

# Nonlocal stochastic-partial-differential-equation limits of spatially correlated noise-driven spin systems derived to sample a canonical distribution

Yuan Gao<sup>1,\*</sup>, Jeremy L. Marzuola<sup>1,†</sup>, Jonathan C. Mattingly<sup>2,‡</sup> and Katherine A. Newhall<sup>1,§</sup>

<sup>1</sup>*Department of Mathematics, University of North Carolina at Chapel Hill, CB 3250, Chapel Hill, North Carolina 27599-3250, USA*

<sup>2</sup>*Department of Mathematics and Department of Statistical Sciences, Duke University, Box 90320, Durham, North Carolina 27708-0320, USA*



(Received 28 July 2020; accepted 12 October 2020; published 9 November 2020)

For a noisy spin system, we derive a nonlocal stochastic version of the overdamped Landau-Lipshitz equation designed to respect the underlying Hamiltonian structure and sample the canonical or Gibbs distribution while being driven by spatially correlated (colored) noise that regularizes the dynamics, making this Stochastic partial differential equation mathematically well-posed. We begin from a microscopic discrete-time model motivated by the Metropolis-Hastings algorithm for a finite number of spins with periodic boundary conditions whose values are distributed on the unit sphere. We thus propose a future state of the system by adding to each spin colored noise projected onto the sphere, and then accept this proposed state with probability given by the ratio of the canonical distribution at the proposed and current states. For uncorrelated (white) noise this process is guaranteed to sample the canonical distribution. We demonstrate that for colored noise, the method used to project the noise onto the sphere and conserve the magnitude of the spins impacts the equilibrium distribution of the system, as coloring projected noise is not equivalent to projecting colored noise. In a specific scenario we show this break in symmetry vanishes with vanishing proposal size; the resulting continuous-time system of Stochastic differential equations samples the canonical distribution and preserves the magnitude of the spins while being driven by colored noise. Taking the continuum limit of infinitely many spins we arrive at the aforementioned version of the overdamped Landau-Lipshitz equation. Numerical simulations are included to verify convergence properties and demonstrate the dynamics.

DOI: [10.1103/PhysRevE.102.052112](https://doi.org/10.1103/PhysRevE.102.052112)

## I. INTRODUCTION

In order to accurately describe noise-induced phenomenon in spatially extended systems, it is important to add fluctuations to continuum models that respect some underlying structure like a Hamiltonian and the sampling of the Gibbs (or Boltzmann or canonical) distribution. Guaranteeing this kind of fluctuation-dissipation relation (a.k.a. detailed balance) is not always obvious, especially in condensed matter physics for which accurate phenomenological models are not always built from first principles. One example is the Landau-Lifshitz-Gilbert equation describing a single magnetic spin requiring multiplicative noise, thereby creating an effective electric field, rather than additive noise to ensure sampling of the Gibbs distribution; see Ref. [1]. In effect, the noise is projected onto the surface of the sphere representing the configuration space of the constant magnitude spin vector. Another example is the regularization of stochastic partial differential equations (SPDEs) by correlating the noise in space. While the corresponding dynamics occur at regularity scales that allow for analysis of the evolution to be treated via now well-understood methods for understanding stochastic

paths in the PDE setting (see, for instance, Refs. [2,3]), entirely different distributions from their un-correlated noise counterparts may be sampled. Although white-noise solutions to SPDEs in situations with much less regularity can be understood with the introduction of regularity structures by Hairer in Ref. [4], there are still dimensional restrictions, even in the case where the deterministic part is parabolic and hence strongly coercive, see for instance the recent work [5] on stochastic harmonic map heat flows. Our goal in this work is to combine the two considerations above related to sampling and regularization, deriving an SPDE model for a spatially extended magnetic spin system with spatially “colored” noise designed to sample an invariant Gibbs measure.

We derive such a continuum model designed to sample an invariant Gibbs measure from a microscopic Metropolis-Hastings (MH) algorithm. The MH algorithm [6,7] allows the random walk dynamics to be separated from the Hamiltonian structure in the invariant measure: a simple random-walk proposal,  $\tilde{X}_i = X_i^n + \varepsilon w_i^n$  with  $i = 1, \dots, N$  indexing space and the  $w_i^n$  independent normally distributed random variables, will sample the Gibbs measure

$$\mu(\vec{X}) = Z^{-1} e^{-\beta H(\vec{X})}, \quad (1)$$

where  $Z$  is the partition function and  $\beta^{-1} = k_B T$ , if an accept probability of

$$\alpha = 1 \wedge e^{-\beta[H(\vec{X}) - H(\vec{X}^n)]},$$

\*gaoyuanmath@gmail.com

†marzuola@math.unc.edu

‡jonm@math.duke.edu

§knewhall@unc.edu

where  $a \wedge b = \min(a, b)$  is used for arbitrary bounded Hamiltonian  $H$  (i.e.,  $\tilde{X}^{n+1} = \tilde{X}$  with probability  $\alpha$  and  $\tilde{X}^n$  otherwise). The stochastic differential equation (SDE)

$$d\tilde{X} = -\nabla H dt + \sqrt{2\beta^{-1}} d\tilde{W}$$

also samples the invariant measure (1). Furthermore the MH dynamics converge to the SDE dynamics in the limit as the proposal size  $\varepsilon \rightarrow 0$ . Thus the limiting MH dynamics can be used to construct SDE models that preserve the invariant measure (1) in more complex situations. For example, if the random walk proposal is changed to  $\tilde{X} = \tilde{X}^n + \varepsilon B \tilde{w}^n$  for constant matrix  $B$ , then the SDE

$$d\tilde{X} = -BB^T \nabla H dt + \sqrt{2\beta^{-1}} B d\tilde{W} \quad (2)$$

also samples the invariant measure (1) (but not for every nonconstant  $B(\tilde{X})$ ; cf. Ref. [8]). In Appendix B 1 we confirm this by direct substitution into the (constant  $B$ ) Fokker-Planck equation

$$\begin{aligned} \partial_t \rho(x, t) = & \sum_{i=1}^{3N} \partial_i [(BB^T \nabla H)_i \rho(x, t)] \\ & + \beta^{-1} \sum_{i,j=1}^{3N} (BB^T)_{ij} \partial_i \partial_j \rho(x, t). \end{aligned} \quad (3)$$

Equation (2), with symmetric, non-negative definite covariance matrix  $BB^T$ , has spatially correlated noise and still samples the Gibbs distribution (1). A continuum limit of the SDE (2) exists if the Hamiltonian  $H$  and covariance matrix  $BB^T$  are appropriately scaled with system size  $N$ .

In this work, we consider a system of  $N$  spins (with periodic boundary conditions), or vectors on  $\mathbb{S}^m$  for some  $m \geq 1$ , thereby introducing a confining geometry and investigate how this interacts with spatially correlated “colored” noise, deriving an appropriately regularized Stochastic partial differential equation that still samples an invariant measure of the form (1). The spatially correlated noise coupled to the geometric constraint will result in a proposal of the form  $\tilde{X} = \tilde{X}^n + \varepsilon B(\tilde{X}^n) \tilde{w}^n$ , where unfortunately the *colored noise proposal is no longer symmetric*. However, we prove that the MH dynamics can still be approximated by an SDE system similar to that of (2), and that for a canonical choice of the matrix  $B$  related to the geometry, that the SDE system still samples the correct invariant measure.

For ease of exposition and physical importance, we will restrict ourselves to  $m = 2$  and work only with spins defined as vectors on  $\mathbb{S}^2$ . We build on our recent work [9] which showed that on a general torus in any dimension,  $\mathbb{T}^d$ , the MH dynamics for a system of spatially uncorrelated “white” noise driven spins with confining geometry converged to the dynamics of an SDE system as  $\varepsilon$ , the proposal size, went to zero. We also considered the  $N \rightarrow \infty$  limit of the dynamics while quenching the noise ( $\beta = N^\gamma$  for  $\gamma$  sufficiently large) to arrive at the harmonic map heat flow equation

$$\partial_t \sigma(x, t) = -\sigma \times (\sigma \times \Delta \sigma). \quad (4)$$

This could also be referred to as the overdamped Landau-Lifshitz-Gilbert (LLG) equation. Quenching the noise was

essential in the derivation due to the known convergence issues with stochastic partial differential equations (SPDE) in spatial dimensions greater than one (cf. [10]). The convergence from the SDE model to a PDE model also relied on the regularity of the harmonic map heat flow equation, which can fail for  $\mathbb{T}^d \rightarrow \mathbb{S}^m$  in finite time for dimensions  $d > 2$  due to bubbling singularities, see [11, 12].

To derive a regularized SPDE limit ( $\beta$  constant with  $N \rightarrow \infty$ ), we begin with a random walk for the MH algorithm that projects now spatially correlated Gaussian noise onto the tangent plane of the underlying geometry. After taking the proposal size  $\varepsilon \rightarrow 0$  arriving at a system of SDEs, we find that unlike the white noise case, the choice of  $\sigma \times$  as the projection is crucial for sampling the desired distribution (1). Therefore, the regularized nonlocal SPDE that samples the Gibbs distribution is

$$\begin{aligned} \partial_t \sigma(x, t) = & -\sigma(x, t) \times \int_{\mathbb{T}^d} C(x-y) (\sigma \times \Delta \sigma)(y, t) dy \\ & + \sqrt{2\beta^{-1}} \sigma(x, t) \times \eta^C(x, t), \end{aligned} \quad (5)$$

where  $C$  is a nonlocal operator to be described below in a variety of cases that encodes the covariance structure of the colored noise,  $\eta^C(x, t)$ , is colored-in-space white-in-time noise [i.e.,  $\mathbb{E}[\eta^C(x, t) \eta^C(y, s)] = C(x-y) \delta(t-s)$ ], interpreted in the Stratonovich sense.

### A. Prior work

Having an appropriately regularized stochastic limit is important to studying thermal effect in ferromagnets such as magnetization reversal [13, 14]. Existing field models continue to use spatially uncorrelated white noise in the stochastic LLG equation so as to maintain the equilibrium distribution, proposing for example weak formulations of the solutions and numerical finite element schemes (cf. chap. 2 of Ref. [15]). Equation (5) is in contrast to regularizing the LLG equation by changing the energy functional to include a term to control the modulus of continuity [16]. It also compliments other works that derive equations to preserve the equilibrium distribution, such as in the case of inhomogeneous magnitude of magnetic spins [17], for temporally colored noise but for finitely many spins [18], and for the stochastic Landau-Lifshitz-Bloch equation [19]. More generally, physical models with confining geometries are natural generalizations of the SPDE limits derived using colored noise for unconstrained random walks in Ref. [20] and more recently in Ref. [21]. See also Ref. [22], which focuses on quantum field theories, but also discusses the effective action of a generic SPDE system through the tools of fluctuation-dissipation and invariant measures, with examples including reaction-diffusion-decay systems, KPZ (noisy Burgers), and purely dissipative SPDEs.

Since our approach starts from the MH algorithm, it is worth pointing out that the MH algorithm itself is widely used in particle statistics and sampling algorithms; see, for instance, Refs. [23–27]. It also arises when adopting the Bayesian approach to inverse problems and signal processing [28, 29]. This has led to the study of optimal scalings for the unconstrained random walk MH algorithm and diffusion limits for certain forms of probability distributions [20, 21, 30–33]. Specifically, for product measures in Ref. [30] and the

Gibbs distribution of a lattice model in Ref. [31], the weak convergence to Langevin diffusions has been shown by comparing generator functions. The pioneering work [20], based in part upon earlier works on sampling [34,35], extended this type of result to nonproduct form measures and demonstrated the weak convergence to a SPDE. Subsequent works [21,32,33] consider scaling limits of systems started away from their equilibrium distributions.

Building on our previous work [9] that studied the limiting dynamics of a geometric MH process with white noise in the proposal, we fill a missing gap in the above results showing strong convergence of trajectories started far from equilibrium to a nonlocal SPDE in a geometric setting, with the underlying dynamics of the process designed to sample an (non product form) invariant measure using colored noise with a given covariance structure. Similar to Ref. [20], we derive a drift term that implicitly is driven by a nonlocal diffusion operator. In the context of random walks, this is related to fractional diffusion operators, but we are interested to see the effects of colored noise on the geometric evolution.

### B. Outline of results

The remainder of the paper is as follows. In Sec. II we lay out the vector notation we adapt for the paper. In Sec. III we review our results from Ref. [9] pointing out a few interesting facts that will be in contrast to the colored noise case. We extend these results to the case of colored noise in Sec. IV, outlining the derivation of the limiting SDE system from the MH dynamics in Sec. IV A (details of the proof are in Appendix A), discussing the correct projection of the noise onto the tangent plane of the underlying geometry to ensure the SDE system samples the desired distribution (1) in Sec. IV B, proving the invariant measure of the MH dynamics converges to this same invariant measure in Sec. IV C, and discussing the Fourier representation of the nonlocal SPDE (5) in Sec. IV D with an outline the well-posedness in Appendix C when the noise is trace class. We support our trajectory-wise convergence results with direct numerical simulations in Sec. V as well as illuminate the differences between the choice of two different projections of the noise onto the tangent plane of the underlying geometry. We give concluding remarks in Sec. VI.

## II. NOTATION

We present our results for the case of one periodic spatial dimension,  $\mathbb{T}$ , and spins that live on  $\mathbb{S}^2$ , although this can be extended to other dimensions for both the periodic domain and the spherical target. It becomes convenient to adopt different notation in different contexts, which we summarize here. The torus with unit length is subdivided with  $x_i = (i-1)/N$  for  $i = 1, \dots, N$  with a spin located at each  $x_i$ . We take  $\sigma_i^n$  for  $i = 1, \dots, N$  as the collection of the  $N$  spins of the MH dynamics at time step  $n$ , each a three-dimensional vector, with components

$$\sigma_i^n = \langle \sigma_{i,x}^n, \sigma_{i,y}^n, \sigma_{i,z}^n \rangle \quad (6)$$

satisfying  $(\sigma_{i,x}^n)^2 + (\sigma_{i,y}^n)^2 + (\sigma_{i,z}^n)^2 = 1$  for each  $i = 1, \dots, N$  and each integer  $n \geq 0$ . The  $3N$ -dimensional vector

$$\vec{\sigma}^n = \langle \sigma_{1,x}^n \dots \sigma_{N,x}^n, \sigma_{1,y}^n \dots \sigma_{N,y}^n, \sigma_{1,z}^n \dots \sigma_{N,z}^n \rangle \quad (7)$$

contains all the components of all the spins. We similarly define  $\vec{\sigma}^n$ ,  $\tilde{\sigma}_i^n$  and  $\tilde{\sigma}_{i,q}^n$   $q \in \{x, y, z\}$  for the MH proposal at time step  $n$ ;  $\vec{w}^n$ ,  $\mathbf{w}_i^n \in \mathbb{R}^3$ , and  $w_{i,q}^n$  for the independent standard Gaussian random variables used to generate the proposal at time step  $n$ ;  $\vec{s}(t)$ ,  $s_i(t)$ ,  $s_{i,q}(t)$  for the solution to the limiting SDE system at time  $t$ . Since the noise will be correlated in each component, it will also be useful to represent it as

$$\vec{w}^n = \langle \mathbf{w}_x^n, \mathbf{w}_y^n, \mathbf{w}_z^n \rangle$$

with each  $N \times 1$  vector

$$\mathbf{w}_q^n = \langle w_{1,q} \dots w_{N,q} \rangle \text{ for } q \in \{x, y, z\}.$$

## III. WHITE NOISE

Here we present an overview of our previous work [9], pointing out a few interesting facts that will be in contrast to the colored noise case. We remind the reader that though we limit our discussion here to the cases  $d = 1, m = 2$  for ease of exposition, all results here extend to  $d \geq 1, m \geq 1$  with small modifications.

