) for optimization. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 4918–4936. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/de-luca22a.html.

DESI Collaboration. The desi experiment part i: Science,targeting, and survey design, 2016.

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., and Lazio, T. J. L. W. The Square Kilometre Array. *IEEE Proceedings*, 97(8):1482–1496, August 2009. doi: 10.1109/JPROC.2009.2021005.

Di Valentino, E. Cosmology intertwined iii: $f\sigma_8$ and s_8. *Astroparticle Physics*, 131:102604, 2021. ISSN 0927-6505. doi: https://doi.org/10.1016/j.astropartphys.2021.102604. URL https://www.sciencedirect.com/science/article/pii/S0927650521000487.

Doré, O., Bock, J., Ashby, M., Capak, P., Cooray, A., de Putter, R., Eifler, T., Flagey, N., Gong, Y., Habib, S., Heitmann, K., Hirata, C., Jeong, W.-S., Katti, R., Korngut, P., Krause, E., Lee, D.-H., Masters, D., Mauskopf, P., Melnick, G., Mennesson, B., Nguyen, H., Öberg, K., Pullen, A., Raccanelli, A., Smith, R., Song, Y.-S., Tolls, V., Unwin, S., Venumadhav, T., Viero, M., Werner, M., and Zemcov, M. Cosmology with the SPHEREX All-Sky Spectral Survey. *arXiv e-prints*, art. arXiv:1412.4872, December 2014. doi: 10.48550/arXiv.1412.4872.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

Girolami, M. and Calderhead, B. Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(2):123–214, 03 2011. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2010.00765.x. URL https://doi.org/10.1111/j.1467-9868.2010.00765.x.

Grenander, U. and Miller, M. I. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(4):549–603, 1994. ISSN 00359246. URL http://www.jstor.org/stable/2346184.

Hoffman, M., Sountsov, P., Dillon, J. V., Langmore, I., Tran, D., and Vasudevan, S. Neutra-lizing bad geometry in hamiltonian monte carlo using neural transport, 2019.

Hoffman, M. D., Gelman, A., et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

Jasche, J. and Wandelt, B. D. Bayesian physical reconstruction of initial conditions from large-scale structure surveys. *MNRAS*, 432(2):894–913, June 2013. doi: 10.1093/mnras/stt449.

Laureijs, R., Amiaux, J., Arduini, S., Auguères, J. ., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., and et al. Euclid Definition Study Report. *ArXiv e-prints*, October 2011.

Li, Y., Lu, L., Modi, C., Jamieson, D., Zhang, Y., Feng, Y., Zhou, W., Kwan, N. P., Lanusse, F., and Greengard, L. pmwd: A differentiable cosmological particle-mesh n-body library, 2022. URL https://arxiv.org/abs/2211.09958.

LSST Science Collaboration. LSST Science Book, Version 2.0. *arXiv e-prints*, art. arXiv:0912.0201, December 2009.

Millea, M. and Seljak, U. Marginal unbiased score expansion and application to CMB lensing. *PRD*, 105(10): 103531, May 2022. doi: 10.1103/PhysRevD.105.103531.

Modi, C., Li, Y., and Blei, D. Reconstructing the universe with variational self-boosted sampling. *JCAP*, 03:059, 2023. doi: 10.1088/1475-7516/2023/03/059.

Neal, R. M. et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

Robnik, J. and Seljak, U. Microcanonical langevin monte carlo, 2023. URL https://arxiv.org/abs/2303.18221.

Robnik, J., Luca, G. B. D., Silverstein, E., and Seljak, U. Microcanonical hamiltonian monte carlo, 2022. URL https://arxiv.org/abs/2212.08549.

Seljak, U., Aslanyan, G., Feng, Y., and Modi, C. Towards optimal extraction of cosmological information from nonlinear data. *Journal of Cosmology and Astroparticle Physics*, 2017(12):009–009, Dec 2017. ISSN 1475-7516. doi: 10.1088/1475-7516/2017/12/009. URL http://dx.doi.org/10.1088/1475-7516/2017/12/009.

Spergel, D., Gehrels, N., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O., Hirata, C., Kalirai, J., Kasdin, N. J., Moos, W., Perlmutter, S., Postman, M., Rauscher, B., Rhodes, J., Wang, Y., Weinberg, D., Centrella, J., Traub, W., Baltay, C., Colbert, J., Bennett, D., Kiessling, A., Macintosh, B., Merten, J., Mortonson, M., Penny, M., Rozo, E., Savransky, D., Stapelfeldt, K., Zu, Y., Baker, C., Cheng, E., Content, D., Dooley, J., Foote, M., Goullioud, R., Grady, K., Jackson, C., Kruk, J., Levine, M., Melton, M., Peddie, C., Ruffa, J., and Shaklan, S. Wide-field infrared survey telescope-astrophysics focused telescope assets wfirst-afta final report, 2013.

Takada, M., Ellis, R. S., Chiba, M., Greene, J. E., Aihara, H., Arimoto, N., Bundy, K., Cohen, J., Doré, O., Graves, G., and et al. Extragalactic science, cosmology, and galactic archaeology with the subaru prime focus spectrograph. *Publications of the Astronomical Society of Japan*, 66(1):R1, Feb 2014. ISSN 2053-051X. doi: 10.1093/pasj/pst019. URL http://dx.doi.org/10.1093/pasj/pst019.

Ver Steeg, G. and Galstyan, A. Hamiltonian dynamics with non-newtonian momentum for rapid sampling. *Advances in Neural Information Processing Systems*, 34:11012–11025, 2021.