To arrive at an appropriate continuum limit, we begin with the standard MH algorithm using independent Gaussian (“white”) noise to propose a new state. The proposed new configuration of the  $N$  spins  $\tilde{\sigma}_i^n, i = 1, 2, \dots, N$  requires picking a random direction in the tangent plane, moving along that direction, and projecting back onto the sphere,

$$\tilde{\sigma}_i^n = \frac{\sigma_i^n + \epsilon \mathbf{v}_i^W}{\|\sigma_i^n + \epsilon \mathbf{v}_i^W\|}, \quad (8)$$

with  $\mathbf{v}_i^W = \mathbf{P}_{\sigma_i^n}^\perp(\mathbf{w}_i^n)$  is a projection of the three-dimensional normal random vector  $\mathbf{w}_i^n$  into the tangent plane of  $\sigma_i^n$ ,  $\mathbf{P}_x^\perp(y) = y - (x \cdot y)x$  or in matrix form  $(I - xx^T)y$ . Defining Hamiltonian

$$H(\vec{\sigma}) = \frac{1}{N} \sum_{i=1}^N \frac{N^2}{2} \|\sigma_{i+1} - \sigma_i\|^2 \quad (9)$$

with  $\sigma_{N+1} = \sigma_1$  for periodic boundary conditions, the accept probability

$$\alpha = 1 \wedge e^{-\beta[H(\vec{\sigma}^n) - H(\vec{\sigma}^n)]} \quad (10)$$

ensures sampling of the Gibbs distribution (1), where  $\vec{\sigma}^n$  and  $\vec{\sigma}^n$  are the  $3N$  vectors of the proposal components and current spin components, respectively. Symmetry in the proposal is crucial for (10) to be the correct accept probability to sample the Gibbs distribution. We discuss this in more detail, pointing out that symmetry is lacking when  $\mathbf{v}_i^W$  in the proposal is replaced with its correlated noise version next in Sec. IV.

By taking the lowest order term in  $\epsilon$  of the mean and added noise, the MH step is approximately equivalent to the Euler-like step

$$\sigma_i^{n+1} - \sigma_i^n \approx -\frac{1}{2} \beta \epsilon^2 \mathbf{P}_{\sigma_i^n}^\perp \left( \frac{\partial H}{\partial \sigma_i^n} \right) - \epsilon^2 \sigma_i^n + \epsilon \mathbf{P}_{\sigma_i^n}^\perp(\mathbf{w}_i^n).$$

Our previous work showed the trajectory-wise convergence as  $\epsilon \rightarrow 0$  of the MH dynamics to the corresponding Itô SDE

$$ds_i = \left[ \mathbf{P}_{s_i}^\perp(\Delta_N s_i) - \frac{2N}{\beta} s_i \right] dt + \sqrt{2\beta^{-1}N} \mathbf{P}_{s_i}^\perp(d\mathbf{W}_i) \quad (11)$$

under the time rescaling  $\delta t = \beta \varepsilon^2 / 2N$  where  $\mathbf{W}_i$  are three-dimensional Brownian motions and

$$\Delta_N \sigma_i = N^2(\sigma_{i+1} - 2\sigma_i + \sigma_{i-1})$$

is the discretized Laplace operator.

In the case of white noise, we point out that other projection operators could be used in place of  $P_x^\perp(y)$  above. The only requirement in the MH algorithm is that white noise is projected onto the tangent plane of  $\sigma_i^n$ . Two other natural choices would be  $\sigma_i^n \times \mathbf{w}_i^n$  and  $-\sigma_i^n \times (\sigma_i^n \times \mathbf{w}_i^n)$ , the later being equivalent to  $P_{\sigma_i^n}^\perp(\mathbf{w}_i^n)$  defined above; both produce white noise in the tangent plane. We will observe in Sec. IV B that this freedom is strongly related to the white noise setting and that care must be taken when moving to the colored noise case.

To see the equivalence of the two natural projection choices of the cross- and cross-cross-product in the white-noise case, we show that the limiting SDE systems for the MH dynamics produce the exact same Fokker-Planck equation in either case, so using either is justified. Define the  $3N \times 1$  vector of independent noises as

$$d\vec{W} = \langle d\mathbf{W}_x \, d\mathbf{W}_y \, d\mathbf{W}_z \rangle \quad (12)$$

with each  $N \times 1$  vector

$$d\mathbf{W}_q = \langle dW_{1,q} \dots dW_{N,q} \rangle \text{ for } q \in \{x, y, z\}$$

so that the  $3N$  (Itô) equations analogous to (11) are

$$d\vec{s} = PP^T \Delta_N \vec{s} \, dt - \frac{2N}{\beta} \vec{s} \, dt + \sqrt{2\beta^{-1}NP} d\vec{W}. \quad (13)$$

We consider two choices for the block-defined projection matrix  $P$  next. Note that both these projection matrices contribute the same factor  $-2\vec{s}$  to the Itô correction term,  $-2N\beta^{-1}\vec{s}$ , in the above SDE. For the single-spin projection  $\sigma_i \times d\mathbf{W}_i$ , the block-defined projection matrix is

$$P_1 = \begin{pmatrix} 0 & -Z & Y \\ Z & 0 & -X \\ -Y & X & 0 \end{pmatrix} \quad (14)$$

and the block-defined projection matrix for  $-\sigma_i \times (\sigma_i \times d\mathbf{W}_i)$  is

$$P_2 = \begin{pmatrix} I - X^2 & -XY & -XZ \\ -XY & I - Y^2 & -YZ \\ -XZ & -YZ & I - Z^2 \end{pmatrix}, \quad (15)$$

where each  $N \times N$  block matrix  $X, Y$  or  $Z$  are the diagonal matrices

$$Q = \begin{pmatrix} \sigma_{1,q} & 0 & \dots & 0 \\ 0 & \sigma_{2,q} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{N,q} \end{pmatrix}$$

for  $Q \in \{X, Y, Z\}$  with corresponding  $q \in \{x, y, z\}$ . The Fokker-Planck equation for (13) is

$$\begin{aligned} \partial_t \rho(\vec{s}, t) = & \sum_{i=1}^{3N} \partial_i [(PP^T \Delta_N \vec{s})_i \rho(\vec{s}, t)] + \frac{2N}{\beta} \sum_{i=1}^{3N} \partial_i [\vec{s}_i \rho(\vec{s}, t)] \\ & + \frac{N}{\beta} \sum_{i,j=1}^{3N} \partial_i \partial_j [(PP^T)_{ij} \rho(\vec{s}, t)]. \end{aligned} \quad (16)$$

Notice that this equation depends only on  $PP^T$ , which is identical for both  $P_1$  and  $P_2$ ,

$$P_1 P_1^T = P_2 P_2^T = P_2, \quad (17)$$

after using that  $\sigma_{i,x}^2 + \sigma_{i,y}^2 + \sigma_{i,z}^2 = 1$  for each  $i$ . Thus, both projections produce statistically equivalent trajectories in the white noise setting, and direct substitution can verify that (1) is an invariant measure for both (see Appendix B 2). The key point when taking colored noise instead of white noise that we will see in Sec. IV B is that *the covariance matrix for the noise and the projection matrix do not commute* and therefore  $PP^T$  does not appear isolated in the colored noise Fokker-Planck equation. The two projection matrices  $P_1$  and  $P_2$  produce statistically different ensembles.

We also point out that the Itô correction term in (13),  $-2N\beta^{-1}\vec{s}$ , is completely independent of choice of projection, the Stratonovich form of (13) being

$$d\vec{s} = PP^T \Delta_N \vec{s} \, dt + \sqrt{2\beta^{-1}NP} \circ d\vec{W}. \quad (18)$$

This fact remains true in the case of colored noise, that the Itô correction term depends only on the covariance matrix of the noise but not the choice of projection (see Sec. IV with details in Appendix A 1 c).

Our previous work also considered the continuum limit of the SDE system (11). Defining a lattice spacing  $\delta x = N^{-1}$  and taking a scaling of  $\beta = N^\gamma$  for  $\gamma$  sufficiently large to quench the noise (numerical simulations verified convergence for  $\beta \sim N^{3/2}$ ), we showed convergence to the (local, deterministic) PDE (4) under some regularity assumptions of the solution to the harmonic map heat flow equation. This convergence holds regardless of the number of spatial dimensions considered, provided we assume regularity of the solution to the corresponding harmonic map heat flow with domain  $\mathbb{T}^d$ ,  $d > 2$ . As mentioned in the Introduction, the regularity of the solution for  $d > 2$  is a delicate issue when considering the fixed  $\beta$  continuum limit to an SPDE, and one may not be guaranteed convergence in the case of white noise.

#### IV. COLORED NOISE

Using spatially correlated noise in the proposal of the MH algorithm to lead to regularized SPDEs in the continuum limit intuitively accounts for the fact that at smaller atomic scales, the true physical system cannot be further subdivided into infinity small units with independent fluctuations. A natural way to introduce correlations in the noise that decay with distance is to “color” the noise, requiring the power in the Fourier representation to decay with frequency. In the discrete setting, to form various covariance matrices satisfying our periodic boundary conditions, we use a periodic Fourier basis with power in each frequency mode that decays with rate  $\kappa$ . We again remind the reader that for ease of exposition we have set  $d = 1$  in this section, but extending to higher dimensions



is just a matter of using higher dimensional discrete Fourier transform machinery. However, in Sec. IV D below about SPDE limits, we will state the limiting equations for general dimension  $d$ .

Specifically we decompose an  $N \times N$  covariance matrix

$$\bar{C}_N = \phi \bar{D}^2 \phi^T \quad (19a)$$

with diagonal matrix  $\bar{D}_{jj} = \lambda_j = d_j^{-\kappa}$  with frequencies  $d_j$  defined as

$$d_j = \begin{cases} 1 & j = 1 \\ 2\pi(j-1) & 2 \leq j \leq \frac{N}{2} + 1 \\ 2\pi(j - \frac{N}{2} - 1) & \frac{N}{2} + 2 \leq j \leq N \end{cases} \quad (19b)$$

and the matrix of Fourier eigenvectors given by

$$\phi_{ij} = \begin{cases} 1 & j = 1 \\ \sqrt{2} \cos[\frac{1}{N} d_j(i-1)] & 2 \leq j \leq \frac{N}{2} \\ \cos[\frac{1}{N} d_j(i-1)] & j = \frac{N}{2} + 1 \\ \sqrt{2} \sin[\frac{1}{N} d_j(i-1)] & \frac{N}{2} + 2 \leq j \leq N \end{cases} \quad (19c)$$

with  $\sum_i \phi_{ij}^2 = N$  for each  $j$ . With this scaling, the eigenvectors  $\phi_{ij}$  converge to the discrete set of Fourier functions  $1, \sqrt{2} \cos(2\pi x), \sqrt{2} \cos(4\pi x), \dots$  and  $\sqrt{2} \sin(2\pi x), \sqrt{2} \sin(4\pi x), \dots$  as  $N \rightarrow \infty$ , forming an orthonormal set, with inner product of two functions defined as  $\int_0^1 f(x)g(x) dx$ . Also due to this scaling  $\text{Tr}(\bar{C}_N) = N \sum_{j=1}^N \lambda_j^2$ . Note that taking  $\kappa = 0$  creates equal power in all modes, reducing  $\bar{C}_N$  to a diagonal matrix with  $N$  on the diagonal representing uncorrelated “white” noise. Increasing  $\kappa$  increases the length scale of the covariance, broadening  $\bar{C}_N$  which is peaked along the diagonal.

For use in the MH algorithm, at time step  $n$ , we form three vectors  $\bar{C}_N^{1/2} \mathbf{w}_q^n$  for  $q \in \{x, y, z\}$  with the  $\mathbf{w}_q^n$  a set of vectors of independent uncorrelated standard Gaussian random variables. The vectors  $\bar{C}_N^{1/2} \mathbf{w}_q^n$  are independent for different  $q$  but spatially correlated with covariance matrices given by  $\bar{C}_N$ . This correlated noise is projected into the tangent plane of the corresponding spin, defining

$$\mathbf{v}_i^n = P_{\sigma_i^n}^\perp [(\bar{C}_N^{1/2} \mathbf{w}_x^n)_i, (\bar{C}_N^{1/2} \mathbf{w}_y^n)_i, (\bar{C}_N^{1/2} \mathbf{w}_z^n)_i].$$

The analogous proposal to (8) is

$$\tilde{\sigma}_i^n = \frac{\sigma_i^n + \varepsilon \mathbf{v}_i^n}{\|\sigma_i^n + \varepsilon \mathbf{v}_i^n\|}. \quad (20)$$

The first thing to note is that using  $\mathbf{v}_i^n$  in place of  $\mathbf{v}_i^W$  in the proposal creates a nonsymmetric proposal and therefore using the accept probability (10) no longer guarantees sampling of the Gibbs distribution (1). In particular, since our noise is now spatially correlated but our projections are completely local, the probability of undoing a particular rotation is not equal to the probability of doing it. In the white noise case, the tangent vector  $\mathbf{v}_i^W$  to get  $\sigma_i^n$  back from the proposal  $\tilde{\sigma}_i^n$  is unique and has the same magnitude as  $\mathbf{v}_i^W$ . Then  $\mathbb{P}(\sigma_i^n | \tilde{\sigma}_i^n) = \mathbb{P}(\tilde{\sigma}_i^n | \sigma_i^n)$  and since the tangent vectors  $\mathbf{v}_i^W$  are independent for different spins  $i$ , the entire proposal in the white noise case is symmetric,

$$\mathbb{P}(\tilde{\sigma}^n | \sigma^n) = \mathbb{P}(\sigma^n | \tilde{\sigma}^n).$$

In the colored noise case, the tangent vectors are correlated and the above symmetry requirement is no longer true. However, as the sphere is locally close to flat, intuitively the projected tangent vectors from  $\sigma_i^n$  and back from the proposal  $\tilde{\sigma}_i^n$  should be almost symmetric, though we observe that it depends upon the projection chosen as to how this asymmetric proposal manifests in the limit of  $\varepsilon \rightarrow 0$ . For the cross-product projection corresponding to  $P_1$ , the nonsymmetric terms appear in higher orders of  $\varepsilon$  and we conjecture they vanish taking similar limits of the (wrongly defined) MH algorithm as we did previously. We revisit this conjecture in Sec. IV C.

We discuss this limit of  $\varepsilon \rightarrow 0$ , arriving at the (Itô) SDE

$$d\vec{s} = P_1 \frac{C_N}{N} P_1^T \Delta_N \vec{s} dt - 2\beta^{-1} \frac{\text{Tr}(\bar{C}_N)}{N} \vec{s} dt + \sqrt{2\beta^{-1}} P_1 C_N^{1/2} d\vec{W} \quad (21)$$

next in Sec. IV A with details appearing in Appendix A. Then in Sec. IV B we discuss why the  $P_1$  projection matrix, corresponding to  $\sigma \times$ , has been selected. In Appendix B 2 we verify that the Gibbs distribution is the invariant measure of (21) by considering the Fokker-Planck equation for the equivalent Stratonovich SDE

$$d\vec{s} = P_1 \frac{C_N}{N} P_1^T \Delta_N \vec{s} dt + \sqrt{2\beta^{-1}} P_1 C_N^{1/2} \circ d\vec{W}. \quad (22)$$

Notice that for the case of uncorrelated noise,  $\kappa = 0$ , the matrix  $C_N$  reduces to a diagonal matrix with  $N$  on the diagonal. The SDE (21) therefore reduces to the white noise SDE (11) as  $\frac{1}{N} C_N$  reduces to the identity matrix,  $\frac{1}{N} \text{Tr}(\bar{C}_N) = N$  and  $C_N^{1/2} d\vec{W} = \sqrt{N} d\vec{W}$ .

### A. Limiting dynamics of Metropolis-Hastings

The idea behind the convergence is to equate one MH step to one Euler-Maruyama numerical integration step of the Itô SDE (21). Following [9,20] we consider the leading order in proposal size  $\varepsilon$  terms for the drift and diffusion of one MH step. At various points we drop higher order terms that are random variables, which are capable of taking on arbitrarily large values, but with small probability. To ensure a true asymptotic convergence, we bound the average pathwise error between MH and SDE trajectories themselves, not the probability distribution governed by a master equation, thus we have a strong, trajectory-wise, convergence result. In what follows, we heuristically explain obtaining the leading order terms for the drift and the diffusion. Expectations,  $\mathbb{E}_n[\cdot]$ , are conditioned on knowing the current MH spin configuration,  $\tilde{\sigma}^n$ . The details of properly bounding the error between the piecewise interpolated MH trajectory and the SDE trajectory are left to Appendix A.

The drift term of the SDE comes from the expectation of one MH step,

$$\mathbb{E}_n[\tilde{\sigma}^{n+1} - \tilde{\sigma}^n] = \mathbb{E}_n[(\tilde{\sigma}^n - \sigma^n)(1 \wedge e^{-\beta(H(\tilde{\sigma}^n) - H(\sigma^n))})], \quad (23)$$

where the elements of the proposal  $\tilde{\sigma}^n$  are each given by (20). Expanding this proposal for small  $\varepsilon$ , we obtain

$$\tilde{\sigma}_i^n - \sigma_i^n \approx \varepsilon \mathbf{v}_i^n - \frac{1}{2} \varepsilon^2 \|\mathbf{v}_i^n\|^2 \sigma_i^n. \quad (24)$$

We evaluate the expectation in (23) for the first term on the right-hand side of (24) first, then the second term.

For the expectation over the first term in the expansion (24), we have

$$\tilde{\sigma}^n - \bar{\sigma}^n \approx \varepsilon PC_N^{1/2} \tilde{w}^n$$

and proceed to compute

$$\mathbb{E}_n[\varepsilon PC_N^{1/2} \tilde{w}^n (1 \wedge e^{-\beta[H(\tilde{\sigma}^n) - H(\bar{\sigma}^n)]})]$$

using the first order expansion of  $H(\tilde{\sigma}^n) - H(\bar{\sigma}^n)$ , which is

$$\delta H \approx \varepsilon (\nabla H)^T PC_N^{1/2} \tilde{w}^n. \quad (25)$$

The first order term in the expansion of  $(1 \wedge e^{-\beta \delta H})$  is one, resulting in the expectation of  $\tilde{w}^n$  which is zero. The next order term comes from using the Lemma 2.4 in Ref. [20], which we state here for convenience.

*Lemma 1 ([20]).* For  $z \sim \mathcal{N}(0, 1)$ ,

$$\mathbb{E}[z(1 \wedge e^{az+b})] = ae^{\frac{a^2}{2}+b} \Phi\left(-\frac{b}{|a|} - |a|\right)$$

for any real constants  $a, b$ , and  $\Phi(\cdot)$  is the CDF for the standard normal random variable.

We apply this lemma on the expectation for a single component of  $\tilde{w}^n$  and corresponding coefficient of  $\delta H$  and then take the expectation over the remaining components of  $\tilde{w}^n$  with further approximations detailed in Appendix A 1 b. The result is the same as taking  $1 \wedge e^{-\beta \delta H} \approx 1 - \beta \delta H$  when  $\delta H < 0$  and 1 otherwise while also assuming  $\delta H$  is mean zero so that each case happens with probability 1/2. Thus,

$$\mathbb{E}_n[\tilde{w}^n (1 \wedge e^{-\beta[H(\tilde{\sigma}^n) - H(\bar{\sigma}^n)]})] \approx \mathbb{E}_n[\tilde{w}^n (\frac{1}{2} \delta H)]$$

and the only expectation that remains is  $\mathbb{E}_n[\tilde{w}^n (\tilde{w}^n)^T] = I$ , the identity matrix. Therefore,

$$\mathbb{E}_n[\tilde{w}^n (1 \wedge e^{-\beta \delta H})] \approx -\varepsilon \frac{\beta}{2} [(\nabla H)^T PC_N^{1/2}]^T \quad (26)$$

and

$$\mathbb{E}_n[\varepsilon PC_N^{1/2} \tilde{w}^n (1 \wedge e^{-\beta \delta H})] \approx -\varepsilon^2 \frac{\beta}{2} PC_N P^T \nabla H. \quad (27)$$

Returning to (23), we consider the second term in the expansion (24), and compute

$$\mathbb{E}_n[\frac{1}{2} \varepsilon^2 \|\mathbf{v}_i^n\|^2 \sigma_i^n (1 \wedge e^{-\beta \delta H})].$$

Here it is convenient to take  $\mathbf{v}_i^n = P_{\sigma_i^n}^\perp \mathbf{u}_i$  and write the three components of  $\mathbf{u}_i$  in terms of the decomposition of the matrix  $\bar{C}_N$  defined in (19) as

$$u_{i,q} = \sum_{j=1}^N \lambda_j \phi_{ji} w_{j,q}^n \quad \text{for } q \in \{x, y, z\}. \quad (28)$$

Unlike above, the first term in the expansion of  $(1 \wedge e^{-\beta \delta H})$  gives nonzero expectation, which is

$$\mathbb{E}_n\left[\frac{1}{2} \varepsilon^2 \|\mathbf{v}_i^n\|^2 \sigma_i^n\right] = \varepsilon^2 \sigma_i^n \sum_{j=1}^N \lambda_j^2 \phi_{ji}^2.$$

We further notice that  $\sum_{j=1}^N \lambda_j^2 \phi_{ji}^2$  is equivalent to  $\frac{1}{N} \text{Tr}(\bar{C}_N)$  for each  $i$  as a result of the chosen Fourier basis to represent  $\bar{C}_N$ . Therefore,

$$\mathbb{E}_n\left[-\frac{1}{2} \varepsilon^2 \|\mathbf{v}_i^n\|^2 \sigma_i^n\right] = -\varepsilon^2 \frac{1}{N} \text{Tr}(\bar{C}_N) \sigma_i^n. \quad (29)$$

In vector form, combining the above with (27), we have that (23) to leading order in  $\varepsilon$  is

$$\mathbb{E}_n[\tilde{\sigma}^{n+1} - \bar{\sigma}^n] \approx -\varepsilon^2 \frac{\beta}{2} PC_N P^T \nabla H - \varepsilon^2 \frac{1}{N} \text{Tr}(\bar{C}_N) \bar{\sigma}^n. \quad (30)$$

The diffusion part of the SDE is the leading-order in  $\varepsilon$  term of the mean-zero noise,

$$\tilde{\sigma}^{n+1} - \bar{\sigma}^n - \mathbb{E}_n[\tilde{\sigma}^{n+1} - \bar{\sigma}^n] \approx \varepsilon PC_N^{1/2} \tilde{w}^n. \quad (31)$$

Recall from above, that the expectation of  $PC_N^{1/2} \tilde{w}^n$  was zero; this is the leading order noise term. Combining with the drift, we have that one step of the MH algorithm to leading order is

$$\begin{aligned} \tilde{\sigma}^{n+1} - \bar{\sigma}^n \approx & -\varepsilon^2 \frac{\beta}{2} PC_N P^T \nabla H \\ & - \varepsilon^2 \frac{1}{N} \text{Tr}(\bar{C}_N) \bar{\sigma}^n + \varepsilon PC_N^{1/2} \tilde{w}^n. \end{aligned} \quad (32)$$

Defining a rescaling of time as  $\delta t = \varepsilon^2 \beta / 2$  the above is

$$\begin{aligned} \tilde{\sigma}^{n+1} - \bar{\sigma}^n \approx & -\delta t P \frac{1}{N} C_N P^T \Delta_N \bar{\sigma}^n \\ & - \delta t \frac{2}{N \beta} \text{Tr}(\bar{C}_N) \bar{\sigma}^n + \sqrt{\frac{2 \delta t}{\beta}} PC_N^{1/2} \tilde{w}^n, \end{aligned} \quad (33)$$

where we have used that for Hamiltonian (9),

$$\nabla H = \frac{1}{N} \Delta_N \bar{\sigma}^n. \quad (34)$$

Equation (33) is one step of the the Euler-Maruyama method for the Stratonovich SDE (22).

The trajectory-wise convergence of the MH dynamics to the solution of (22) is summarized in the following statement and proved in Appendix A 3.

*Theorem 1.* Define the piecewise constant interpolation of the MH dynamics as  $\bar{\sigma}(t)$ ,

$$\bar{\sigma}(t) = \bar{\sigma}^n \text{ for } n \delta t \leq t < (n+1) \delta t, \quad (35)$$

where  $\delta t = \frac{\beta \varepsilon^2}{2}$  is the time step size of the MH dynamics, and  $\bar{\sigma}(t)$  is the solution to the SDE system (22) with initial conditions  $\bar{\sigma}(0) = \bar{\sigma}^0$  and  $\|\sigma_i(0)\| = 1, 1 \leq i \leq N$ . If the proposal noise in the MH step is generated by the same  $3N$  Wiener processes in (22) as

$$\varepsilon \mathbf{w}_i^n = \sqrt{2 \beta^{-1}} [\mathbf{W}_i((n+1) \delta t) - \mathbf{W}_i(n \delta t)],$$

for  $i = 1 \dots N$ , then we have the following strong convergence result:

$$\mathbb{E}\left[\sup_{0 \leq \tau \leq T} \|\bar{\sigma}(\tau) - \bar{\sigma}(\tau)\|^2\right] \leq c_1 \sqrt{\delta t} \exp(c_2 T) \quad (36)$$

for any  $T \in (0, \infty)$ , where  $c_1$  and  $c_2$  are functions of  $N, \beta, T, \text{Tr}(C_N)$  and independent of the choice of  $\delta t$ .

This convergence result holds regardless of the projection matrix  $P$ . What remains to be determined is if the SDE (22) has the Gibbs distribution (1) as its invariant measure.

### B. Choosing a projection

Having established convergence of the MH dynamics, we are left to show that the system of SDEs (22) has the Gibbs distribution (1) as its invariant measure. This SDE is in the form of (2) with the matrix  $B = PC_N^{1/2}$  being nonconstant. For generic nonconstant  $B$  in (2), the Gibbs distribution (1) is no longer an invariant measure, however in a few special cases it is. For example, in the case of the white noise SDE (11) using either  $\sigma_i \times dW_i$  or  $-\sigma_i \times (\sigma_i \times dW_i)$ , so that  $B = P_1$  or  $B = P_2$ , it is, as we show by direct computation in Appendix B 2. However, when considering colored noise, only the projection of the form  $\sigma_i \times (\cdot)$  corresponding to  $B = P_1 C_N^{1/2}$ , and not  $B = P_2 C_N^{1/2}$ , has (1) as an invariant measure, as we show by direct computation in Appendix B 2. Unlike the white noise case, since the colored noise matrix and projection matrix do not commute,  $P_1 C_N P_1^T \neq P_2 C_N P_2^T$ , and the two projections of the noise into the tangent plane produce statistically different trajectories. We explore this idea further numerically in Sec. V, showing that the cross-cross-projection samples something further and further from the Gibbs distribution as the noise becomes more correlated.

### C. Convergence of the invariant measure

In this section, we justify a statement said earlier in Sec. IV that the nonsymmetric terms in the MH proposal (20) appear in higher orders of the proposal size  $\varepsilon$ . In particular, we show that the invariant measure of the MH dynamics with colored noise in the proposal and cross-product projection is close to the desired invariant Gibbs distribution, converging to it in the  $\varepsilon \rightarrow 0$  limit. We apply similar ideas to those of Ref. [36] which consider invariant measures of numerical approximations of SDE solutions. We start with Dynkin's formula over one time step of the SDE, and then replace the integral over the SDE solution with the MH solution, bounding the difference. Summing over multiple time steps and noticing a telescoping series, we show the long-time average over the MH solution converges to the average over the invariant measure of the SDE, which is the Gibbs distribution. Therefore, as in Ref. [36], we find that the difference between the invariant measures is the same order of magnitude as the error between the MH dynamics and the solution to the SDE (22) on a finite time interval, given by (36).

Our goal is to show that the long-time average of a  $C^\infty$  test function  $\varphi$

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \frac{1}{n} \sum_{k=0}^{n-1} \varphi(\bar{\sigma}^k) \right],$$

where  $\bar{\sigma}^k$  is the  $k$ th MH step with the inaccurate accept rate (10) and any projection to form  $\mathbf{v}_i^n$  in (20), converges to the stationary average  $\bar{\varphi}$  with respect to the invariant measure  $\mu$  of the SDE (22) with corresponding projection,

$$\bar{\varphi} = \int \varphi(\bar{\sigma}) \mu(\bar{\sigma}) d\bar{\sigma}. \quad (37)$$

We build on the fact that the MH algorithm has a unique stationary distribution, that is not the Gibbs distribution, and that the SDE has a unique stationary measure  $\mu$  because the generator  $\mathcal{L}$  of the SDE (22) is hypoelliptic; its second-order

term is

$$\sum_{i,j} (PC_N P^T)_{ij} \partial_i \partial_j = (P^T D)^T C_N (P^T D) - \sum_{i,j} \partial_i (PC_N P^T)_{ij} \partial_j,$$

where  $D$  is the diagonal matrix with  $\bar{D}$  repeated three times along the diagonal and the system of vector fields  $(P^T D)^T C_N (P^T D)$  covers  $\mathbb{T}(\mathbb{S}^2)^N$  as in Ref. [15] and the second term on the right hand side is first order. In the special case that the cross-product projection matrix  $P_1$  in (14) is used, then the SDE has the known invariant measure of the Gibbs measure  $\mu$  in (1). Our argument will therefore show that the MH algorithm with cross-product projection samples a distribution that converges to the Gibbs measure as the proposal size  $\varepsilon \rightarrow 0$ .

We start with Dynkin's formula [37] for the SDE (22), with generator  $\mathcal{L}$ , over a time step  $\delta t$ ,

$$\begin{aligned} & \mathbb{E}[\psi(\bar{\sigma}((k+1)\delta t))] - \mathbb{E}[\psi(\bar{\sigma}(k\delta t))] \\ &= \mathbb{E} \left[ \int_{k\delta t}^{(k+1)\delta t} \mathcal{L}\psi(\bar{\sigma}(t)) dt \right]. \end{aligned} \quad (38)$$

Consider that  $\psi$  solves a Poisson equation for  $C^\infty$  test function  $\varphi$ ,

$$\mathcal{L}\psi = \varphi - \bar{\varphi}, \quad (39)$$

where the stationary average  $\bar{\varphi}$  is defined in (37). Using (39) in the right-hand side of (38), we have that

$$\begin{aligned} & \mathbb{E}[\psi(\bar{\sigma}((k+1)\delta t))] - \mathbb{E}[\psi(\bar{\sigma}(k\delta t))] \\ &= \mathbb{E} \left[ \int_{k\delta t}^{(k+1)\delta t} \varphi(\bar{\sigma}(t)) dt \right] - \bar{\varphi} \delta t. \end{aligned} \quad (40)$$

The integral term can be bounded by

$$\left| \mathbb{E} \left[ \int_{k\delta t}^{(k+1)\delta t} \varphi(\bar{\sigma}(t)) dt - \varphi(\bar{\sigma}(k\delta t)) \delta t \right] \right| \leq c \delta t^2 \quad (41)$$

for some constant  $c$  independent of  $\delta t$  by Riemann sum approximations of integrals. From Theorem 1, the difference between the SDE solution  $\bar{\sigma}(k\delta t)$  and the MH step  $\bar{\sigma}^k$  is bounded by

$$\mathbb{E}[\|\bar{\sigma}^k - \bar{\sigma}(k\delta t)\|] \leq c_3 \delta t^{1/4},$$

and therefore for smooth test functions

$$|\mathbb{E}[\varphi(\bar{\sigma}(k\delta t))] - \mathbb{E}[\varphi(\bar{\sigma}^k)]| \leq c_4 \delta t^{1/4}. \quad (42)$$

Using bounds (41) and then (42) we have that (40) can be written as

$$\begin{aligned} & \mathbb{E}[\psi(\bar{\sigma}((k+1)\delta t))] - \mathbb{E}[\psi(\bar{\sigma}(k\delta t))] \\ &= \mathbb{E}[\varphi(\bar{\sigma}^k)] \delta t - \bar{\varphi} \delta t + e_1, \end{aligned}$$

where  $|e_1| \leq c \delta t \delta t^{1/4}$  for some constant  $c$  independent of  $\delta t$ . Rearranging and dividing by  $\delta t$  we have that

$$\mathbb{E}[\varphi(\bar{\sigma}^k)] - \bar{\varphi} = \frac{1}{\delta t} \mathbb{E}[\psi(\bar{\sigma}((k+1)\delta t)) - \psi(\bar{\sigma}(k\delta t))] + e_2,$$

where  $|e_2| \leq c \delta t^{1/4}$  for some constant  $c$  independent of  $\delta t$ . Summing over  $n$  values of  $k$  and dividing by  $n$  we

have that

$$\begin{aligned} & \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}[\varphi(\vec{\sigma}^k)] - \bar{\varphi} \\ &= \frac{1}{n\delta t} \sum_{k=0}^{n-1} \mathbb{E}[\psi(\vec{s}((k+1)\delta t)) - \psi(\vec{s}(k\delta t))] + e_2, \end{aligned} \quad (43)$$

which has a telescoping sum on the right-hand side. By defining  $T = n\delta t$  the above is equivalent to

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}[\varphi(\vec{\sigma}^k)] - \bar{\varphi} = \frac{1}{T} \mathbb{E}[\psi(\vec{s}(T)) - \psi(\vec{s}(0))] + e_2. \quad (44)$$

Recall that  $\psi$  is the unique solution to the Poisson equation (39) therefore it is smooth because  $\varphi$  is smooth. Indeed, the theory of hypoelliptic operators is precisely such that  $\mathcal{L}u \in C^\infty$  implies  $u \in C^\infty$ ; see Ref. [38] or chap. XI of Ref. [39]. Since we are operating on a compact space overall,  $\psi$  is thus bounded and the convergence result follows. Thus, the  $1/T$  term goes to zero as  $T \rightarrow \infty$  ( $n \rightarrow \infty$ ). We therefore conclude that the MH long-time average converges to the stationary average with respect to the SDE invariant measure  $\bar{\varphi}$  as  $\delta t \rightarrow 0$  ( $\varepsilon \rightarrow 0$ ) with order  $\delta t^{1/4}$  and as  $n \rightarrow \infty$ , and the following convergence results holds:

**Theorem 2.** Define  $\vec{\sigma}^n$  as the  $n$ th step of the MH dynamics with colored noise proposal given in (20), either the cross- or cross-cross-projection, accept rate given in (10), and let  $\mu(\vec{s})$  be the invariant measure of the corresponding SDE (22) with the same projection. Then

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}[\varphi(\vec{\sigma}^k)] - \int \varphi(\vec{s}) \mu(\vec{s}) d\vec{s} \right| \leq \frac{c_1}{n\delta t} + c_2 \delta t^{1/4}$$

for time step  $\delta t = \varepsilon^2 \beta / 2$  and constants  $c_1$  and  $c_2$  independent of  $n$ , and  $\delta t$ .

**Remark 1.** A nearly identical argument can be used to show that the invariant measure for the SDE with the  $P_2$  projection will converge to that of the SDE with the  $P_1$  projection as the colored noise converges to white noise, thus both SDEs sample the Gibbs measure. In other words, as  $\kappa \rightarrow 0$  the covariance matrix  $C \rightarrow I$  in a uniform sense in our definition of (19) as an operator on  $\ell^2$  (and hence smooth) functions  $(\mathbb{S}^2)^N$ .

#### D. A new nonlocal SPDE limit

In this section, we discuss the extension of the nonlocal SPDE (5) to the case  $\mathbb{T}^d \rightarrow \mathbb{S}^2$  with  $d > 2$  obtained by taking the limit as  $N \rightarrow \infty$  (with  $\beta$  constant) of the SDE (22) and remark briefly on properties of the corresponding solutions. In particular, formally taking the limit of (22), we arrive at a nonlocal stochastic version of the harmonic map heat flow equation given by

$$\begin{aligned} d\sigma &= \{-\sigma \times [M_\kappa(D)](\sigma \times \Delta\sigma)\}dt + \sigma \\ &\quad \times \{\mathcal{F}^{-1}[m(k)]^{-\kappa} \circ dW(k)\}, \end{aligned} \quad (45)$$

where we let  $m(k) = 2\pi k$  if  $|k| \neq 0$ , and  $m(k) = 1$  for  $k = 0$ ,  $\mathcal{F}$  is the Fourier transform on  $\mathbb{T}^d$ ,  $M_\kappa(D)$  is the Fourier

multiplier such that

$$M_\kappa(D)f = \mathcal{F}^{-1}[|m(k)|]^{-2\kappa} \mathcal{F}f$$

and  $dW(k)$  are a set of independent standard Gaussian noises for each corresponding Fourier mode in frequency space. Note that many other forms of the covariance structure could easily work here, such as  $m(k) = \langle k \rangle = \sqrt{1 + |2\pi k|^2}$ . Also note that we can write

$$[(I - \Delta)^{-\kappa} f](x) = \int K_\kappa(x, y) f(y) dy$$

with the integral kernel given by

$$K_\kappa(x, y) = \frac{1}{(2\pi)^d} \sum_{k \in \mathbb{Z}^d} \int e^{-i2\pi k(x-y)} \langle k \rangle^\kappa.$$

Then, if  $\kappa$  is chosen such that  $M_\kappa(D)$  is trace class with a weight relating to the regularity required ( $\int K(x, y) dx, \int K(x, y) dy < \infty$  as well as integrals of derivatives of  $K$ ), we can use canonical results on stochastic PDEs coupled with existence arguments for quasilinear heat equations. We will follow somewhat the ideas in Refs. [2,40] for stochastic PDEs with multiplicative noise (mostly in the context of motivating the Itô formulation in the former and for using energy estimates to handle degenerate SPDE models in the latter). For the key energy estimates on the deterministic piece, we cite the general theory of well-posedness for quasilinear heat equations developed in Chap. 15 of Ref. [41]. For possible extensions to nontrace class covariance structure, see the recent work of Ref. [5] where a renormalization is proposed. It will be a topic of further work to explore the place of our colored noise model within this context.

Using the regularity of the colored noise, we provide a brief outline of existence for solutions to (45) in Appendix C. However, as the results are fairly standard with sufficiently regular noise, we proceed with a detailed numerical study of convergence of the MH model and dynamics.

#### V. NUMERICAL RESULTS

In this section, we perform numerical simulations to support our convergence results and demonstrate the discussed differences when using different projections. All the simulations are from the one-dimensional periodic lattice  $\mathbb{T}^1$  to the unit sphere  $\mathbb{S}^2$ . The MH dynamics are simulated as explained in Sec. IV. To numerically solve the SDE (21), written in the Itô form, we use the stochastic Euler's method combined with a normalizing step to project the spins back onto the sphere after each time step.

We start by showing a trajectory-wise comparison in Fig. 1(a) of the MH dynamics and the SDE dynamics generated utilizing the same random noise for the proposal in the MH as the diffusion term in the SDE. Each spin is plotted on the same sphere, with lines connecting nearest neighbors. Figures 1(b) and 1(c) show the strong order of convergence for the error between the MH algorithm and the SDE (22) with respect to the time step size  $\delta t$ , for which the equivalent MH proposal size is  $\varepsilon = \sqrt{2\delta t/\beta}$ . The error is calculated at



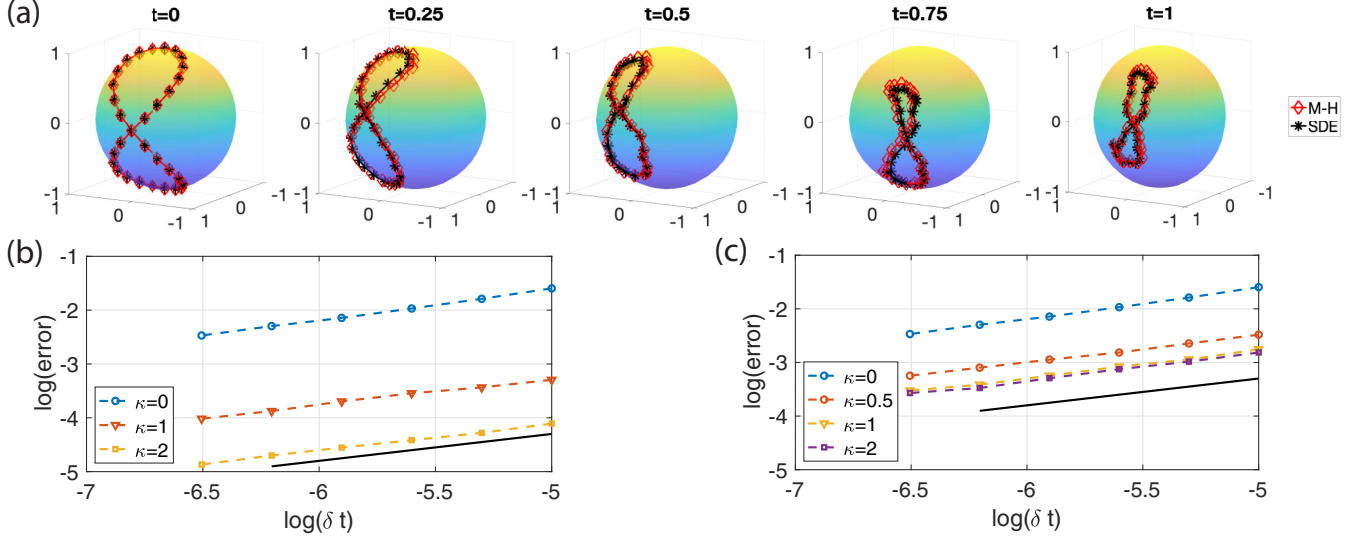


FIG. 1. (a) Dynamics of MH algorithm and SDE (22) at the indicated values of time,  $t$  (recall the relationship between SDE time step and MH proposal size:  $\delta t = 2\beta\epsilon^2$ ). Parameters: number of spins  $N = 32$ , inverse temperature  $\beta = 10$ , order of eigenvalues  $\kappa = 1$ , time step size  $\delta t = 10^{-5}$ . (b) Strong order of convergence with respect to time step size for the error between the MH algorithm and SDE (22), both using the cross-projection matrix (14). The solid black line with slope 1/2 indicates the order is approximately that given in Theorem 1. (c) The same as (b) but using the cross-cross-projection matrix (15). (b) and (c) parameters:  $N = 16$ ,  $\beta = 5$ , error calculated at time  $T = 0.05$ , and averaged over 400 simulations.

fixed time  $T$  as

$$\mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \left\| \sigma_i^n - s_i(n\delta t) \right\|^2 \right], \quad (46)$$

where the expectation is taken over multiple realizations. The numerical convergence order is approximately  $\frac{1}{2}$ , supporting Theorem 1 as a tight bound on the error regardless of choosing the cross-projection matrix (14) or the cross-cross-projection matrix (15).

Next we show the effect of the different projection matrices on the invariant measure of the SDE system (22). Since the desired invariant measure is high-dimensional, we instead plot the empirical cumulative distribution function (cdf) of the energy over time. Figure 2(a) shows that for the case of white noise,  $\kappa = 0$ , utilizing either the cross-projection matrix (14) or the cross-cross-projection matrix (15) results in indistinguishable invariant distributions of the energy; both versions have the Gibbs distribution as an invariant measure. However, when coloring the noise by increasing  $\kappa$ , it is only the cross-projection matrix (14) that maintains an energy distribution indistinguishable from the white noise case. Figure 2(b) supports that the color noise SDE (22) with the cross-projection matrix (14) is ergodic with respect to the correct Gibbs distribution, despite being the limit of our incorrect MH scheme in Sec. IV. Figure 2(c) shows that the SDE system with the cross-cross-projection matrix (15) has lower energy on average as the correlations in the colored noise increase with increasing  $\kappa$ .

To further illuminate this interaction of the projection matrix and the correlated noise, we look how each term in the SDE affects the energy of system when in equilibrium. The

energy,  $H$  given by (9), evolves according to the Itô SDE

$$dH = \frac{1}{N} \sum_{i=1}^N N^2 (s_{i+1} - s_i) \cdot (ds_{i+1} - ds_i) + N\beta^{-1} \text{Tr}(C_N^{1/2} P^T A P C_N^{1/2}) dt, \quad (47)$$

where  $A$  is the tridiagonal matrix with 2 on the diagonal and  $-1$  on the sub- and superdiagonals (taking into account periodicity),  $ds_i$  is given by (21), and  $s_{N+1} = s_1$ . Note that since  $\text{Tr}(XY) = \text{Tr}(YX)$  for two  $n \times n$  matrices  $X$  and  $Y$ , the trace term

$$\begin{aligned} \text{Tr}(P C_N^{1/2} C_N^{1/2} P^T A) &= \text{Tr}(P P^T C_N A) \\ &= \text{Tr}(P P^T \phi \bar{D}^2 \phi^T A) = \text{Tr}(P P^T \phi \phi^T \bar{D}^2 A) \\ &= \text{Tr}(P \bar{D}^2 A) = 4N \sum_{i=1}^N \lambda_i^2 \end{aligned}$$

is a constant independent of the choice of projection matrix. We therefore ignore this term and proceed to decompose  $ds_i$  given by (21) over one  $\delta t$  time step of numerical integration as

$$s_i^{n+1} - s_i^n = p_i^n \delta t - 2\beta^{-1} \frac{\text{Tr}(\bar{C}_N)}{N} s_i^n \delta t + q_i^n \sqrt{\delta t}, \quad (48)$$

where we define

$$\bar{p}^n = P \frac{1}{N} C_N P^T \Delta_N \bar{s}^n \quad \text{and} \quad \bar{q}^n = P C_N^{1/2} \bar{w}^n$$

as well as take  $s_{N+1}^n = s_1^n$ ,  $p_{N+1}^n = p_1^n$ , and  $q_{N+1}^n = q_1^n$  for the periodic boundary conditions. The trace term in (48) is also of a form independent of the choice of projection matrix. Therefore, to illuminate the interaction of the projection matrix and the correlated noise we consider only the contributions to (47),

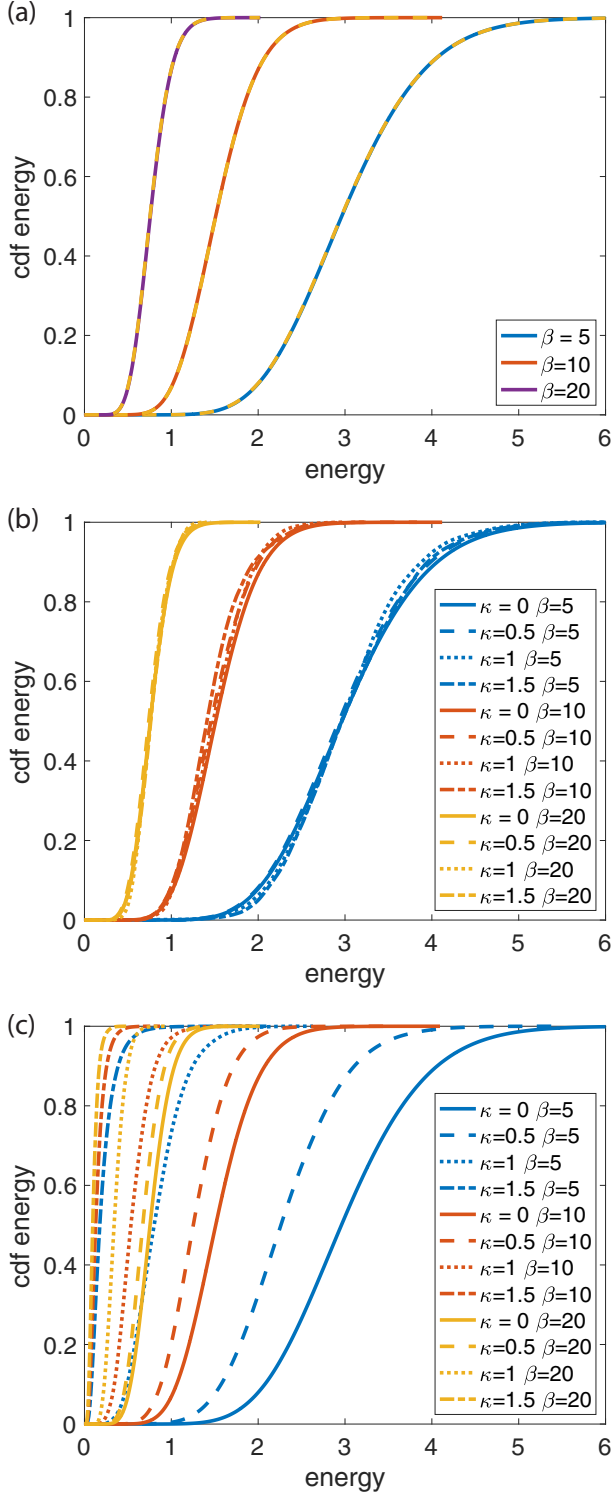


FIG. 2. (a) Energy distribution in equilibrium for the indicated values of  $\beta$  when “white noise”  $\kappa = 0$  is used. The solid lines are for the cross-product projection, the yellow dashed lines are for the cross-cross-product projection. They agree entirely; Gibbs is being sampled in all cases. (b) For the cross-product, at the indicated values of  $\kappa$ , the same distribution is being sampled at each of the three values of  $\beta = 5, 10$ , and  $20$ . (c) For the cross-cross-product, different distributions are being sampled for different values of  $\kappa$ , consistent with Gibbs not being the invariant measure of the SDE when  $\kappa \neq 0$ . Parameters:  $N = 16$ ,  $\delta t = 10^{-4}$ , and  $10^5$  different time points.

the change in energy, given by the  $\vec{p}^n$  and  $\vec{q}^n$  terms over each time step of the numerical integration of the SDE, calculated as

$$\delta H_{\text{drift}}^n = \sum_{i=1}^N (s_{i+1}^n - s_i^n) \cdot (\vec{p}_{i+1}^n - \vec{p}_i^n) \delta t \quad (49)$$

and

$$\delta H_{\text{noise}}^n = N \sqrt{\frac{2}{\beta}} \sum_{i=1}^N (s_{i+1}^n - s_i^n) \cdot (\vec{q}_{i+1}^n - \vec{q}_i^n) \sqrt{\delta t}. \quad (50)$$

In Fig. 3 we plot the distribution of  $\delta H_{\text{drift}}^n$  and  $\delta H_{\text{noise}}^n$  for both the  $P_1$  (cross-product) and  $P_2$  (cross-cross-product) projections over the course of one simulation using each of the indicated values of  $\kappa$  to form  $\tilde{C}_N$ . We see that as  $\kappa$  increases, the differences between these distributions increase, consistent with Fig. 2(c) showing more deviation from the Gibbs distribution with increasing  $\kappa$ . This difference is more pronounced in the deterministic drift contribution to the energy,  $\delta H_{\text{drift}}^n$ , than the diffusion contribution,  $\delta H_{\text{noise}}^n$ . It suggests the random-walk nature of the dynamics remains relatively unaffected by the choice of projection, while the cross-cross-projection produces long tails to lower values of  $\delta H_{\text{drift}}^n$  possibly explaining the shift in average energy to lower energies seen in Fig. 2(c).

In Fig. 4 we verify convergence of the SDE system to the SPDE (5). First, in Fig. 4(a), for just the deterministic drift part of this system, we show convergence of the finite difference ODE approximation of the nonlocal PDE ( $\beta^{-1} = 0$ ). We compute the error at fixed time  $T$  between each coarser scale,  $N = N_c$ , with the finest scale,  $N = N_f$ , as

$$\frac{1}{N_c} \sum_{i=1}^{N_c} \left\| s_i^{\text{Coarse}}(T) - s_{1+(i-1)\frac{N_f}{N_c}}^{\text{Fine}}(T) \right\|^2. \quad (51)$$

Then, in Fig. 4(b) we shown the strong convergence of the SDE, taking the expectation of the above error over realizations. Note the convergence rate even for the white noise case of  $\kappa = 0$ , which is not guaranteed if more than one spatial dimension of this SPDE was considered due to the potential breakdown of regularity of the deterministic solution in that case. The deterministic convergence of order 4 is twice that of the noisy system, which is approximately order 2.

Last, we look at some of the behavior of the new nonlocal (deterministic) PDE. In Fig. 5(a) we show the evolution toward equilibrium of the spins for different values of  $\kappa$  highlighting the different time scales. By considering the covariance operator as a fractional Laplacian, acting similarly to the harmonic map heat flow equation, we conjecture the time rescaling being related to the diffusion time scaling of the underlying nonlocal heat equation

$$u_t = M_\kappa(D)\Delta u,$$

which decays to its equilibrium on the timescale  $e^{-\lambda_1^{2(1-\kappa)}}$  for  $\lambda_1$  the first nontrivial eigenvalue of the Laplacian on  $\mathbb{T}^d$ . The nonlocal form of the operator we consider here does not immediately present a leading order linear operator of this form as occurs in the cross-cross-projection, however we will see that this timescale still arises in Fig. 5. Figures 5(c) and

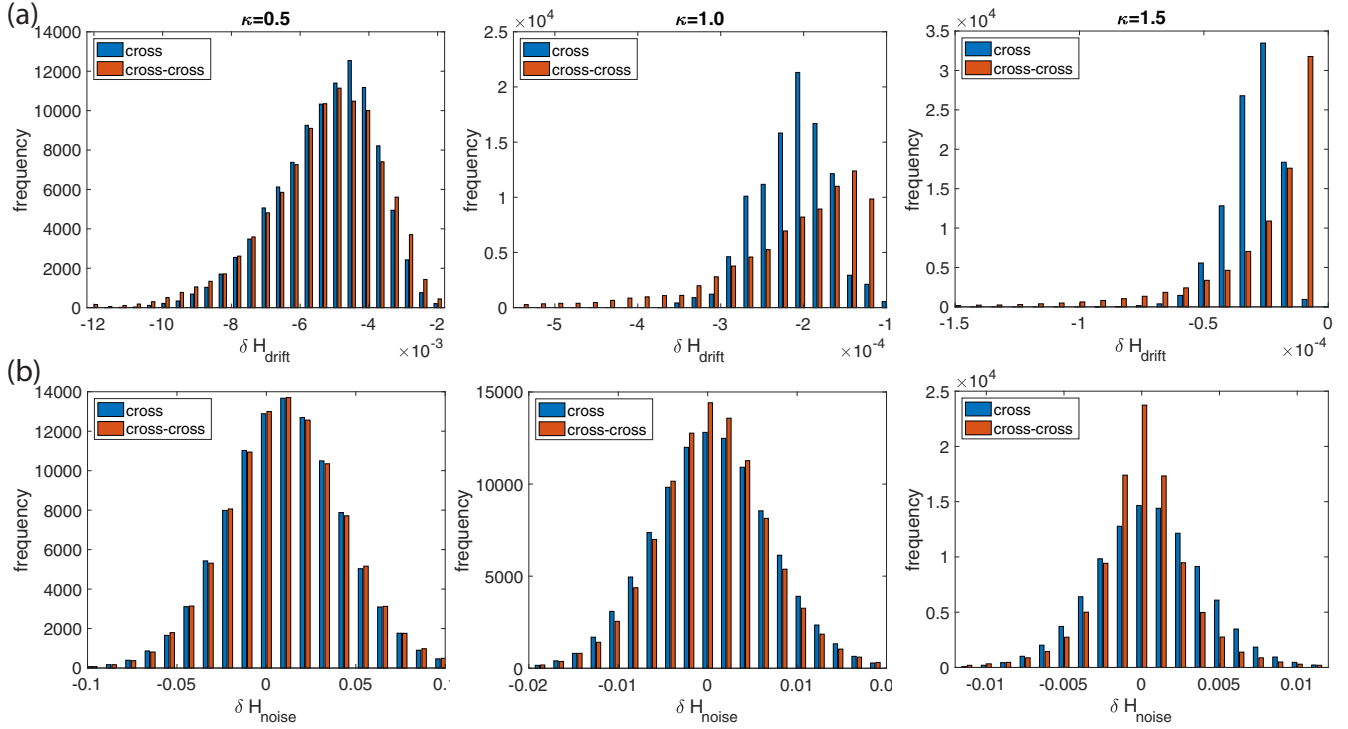


FIG. 3. Comparison of the effect on the energy for the cross- and the cross-cross-projection matrix in SDE (22). (a) The empirical histogram of  $\delta H_{\text{drift}}^n$  from (49) taken at each time point of one simulation of the SDE in equilibrium for the indicated value of  $\kappa$ . (b) Same as (a) but for  $\delta H_{\text{noise}}^n$  from (50). Parameters:  $N = 16$ ,  $\delta t = 10^{-4}$ ,  $\beta = 10$ , and  $10^5$  different time points.

5(d) also show the effect of this time rescaling when looking at the evolution of the energy of the system.

## VI. CONCLUSIONS AND DISCUSSION

We establish here a new stochastic partial differential equation as the limit of a set of sampling algorithms where the proposal is taken with spatially correlated colored noise, thereby deriving a mesoscopic model of fluctuations for spin systems in a principled way. The geometric nature of our system means that the nonlocal form of the drift arises in a manner that we have not seen before in the literature. In order to ensure that the system samples the desired Gibbs measure, we have to be careful with the manner by which we project the noise into the geometric setting. Specifically, we show using a cross-product projection samples the Gibbs measure while a cross-cross-product projection samples an invariant measure that is shifted to lower energy than the Gibbs measure. This shift increases as the correlation length scale of the noise is increased, and is shown numerically to be related to the deterministic effect on the energy of the system rather than fluctuations in the energy. In addition to finding convergence rates, numerical simulations are also used to show that the nonlocal drift term of the new SPDE exhibits the same timescales for relaxation to equilibrium as a fractional Laplacian, thereby acting similarly to the harmonic map heat flow equation. Future work will involve considering other geometries beyond the sphere, performing a more careful analysis of the resulting SDE and SPDE systems following for instance

the recent developments on geometric renormalization tools in Ref. [5].

## ACKNOWLEDGMENTS

Y.G. and J.L.M were supported in part by NSF Applied Math Grant DMS-1312874 and NSF CAREER Grant DMS-1352353. J.C.M. was partially supported by DMS-1613337 from the National Science Foundation and a Simons Travel grant. The authors thank Amarjit Budhiraja for helpful conversations during the preparation of this work. J.L.M. acknowledges Duke University and Mathematical Sciences Research Institute (MSRI) where he was hosted during part of the completion of this project.

## APPENDIX A: CONVERGENCE PROOF OF MH TO SDE

This Appendix compliments Sec. IV A, containing the details of the derivation and proof of Theorem 1. In Sec. A 1 we derive the expansion of the mean drift of one MH step, (30), and bound the remainder terms. In Sec. A 2 we bound the error of the diffusion approximation of one MH step, (31). Then in Sec. A 3 we argue the existence and uniqueness of the solution to SDE (22) and complete the proof of Theorem 1 giving the ideas for bounding the error between the MH and SDE dynamics. Throughout this Appendix we use  $c$  or  $c(\cdot)$  to indicate a positive constant, potentially different constant every time the symbol is used, independent of the parameters we are bounding quantities in, either  $\varepsilon$  or  $\delta t$ . The functional representation  $c(\cdot)$  indicates precisely which model parameters the constant is dependent on.

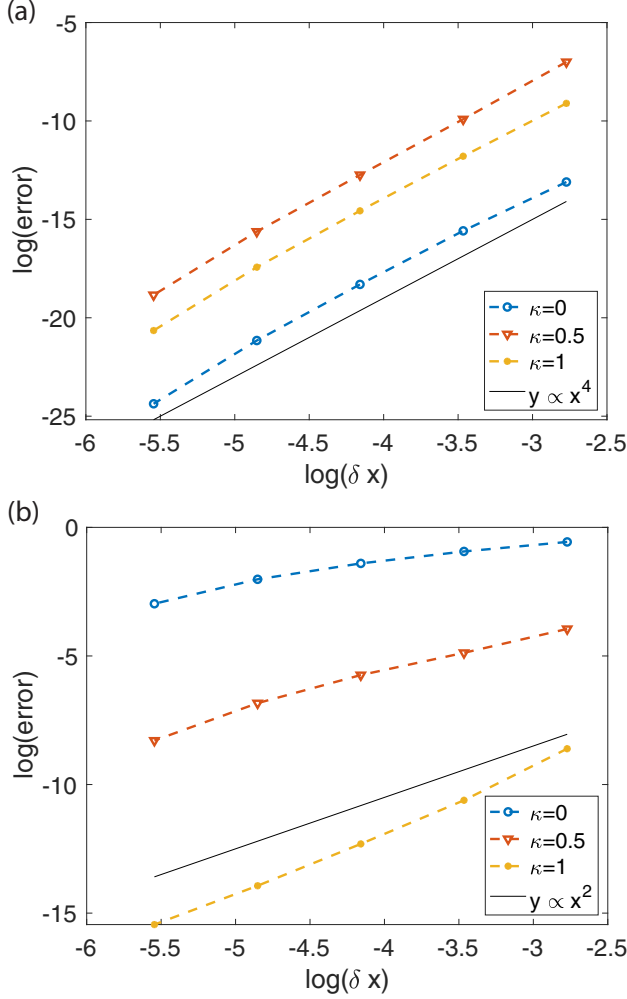


FIG. 4. (a) Convergence plot for the deterministic finite difference approximation of the nonlocal PDE (22) with  $\beta^{-1} = 0$ . (b) Convergence plot for the stochastic finite difference approximation of the nonlocal SPDE (22) with  $\beta = 5$ , averaged over 100 simulations. In both panels, dynamics are simulated until  $T = 0.125$  with  $\delta t = \frac{1}{2}\delta x^2$ .

### 1. Expectation of one MH step

In this section, we calculate the leading order in  $\varepsilon$  terms of the expectation of one MH step, Eq. (30), and bound the expectation of the error of each expansion and approximation we utilize. The leading terms which determine the dynamics are of size  $\varepsilon^2$  while the mean-squared error is bounded by the next order terms of size  $\varepsilon^6$ . For brevity, throughout this section we drop the superscript  $n$  on all terms. The first error term, denoted  $\tilde{r}_1$  and defined coordinatewise as  $(\mathbf{r}_1)_i \in \mathbb{R}^3$ ,  $i = 1, \dots, N$ , arises from utilizing the expansion (24) of the MH proposal as

$$(\mathbf{r}_1)_i = \mathbb{E}_n[d_i(1 \wedge e^{-\beta[H(\tilde{\sigma}) - H(\sigma)]})], \quad (\text{A1})$$

where

$$d_i \equiv \tilde{\sigma}_i - \left( \sigma_i + \varepsilon \mathbf{v}_i - \frac{\varepsilon^2}{2} \|\mathbf{v}_i\|^2 \sigma_i \right) \quad (\text{A2})$$

is the difference between the true proposal and its approximation. Other error terms will come from computing the expectations of the two terms of the expansion of the proposal. Namely, by writing

$$\begin{aligned} & \mathbb{E}_n[(\tilde{\sigma}_i - \sigma_i)(1 \wedge e^{-\beta[H(\tilde{\sigma}) - H(\sigma)]})] \\ &= \mathbb{E}_n[\varepsilon \mathbf{v}_i(1 \wedge e^{-\beta[H(\tilde{\sigma}) - H(\sigma)]})] \\ & \quad - \mathbb{E}_n[\frac{1}{2}\varepsilon^2 \|\mathbf{v}_i\|^2 \sigma_i(1 \wedge e^{-\beta[H(\tilde{\sigma}) - H(\sigma)]})] + (\mathbf{r}_1)_i, \end{aligned} \quad (\text{A3})$$

we now can move to defining the error term denoted  $\tilde{r}_2$ , which bounds the difference of the expectations of the  $\mathbf{v}_i$  term above when using the actual difference in the Hamiltonians of the proposal and current step versus its leading order approximation (25). Namely, we have

$$\begin{aligned} \tilde{r}_2 &= \mathbb{E}_n[\varepsilon PC_N^{1/2} \tilde{w}(1 \wedge e^{-\beta[H(\tilde{\sigma}) - H(\sigma)]})] \\ & \quad - \mathbb{E}_n[\varepsilon PC_N^{1/2} \tilde{w}(1 \wedge e^{-\beta\varepsilon(PC_N^{1/2} \tilde{w})^T \nabla H})]. \end{aligned} \quad (\text{A4})$$

This error term  $\tilde{r}_2$  is bound in Sec. A 1 b. We further bound the error of the approximation (27) in the main text by bounding the term denoted by  $\tilde{r}_3$

$$\mathbb{E}_n[\varepsilon PC_N^{1/2} \tilde{w}(1 \wedge e^{-\beta\delta H})] = -\frac{1}{2}\varepsilon^2 \beta PC_N P^T \nabla H + \tilde{r}_3 \quad (\text{A5})$$

to be specifically defined later but that results from approximations applying Lemma 1. A similar bound on the second, order  $\varepsilon^2$ , term in (A3) is handled in Sec. A 1 c, defining

$$\begin{aligned} (\mathbf{r}_4)_i &= \mathbb{E}_n[-\frac{1}{2}\varepsilon^2 \|\mathbf{v}_i\|^2 \sigma_i(1 \wedge e^{-\beta[H(\tilde{\sigma}) - H(\sigma)]})] \\ & \quad - \mathbb{E}_n[-\frac{1}{2}\varepsilon^2 \|\mathbf{v}_i\|^2 \sigma_i]. \end{aligned} \quad (\text{A6})$$

We also show in Sec. A 1 c that this term is the Itô correction term to the drift between the Stratonovich and Itô representations of the SDE, Eqs. (22) and (29), respectively.

Combining the to be presented error bounds in Eqs. (A11), (A16), (A25), and (A28) below, we arrive at the global bound of Eq. (30), namely, by defining

$$\begin{aligned} & \mathbb{E}_n[(\tilde{\sigma}^n - \sigma^n)(1 \wedge e^{-\beta[H(\tilde{\sigma}^n) - H(\sigma^n)]})] \\ &= -\frac{1}{2}\beta\varepsilon^2 PC_N P^T \nabla H - \varepsilon^2 \frac{1}{N} \text{Tr}(\tilde{C}_N) \sigma + \sum_{k=1}^4 \tilde{r}_k, \end{aligned}$$

the error is bounded by

$$\mathbb{E} \left[ \left\| \sum_{k=1}^4 \tilde{r}_k \right\|^2 \right] \leq c N^{13} \beta^6 [\text{Tr}(C_N)]^3 \varepsilon^6, \quad (\text{A7})$$

where we will repeatedly use the very crude bound

$$\|\nabla H\| \lesssim N^2$$

from (34).

*Remark 2.* Intuitively the term  $\nabla H$  should have a better bound in some probability sense, but the current bound might be the best we could hope for so far since we are not able to specify a probability distribution for  $H$  with the wrong MH step setting.

#### a. Bounding the proposal expansion remainder

In this section we bound the error of the remainder terms,  $(\mathbf{r}_1)_i$ , defined in (A1).



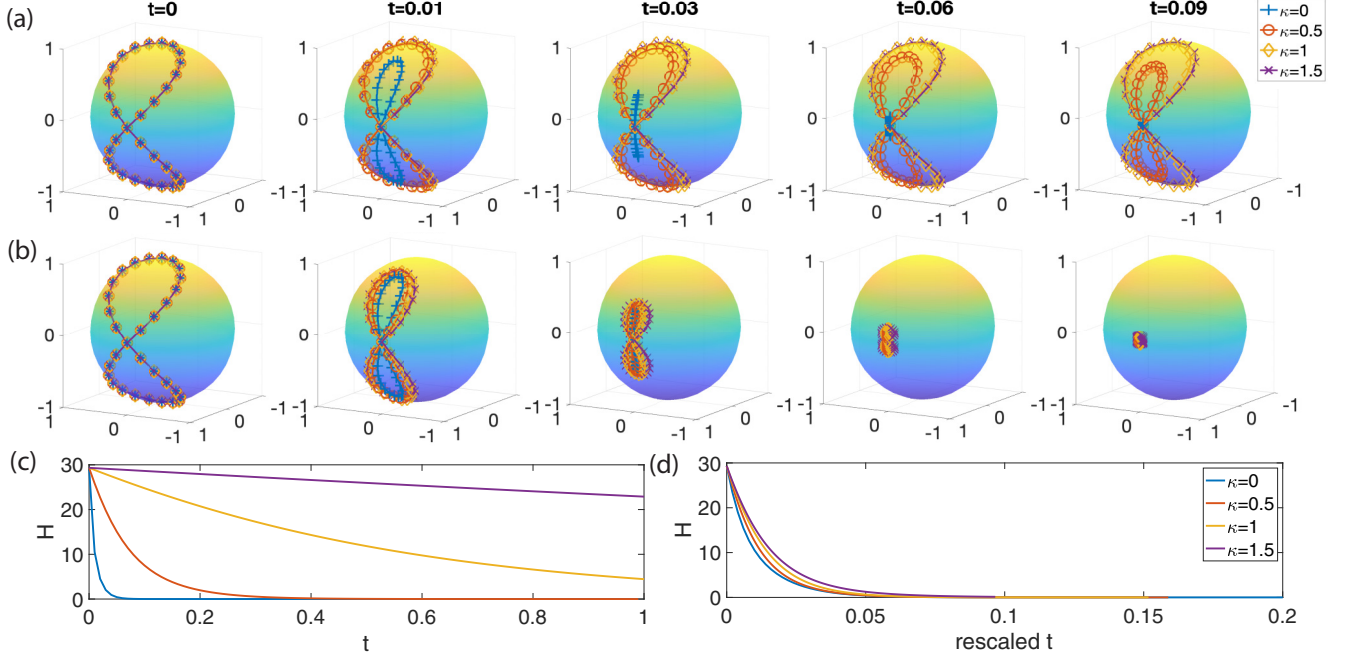


FIG. 5. (a) Dynamics of the nonlocal PDE [Eq. (5) with  $\beta^{-1} = 0$ ] at the indicated values of time,  $t$ , with original timescale and (b) with rescaled time  $\tilde{t} = (2\pi)^{2\kappa} t$  for  $\kappa \in \{0, 0.5, 1, 1.5\}$ . (c) Evolution of the energy with the original timescale and (d) with rescaled time. Legend applies to all panels. Parameters: number of spins  $N = 32$ , original time step size  $\delta t = 10^{-5}$ .

We begin by bounding  $d_i$  in (A2) by first bounding the size of the tangent vector  $\tilde{v} = PC_N^{1/2} \tilde{w}$ . Since the projection matrix  $P$  acts either as the cross-product, or the cross-cross-product, the magnitude  $\|Px\|$  is smaller than  $\|x\|$  for each single spin vector, and we have that

$$\|\tilde{v}\|^2 = \|PC_N^{1/2} \tilde{w}\|^2 \leq \|C_N^{1/2} \tilde{w}\|^2 = (\tilde{w})^T C_N \tilde{w}. \quad (\text{A8})$$

The expectation of  $(\tilde{w})^T C_N \tilde{w}$  is equivalent to the expectation of  $N \sum_{q \in \{x,y,z\}} (w_q)^T \bar{D}^2 w_q$  with the way we have defined  $\bar{C}_N$  in (19). Since each component of  $w$  is independent and identically distributed,

$$\mathbb{E} \left[ \sum_{q \in \{x,y,z\}} (w_q)^T \bar{D}^2 w_q \right] = \mathbb{E} \left[ 3 \sum_{i=1}^N \lambda_i^2 w_{ix}^2 \right].$$

Bounding the expectation of  $\|\tilde{v}\|^{2k}$  for any positive integer  $k$ , we have that

$$\begin{aligned} \mathbb{E}[\|\tilde{v}\|^{2k}] &\leq \mathbb{E} \left[ \left( 3N \sum_{i=1}^N \lambda_i^2 w_{ix}^2 \right)^k \right] \\ &= \mathbb{E} \left[ \sum_{i=1}^N N^k 3^k \lambda_i^{2k} w_{ix}^{2k} + \text{cross-terms with } w_{i,x}^{2l} w_{j,x}^{2(k-l)} \right] \\ &\leq \mathbb{E}[w_{i,x}^{2k}] \left( 3N \sum_{i=1}^N \lambda_i^2 \right)^k = [(2k-1)!!] \left( 3N \sum_{i=1}^N \lambda_i^2 \right)^k \end{aligned}$$

since  $\mathbb{E}[w_{i,x}^{2l} w_{j,x}^{2(k-l)}] = (2l-1)!!(2k-2l-1)!! \leq \mathbb{E}[w_{i,x}^{2k}] = (2k-1)!!$ . Therefore

$$\mathbb{E}[\|\tilde{v}\|^{2k}] \leq [(2k-1)!!] \left( 3 \sum_{i=1}^N N \lambda_i^2 \right)^k \leq c[\text{Tr}(\bar{C}_N)]^k. \quad (\text{A9})$$

Now  $d_i$  can be bounded as a function of  $v_i$  and the remainder from the Taylor expansion of  $f(x) = (1+x)^{-1/2} = 1 - \frac{x}{2} + \frac{3}{8}(1+\xi)^{-5/2}x^2$  for some  $\xi \in [0, \infty)$ . Writing

$$(1 + \varepsilon^2 \|v_i^n\|^2)^{-1/2} = 1 - \frac{\varepsilon^2}{2} \|v_i^n\|^2 + \eta_i^n$$

we see that

$$\|\eta_i^n\| \leq \frac{3}{8} \varepsilon^4 \|v_i^n\|^4$$

and

$$d_i^n = -\frac{\varepsilon^3}{2} \|v_i^n\|^2 v_i^n + (\sigma_i^n + \varepsilon v_i^n) \eta_i^n. \quad (\text{A10})$$

We bound

$$\begin{aligned} \sum_{i=1}^N \mathbb{E}[\|d_i\|^{2k}] &\leq \sum_{i=1}^N 4^k \mathbb{E} \left[ \left\| -\frac{\varepsilon^3}{2} \|v_i\|^2 v_i \right\|^{2k} \right] \\ &\quad + 4^k \mathbb{E}[\|(\sigma_i + \varepsilon v_i^n) \eta_i\|^{2k}], \end{aligned}$$

then bound the first term

$$\sum_{i=1}^N 4^k \mathbb{E} \left[ \left\| -\frac{\varepsilon^3}{2} \|v_i\|^2 v_i \right\|^{2k} \right] \leq \sum_{i=1}^N \varepsilon^{6k} \mathbb{E}[\|v_i\|^{6k}]$$

and the second term

$$4^k \mathbb{E}[\|(\boldsymbol{\sigma}_i + \varepsilon \mathbf{v}_i) \boldsymbol{\eta}_i\|^{2k}] \leq 4^k (\mathbb{E}[\|\boldsymbol{\sigma}_i + \varepsilon \mathbf{v}_i\|^{4k}])^{1/2} \{\mathbb{E}[(\boldsymbol{\eta}_i)^{4k}]\}^{1/2}$$

to see that

$$\begin{aligned} \sum_{i=1}^N \mathbb{E}[\|\mathbf{d}_i\|^{2k}] &\leq \sum_{i=1}^N \varepsilon^{6k} E[\|\mathbf{v}_i\|^{6k}] \\ &\quad + 4^k (\mathbb{E}[4^k + 4^k \varepsilon^{4k} \|\mathbf{v}_i\|^{4k}])^{1/2} \\ &\quad \times \left\{ \left( \frac{3}{8} \right)^{4k} \varepsilon^{16k} \mathbb{E}[\|\mathbf{v}_i\|^{16k}] \right\}^{1/2} \\ &\leq c(k) [\text{Tr}(C_N)]^{3k} \varepsilon^{6k}. \end{aligned}$$

Together with the fact that  $(1 \wedge e^{-\beta[H(\tilde{\sigma}) - H(\sigma)])} \leq 1$  we have that

$$\mathbb{E}[\|\tilde{\mathbf{r}}_1\|^2] \leq \sum_{i=1}^N \mathbb{E}[\|\mathbf{d}_i\|^2] \leq cN [\text{Tr}(C_N)]^3 \varepsilon^6. \quad (\text{A11})$$

### b. The drift term

Here we bound  $\tilde{\mathbf{r}}_2$  defined in (A4), bounding the error of approximating the change in the Hamiltonian between the current step and the proposal as in Eq. (25). This further allows us to establish the leading order approximation (26). We bound the errors  $\tilde{\mathbf{r}}_3$  and  $\tilde{\mathbf{r}}_4$

We start with the approximation (25), a combination of a Taylor expansion in  $\varepsilon$  and the approximation of  $\tilde{\sigma} \approx \sigma + \varepsilon \tilde{\mathbf{v}}$ , the error of which is

$$\begin{aligned} g &\equiv \sum_{i=1}^N \frac{\partial H}{\partial \boldsymbol{\sigma}_i} \cdot \mathbf{f}_i + N \sum_{i=1}^N (\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i) \cdot (\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i) \\ &\quad - \frac{1}{2} N \sum_{i=1}^N (\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i) \cdot (\tilde{\boldsymbol{\sigma}}_{i+1} - \boldsymbol{\sigma}_{i+1} + \tilde{\boldsymbol{\sigma}}_{i-1} - \boldsymbol{\sigma}_{i-1}), \end{aligned} \quad (\text{A12})$$

where the error  $\mathbf{f}_i$  is from the first-order Taylor expansion of the proposal,

$$\mathbf{f}_i \equiv \tilde{\boldsymbol{\sigma}}_i - (\boldsymbol{\sigma}_i + \varepsilon \mathbf{v}_i). \quad (\text{A13})$$

To bound  $\mathbf{f}_i$ , note that  $\mathbf{f}_i = \mathbf{d}_i - \frac{\varepsilon^2}{2} \|\mathbf{v}_i\|^2 \boldsymbol{\sigma}_i$ , and therefore for positive integer  $k$ ,

$$\begin{aligned} \sum_{i=1}^N \mathbb{E}[\|\mathbf{f}_i\|^{2k}] &\leq \sum_{i=1}^N 4^k \left\{ \left( \frac{\varepsilon^2}{2} \right)^{2k} \mathbb{E}[\|\mathbf{v}_i\|^{4k}] + \mathbb{E}[\|\mathbf{d}_i\|^{2k}] \right\} \\ &\leq c(k) \{ [\text{Tr}(C_N)]^{2k} \varepsilon^{4k} + [\text{Tr}(C_N)]^{3k} \varepsilon^{6k} \} \\ &\leq c(k) [\text{Tr}(C_N)]^{2k} \varepsilon^{4k}. \end{aligned} \quad (\text{A14})$$

With the bound (A14), we proceed to bound  $g$  in (A12) by bounding its three summations,

$$\begin{aligned} \mathbb{E}[|g|^4] &\leq c \left\{ \mathbb{E} \left[ \left| \sum_{i=1}^N \frac{\partial H}{\partial \boldsymbol{\sigma}_i} \cdot \mathbf{f}_i \right|^4 \right] \right. \\ &\quad + \mathbb{E} \left[ \left| \frac{N}{2} \sum_{i=1}^N (\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i) \cdot (\tilde{\boldsymbol{\sigma}}_{i+1} - \boldsymbol{\sigma}_{i+1} + \tilde{\boldsymbol{\sigma}}_{i-1} - \boldsymbol{\sigma}_{i-1}) \right|^4 \right] \\ &\quad \left. + \mathbb{E} \left[ \left| N \sum_{i=1}^N (\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i) \cdot (\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i) \right|^4 \right] \right\} \\ &\leq c(k, N) N^8 \sum_i (\mathbb{E}[\|\mathbf{f}_i\|^4] + \mathbb{E}[\|\tilde{\boldsymbol{\sigma}}_i - \boldsymbol{\sigma}_i\|^8]). \end{aligned}$$

We then conclude that

$$\mathbb{E}[|g|^4] \leq cN^9 \varepsilon^8 \{2[\text{Tr}(C_N)]^4 + \varepsilon^8 [\text{Tr}(C_N)]^8\}. \quad (\text{A15})$$

Together with the fact that  $1 \wedge x$  is 1-Lipschitz, we can now bound the remainder  $\tilde{\mathbf{r}}_2$  appearing in (A4) as

$$\begin{aligned} \mathbb{E}[\|\tilde{\mathbf{r}}_2\|^2] &\leq \mathbb{E}[\|\varepsilon PC_N^{1/2} \tilde{\mathbf{w}} g\|^2] \\ &\leq \varepsilon^2 \{ \mathbb{E}[\|PC_N^{1/2} \tilde{\mathbf{w}}\|^4] \}^{1/2} \{ \mathbb{E}[|g|^4] \}^{1/2} \\ &\leq c\varepsilon^2 \text{Tr}(\tilde{C}_N) N^{9/2} \varepsilon^4 [\text{Tr}(\tilde{C}_N)]^2. \end{aligned} \quad (\text{A16})$$

We now proceed to compute the leading order term of

$$\mathbb{E}_n[\varepsilon \tilde{\mathbf{v}} (1 \wedge e^{-\beta \delta H})] = \varepsilon PC_N^{1/2} \mathbb{E}_n[\tilde{\mathbf{w}} (1 \wedge e^{-\beta \delta H})].$$

Specifically, we derive the approximation (26) and bound the error of each approximation used. The calculation utilizes Lemma 2.4 in Ref. [20] as stated in Lemma 1.

For each component of the noise  $w_{i,q}$  with  $i = 1, \dots, N$  and  $q \in \{x, y, z\}$ , we compute the expectation in two steps,

$$\mathbb{E}_n[w_{i,q} (1 \wedge e^{-\beta \delta H})] = \mathbb{E}_n[\mathbb{E}_n[w_{i,q} (1 \wedge e^{-\beta \delta H}) | \tilde{\mathbf{w}} \setminus w_{i,q}],$$

first taking the expectation over  $w_{i,q}$  using the above lemma, then over the remainder of the components of  $\tilde{\mathbf{w}}$ . To apply Lemma 1 for the first expectation, take

$$\begin{aligned} a &= a_{i,q} = -\beta \varepsilon ((\nabla H)^T PC_N^{1/2})_{i,q}, \\ b &= b_{i,q} = -\beta \varepsilon (PC_N^{1/2} \tilde{\mathbf{w}})^T \nabla H - a_{i,q} w_{i,q}, \end{aligned} \quad (\text{A17})$$

leaving the need to calculate the error denoted

$$\tilde{\mathbf{r}}_3 = \varepsilon PC_N^{1/2} \tilde{\mathbf{r}}_3 \quad (\text{A18})$$

with

$$(\tilde{\mathbf{r}}_3)_{i,q} = a_{i,q} \mathbb{E}_n \left[ e^{\frac{a_{i,q}^2}{2} + b_{i,q}} \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} - |a_{i,q}| \right) \right]. \quad (\text{A19})$$

We approximate

$$e^{\frac{a_{i,q}^2}{2} + b_{i,q}} \approx 1 \quad (\text{A20})$$

and bound the error of this approximation,

$$(\tilde{r}_{3,1})_{i,q} \equiv \mathbb{E}_n \left[ \left( e^{\frac{a_{i,q}^2}{2} + b_{i,q}} - 1 \right) \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} - |a_{i,q}| \right) \right], \quad (\text{A21})$$

next.

To bound  $(\tilde{r}_{3,1})_{i,q}$ , we use that for  $z \sim \mathcal{N}(\mu, \varsigma^2)$ ,

$$\mathbb{E}[e^z - 1]^k = \mathbb{E}[|e^z - 1|^k \mathbb{1}(z \leq 2)] + \mathbb{E}[|e^z - 1|^k \mathbb{1}(z > 2)], \quad (\text{A22})$$

where the indicator function  $\mathbb{1}$  return one if the statement is true and zero otherwise. To bound the first term in (A22), since  $|e^z - 1| \leq e^2 z$  for  $z \leq 2$ , we have that

$$\mathbb{E}[|e^z - 1|^k \mathbb{1}(z \leq 2)] \leq e^{2k} \mathbb{E}[z^k].$$

To bound the second term in (A22), we note that  $|e^z - 1|^k \leq e^{kz}$  when  $z > 2$ , and if  $2 + \mu + k\varsigma^2 \geq 1$  then we have that

$$\begin{aligned} \mathbb{E}[e^{kz} \mathbb{1}(z > 2)] &= \int_2^\infty \frac{1}{\sqrt{2\pi\varsigma^2}} e^{kz} e^{-\frac{(z-\mu)^2}{2\varsigma^2}} dz \\ &= e^{k\mu + \frac{k^2\varsigma^2}{2}} \int_{2+\mu+k\varsigma^2}^\infty \frac{1}{\sqrt{2\pi\varsigma^2}} e^{-\frac{x^2}{2\varsigma^2}} dx \end{aligned}$$

under the change of variables  $x = z - \mu - k\varsigma^2$ . This Gaussian integral is bounded by the exponential integral as

$$\int_{2+\mu+k\varsigma^2}^\infty e^{-\frac{x^2}{2\varsigma^2}} dx \leq \int_{2+\mu+k\varsigma^2}^\infty e^{-\frac{x}{2\varsigma^2}} dx = 2\varsigma^2 e^{-\frac{2+\mu+k\varsigma^2}{2\varsigma^2}}$$

and  $e^{-(2+\mu+k\varsigma^2)/2\varsigma^2} \leq e^{-1/2\varsigma^2}$ . Therefore, we arrive at the bound

$$\mathbb{E}[e^{kz} \mathbb{1}(z > 2)] \leq e^{k\mu + \frac{k^2\varsigma^2}{2}} \sqrt{\frac{2}{\pi}} \varsigma e^{-\frac{1}{2\varsigma^2}}.$$

Notice that  $\frac{a_{i,q}^2}{2} + b_{i,q} \sim \mathcal{N}(g_1\epsilon^2, g_2\epsilon^2)$ , where

$$\begin{aligned} g_1 &= \frac{1}{2}\beta^2 [(\nabla H)^T PC_N^{1/2}]_{i,q}^2 \\ g_2 &= (\nabla H)^T PC_N P^T \nabla H - a_{i,q}^2. \end{aligned}$$

Therefore the condition  $2 + \mu + k\varsigma^2 = 2 + g_1\epsilon^2 + kg_2\epsilon^2 \geq 1$  is met when we take  $\epsilon$  small enough, and applying the above derived bounds for the two terms in (A22), noting that for  $k = 2$ ,  $\mathbb{E}[z^k] = \mu^2 + \varsigma^2$ , we arrive at

$$\begin{aligned} \mathbb{E} \left[ \left( e^{\frac{a_{i,q}^2}{2} + b_{i,q}} - 1 \right)^2 \right] &\leq e^8 g_2 \epsilon^2 + \sqrt{\frac{2g_2}{\pi}} \epsilon e^{-\frac{1}{2g_2\epsilon^2}} \\ &\leq c\epsilon^2 (\nabla H)^T PC_N P^T \nabla H \end{aligned}$$

as the term  $e^{-\frac{1}{2g_2\epsilon^2}}$  decays faster than any polynomial of  $\epsilon$  as  $\epsilon \rightarrow 0$ . The bound

$$|(\tilde{r}_{3,1})_{i,q}|^2 \leq c\epsilon^2 (\nabla H)^T PC_N P^T \nabla H \quad (\text{A23})$$

follows.

We return to bounding (A19) using approximation (A20) and consider

$$\mathbb{E}_n \left[ \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} - |a_{i,q}| \right) \right].$$

Since both  $a_{i,q}$  and  $b_{i,q}$  are both proportional to  $\epsilon$ , the ratio  $b_{i,q}/|a_{i,q}|$  is large relative to  $|a_{i,q}|$  and we approximate

$$\mathbb{E}_n \left[ \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} - |a_{i,q}| \right) \right] \approx \mathbb{E}_n \left[ \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} \right) \right].$$

We bound the error of this approximation,

$$(\tilde{r}_{3,2})_{i,q} = \mathbb{E}_n \left[ \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} - |a_{i,q}| \right) \right] - \mathbb{E}_n \left[ \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} \right) \right]$$

by noting that

$$\left| \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} - |a_{i,q}| \right) - \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} \right) \right| \leq \frac{1}{\sqrt{2\pi}} |a_{i,q}|,$$

therefore

$$|(\tilde{r}_{3,2})_{i,q}|^2 \leq c\beta^2 \epsilon^2. \quad (\text{A24})$$

We calculate

$$\mathbb{E}_n \left[ \Phi \left( -\frac{b_{i,q}}{|a_{i,q}|} \right) \right] = \frac{1}{2}$$

by noting that for  $z \sim \mathcal{N}(0, \varsigma^2)$ ,

$$\mathbb{E}[\Phi(z)] = \mathbb{E} \left[ \left( \Phi(z) - \frac{1}{2} \right) + \frac{1}{2} \right] = \frac{1}{2}.$$

Retracing our steps, we see that

$$\mathbb{E}[w_{i,q} (1 \wedge e^{-\beta[H(\vec{\sigma}) - H(\vec{\sigma})]})] \approx \frac{a_{i,q}}{2}$$

and (26) follows.

Thus,  $\tilde{r}_3$ , the error of the approximation in (26), is bounded as

$$\mathbb{E}_n[\epsilon PC_N^{1/2} w (1 \wedge e^{-\beta\epsilon(PC_N^{1/2} w)^T \nabla H})] - \left( -\epsilon^2 \frac{\beta}{2} PC_N P^T \nabla H \right),$$

then

$$\mathbb{E}_n[\|\tilde{r}_3\|^2] \leq cN\epsilon^6 \beta^6 \text{Tr}(C_N)(\|\nabla H\|^4 + \|\nabla H\|^6) \quad (\text{A25})$$

so that its components involve  $a_{i,q}$ , a term of size  $\epsilon$ , times the error accumulated in the approximations bounded by  $(\tilde{r}_{3,1})_{i,q}$  in (A23) and  $(\tilde{r}_{3,2})_{i,q}$  in (A24).

### c. The Itô correction term

Here we consider the approximation

$$\mathbb{E}_n \left[ \frac{1}{2} \epsilon^2 \|\mathbf{v}_i\|^2 \sigma_i (1 \wedge e^{-\beta[H(\vec{\sigma}) - H(\vec{\sigma})]}) \right] \approx \mathbb{E}_n \left[ \frac{1}{2} \epsilon^2 \|\mathbf{v}_i\|^2 \sigma_i \right]$$

of the second term on the right-hand side of Eq. (A3), bounding the error term given in (A6). This includes computing the right-hand side of (29) and showing that it corresponds to the Itô correction of the Stratonovich SDE (22).

First we approximate  $H(\vec{\sigma}) - H(\vec{\sigma})$  by  $\delta H$  given in (25) and then compute

$$\mathbb{E}_n \left[ -\frac{1}{2} \epsilon^2 \|\mathbf{v}_i\|^2 \sigma_i (1 \wedge e^{-\beta\delta H}) \right] \approx \mathbb{E}_n \left[ -\frac{1}{2} \epsilon^2 \|\mathbf{v}_i\|^2 \sigma_i \right], \quad (\text{A26})$$

finding it has the same value regardless if the cross-product projection or the cross-cross-product projection is used to obtain  $\mathbf{v}_i$ . With  $u_{i,q}$  defined in (28), consider the cross-product projection,  $\mathbf{v}_i = \boldsymbol{\sigma}_i \times \mathbf{u}_i$ , then

$$\begin{aligned} \mathbb{E}_n \left[ -\frac{1}{2} \varepsilon^2 \|\mathbf{v}_i\|^2 \boldsymbol{\sigma}_i \right] \\ = -\frac{1}{2} \varepsilon^2 \mathbb{E}_n [(\sigma_{i,z} u_{i,y})^2 + (\sigma_{i,y} u_{i,z})^2 + (\sigma_{i,x} u_{i,z})^2 \\ + (\sigma_{i,z} u_{i,x})^2 + (\sigma_{i,y} u_{i,x})^2 + (\sigma_{i,x} u_{i,y})^2] \boldsymbol{\sigma}_i, \end{aligned}$$

where we have used that  $u_{i,x}$ ,  $u_{i,y}$ , and  $u_{i,z}$  are independent and mean zero. For the expectation of one  $(u_{i,q})^2$  we have that

$$\mathbb{E}_n [(u_{i,q})^2] = \sum_{j=1}^N \mathbb{E}_n [\lambda_j^2 \phi_{ji}^2 (w_{j,q})^2] = \sum_{j=1}^N \lambda_j^2 \phi_{ji}^2,$$

where we have used that the  $w_{j,x}$ ,  $w_{j,y}$ , and  $w_{j,z}$  are independent and mean zero for each  $j = 1, \dots, N$ . Together,

$$\mathbb{E}_n \left[ -\frac{1}{2} \varepsilon^2 \|\mathbf{v}_i\|^2 \boldsymbol{\sigma}_i \right] = -\varepsilon^2 \sum_{j=1}^N \lambda_j^2 \phi_{ji}^2 \boldsymbol{\sigma}_i, \quad (\text{A27})$$

where we have used the identity  $\sigma_{i,x}^2 + \sigma_{i,y}^2 + \sigma_{i,z}^2 = 1$ .

Now consider the cross-cross-product projection  $\mathbf{v}_i = -\boldsymbol{\sigma}_i \times (\boldsymbol{\sigma}_i \times \mathbf{u}_i)$ . Note that

$$\|\mathbf{v}_i\|^2 = \|-\boldsymbol{\sigma}_i \times (\boldsymbol{\sigma}_i \times \mathbf{u}_i)\|^2 = \|\boldsymbol{\sigma}_i \times \mathbf{u}_i\|^2$$

under the assumption that  $\|\boldsymbol{\sigma}_i\|^2 = 1$ . Therefore, for  $\mathbf{v}_i = -\boldsymbol{\sigma}_i \times (\boldsymbol{\sigma}_i \times \mathbf{u}_i)$  Eq. (A27) also holds.

Returning to (A6), we bound the error by

$$\begin{aligned} \mathbb{E} [\|(\mathbf{r}_4)_i\|^2] &\leq \left\{ \mathbb{E} \left[ \left( -\frac{1}{2} \varepsilon^2 \|\mathbf{v}_i\|^2 \boldsymbol{\sigma}_i \right)^4 \right] \right\}^{1/2} \{ \mathbb{E} [(-\beta \delta H)^4] \}^{1/2} \\ &\leq c \varepsilon^6 \{ \mathbb{E} [\|\mathbf{v}_i\|^8] \}^{1/2} [(\nabla H)^T P C_N P \nabla H]^2. \end{aligned}$$

Bounding the magnitude of this vector, we have that

$$\mathbb{E} [\|\tilde{\mathbf{r}}_4\|^2] \leq c \varepsilon^6 \text{Tr}(\tilde{C}_N) [(\nabla H)^T P C_N P \nabla H]^2 \quad (\text{A28})$$

regardless of the block projection matrix used,  $P$ .

Last, we will calculate the Itô correction for (22) and show it is equivalent to the right-hand side of (29). For the Stratonovich SDE of the form  $dX_t = \mu(t, X_t)dt + B(t, X_t) \circ dW_t$ , the corresponding Itô SDE is [37]

$$dX_t = \tilde{\mu}(t, X_t)dt + B(t, X_t)dW_t,$$

where

$$\tilde{\mu}_i(t, x) = \mu_i(t, x) + \frac{1}{2} \sum_j \sum_k \frac{\partial B_{ij}}{\partial x_k} B_{kj}.$$

Ignoring the constant coefficient  $\sqrt{2/\beta}$  for now, the Stratonovich SDE for a single spin taken from (22) can be written as

$$\begin{aligned} d\sigma_{i,x} &= \mu_{i,x}dt + \sigma_{i,y}dU_{i,z} - \sigma_{i,z}dU_{i,y}, \\ d\sigma_{i,z} &= \mu_{i,z}dt + \sigma_{i,x}dU_{i,y} - \sigma_{i,y}dU_{i,x}, \end{aligned} \quad (\text{A29})$$

where  $dU_{i,q} = \sum_{j=1}^N \lambda_j \phi_{ji} dW_{j,q}$  for  $q \in \{x, y, z\}$ . Consider first  $\sigma_{i,x}$

$$B_{ix,jx} = 0, \quad B_{ix,jy} = -\sigma_{i,z} \lambda_j \phi_{ji}, \quad B_{ix,jz} = \sigma_{i,y} \lambda_j \phi_{ji}.$$

Since  $B_{ix,jx} = 0$ , all the partial derivatives in the Itô correction are zero. For  $B_{ix,jy}$ , only  $\frac{\partial B_{ix,jy}}{\partial \sigma_{i,z}} = -\lambda_j \phi_{ji} \neq 0$  and the corresponding  $B_{iz,jy}$  are  $\sigma_{i,x} \lambda_j \phi_{ji}$ . Therefore, we have that

$$\sum_{j=1}^N \frac{\partial B_{ix,jy}}{\partial \sigma_{i,z}} B_{iz,jy} = -\sum_{j=1}^N \lambda_j^2 \phi_{ji}^2 \sigma_{i,x}.$$

Similarly, for  $B_{ix,jz}$ , only  $\frac{\partial B_{ix,jz}}{\partial \sigma_{i,y}} = \lambda_j \phi_{ji} \neq 0$  and the corresponding  $B_{iy,jz} = -\sigma_{i,x} \lambda_j \phi_{ji}$ . Therefore, we have that

$$\sum_{j=1}^N \frac{\partial B_{ix,jz}}{\partial \sigma_{i,y}} B_{iy,jz} = -\sum_{j=1}^N \lambda_j^2 \phi_{ji}^2 \sigma_{i,x}.$$

The above is for the cross-product. For the cross-cross-product,

$$B_{ix,jx} = (1 - \sigma_{ix}^2) \lambda_j \phi_{ji},$$

$$B_{ix,jy} = -\sigma_{ix} \sigma_{iy} \lambda_j \phi_{ji},$$

$$B_{ix,jz} = -\sigma_{ix} \sigma_{iz} \lambda_j \phi_{ji},$$

and  $B_{iy,jq}$ ,  $B_{iz,jq}$  follow similarly. For  $\sigma_{ix}$ ,

$$\frac{\partial B_{ix,jx}}{\partial \sigma_{ix}} B_{ix,jx} = -2\sigma_{ix} (1 - \sigma_{ix}^2) \lambda_j^2 \phi_{ji}^2,$$

$$\frac{\partial B_{ix,jy}}{\partial \sigma_{ix}} B_{ix,jy} = -\sigma_{iy} (-\sigma_{ix} \sigma_{iy}) \lambda_j^2 \phi_{ji}^2,$$

$$\frac{\partial B_{ix,jy}}{\partial \sigma_{iy}} B_{iy,jy} = -\sigma_{ix} (1 - \sigma_{iy}^2) \lambda_j^2 \phi_{ji}^2,$$

$$\frac{\partial B_{ix,jz}}{\partial \sigma_{ix}} B_{ix,jz} = -\sigma_{iz} (-\sigma_{ix} \sigma_{iz}) \lambda_j^2 \phi_{ji}^2,$$

$$\frac{\partial B_{ix,jz}}{\partial \sigma_{iz}} B_{iz,jz} = -\sigma_{ix} (1 - \sigma_{iz}^2) \lambda_j^2 \phi_{ji}^2.$$

Summing the above, we have for  $\sigma_{ix}$  the Itô correction is

$$\begin{aligned} &\frac{1}{2} \sum_j \lambda_j^2 \phi_{ji}^2 (-\sigma_{ix}) \\ &\quad \times [2(1 - \sigma_{ix}^2) + (1 - \sigma_{iy}^2) + (1 - \sigma_{iz}^2) - \sigma_{iy}^2 - \sigma_{iz}^2] \\ &= -\sum_j \lambda_j^2 \phi_{ji}^2 \sigma_{ix}. \end{aligned}$$

The above shows that the Itô form of the SDE has drift coefficient

$$\tilde{\mu}_{ix} = \mu_{ix} - \sum_{j=1}^N \lambda_j^2 \phi_{ji}^2 \sigma_{i,x}. \quad (\text{A30})$$

The calculations for  $\tilde{\mu}_{iy}$  and  $\tilde{\mu}_{iz}$  follow similarly.

With  $\sum_{j=1}^N \lambda_j^2 \phi_{ji}^2 = \frac{1}{N} \text{Tr}(\tilde{C}_N)$  for each  $i$ , and adding in the coefficient  $\sqrt{2/\beta}$ , the Itô drift for the equivalent equation to (22) is

$$\tilde{\mu} = \mu - \frac{2}{\beta} \frac{1}{N} \text{Tr}(\tilde{C}_N) \vec{\sigma}. \quad (\text{A31})$$

Recall the time change to arrive at the SDE,  $\delta t = \beta \varepsilon^2 / 2$ , with which we see that the above addition to the drift is equivalent to the calculated term in (A27).



## 2. The diffusion of one MH step

In this section, we bound the error of approximating the diffusion part of one step of the MH algorithm as in Eq. (31),

$$\zeta_i^n \equiv \sigma_i^{n+1} - \sigma_i^n - \mathbb{E}_n[\sigma_i^{n+1} - \sigma_i^n] - \varepsilon \mathbf{v}_i^n. \quad (\text{A32})$$

This random variable  $\zeta_i^n$  takes the values

$$\zeta_i^n = \begin{cases} f_i - \mathbb{E}_n[\sigma_i^{n+1} - \sigma_i^n] & \text{with prob. } \bar{\alpha} \\ -\mathbb{E}_n[\sigma_i^{n+1} - \sigma_i^n] - \varepsilon \mathbf{v}_i^n & \text{with prob. } 1 - \bar{\alpha} \end{cases},$$

where  $\bar{\alpha} = 1 \wedge e^{-\beta \delta H}$ ,  $f_i$  is defined in Eq. (A13) and  $\delta H$  is given in Eq. (25). This error can be bounded by

$$\begin{aligned} \mathbb{E}[\|\zeta_i^n\|^2] &= \mathbb{E}[\|f_i - \mathbb{E}_n[\sigma_i^{n+1} - \sigma_i^n]\|^2 (1 \wedge e^{-\beta \delta H})] \\ &\quad + \mathbb{E}[\|\mathbb{E}_n[\sigma_i^{n+1} - \sigma_i^n] - \varepsilon \mathbf{v}_i^n\|^2 [1 - (1 \wedge e^{-\beta \delta H})]] \\ &\leq \mathbb{E}[\|f_i\|^2] + \{\mathbb{E}[(\mathbb{E}_n[\sigma_i^{n+1} - \sigma_i^n] - \varepsilon \mathbf{v}_i^n)^4]\}^{1/2} \\ &\quad \{\mathbb{E}[(-\beta \delta H)^2]\}^{1/2} \leq c\varepsilon^3. \end{aligned} \quad (\text{A33})$$

The covariance of the error at different time steps  $n > m$  when is

$$\begin{aligned} \mathbb{E}[\zeta_{i,p}^n \zeta_{i,q}^m] &= \mathbb{E}[\mathbb{E}_n[\zeta_{i,p}^n \zeta_{i,q}^m]] = \mathbb{E}[\zeta_{i,p}^m \mathbb{E}_n[\zeta_{i,q}^n]] \\ &= \mathbb{E}[\zeta_{i,p}^m \cdot 0] = 0 \end{aligned} \quad (\text{A34})$$

for any  $i = 1, \dots, N$  and  $p, q \in \{x, y, z\}$ .

## 3. Completion of the proof

The Itô SDE (21) has a unique solution before proceeding in the next section to bound the error between the MH dynamics and this unique SDE solution. We apply Theorem 5.2.1 in Ref. [37] for an (Itô) SDE of the form  $dx = \mu(x)dt + B(x)dW$  by showing the SDE coefficients

$$\begin{aligned} \mu(x) &= \frac{1}{N} P_x C_N P_x^T \Delta_N x - 2\beta^{-1} \frac{1}{N} \text{Tr}(\bar{C}_N) x, \\ B(x) &= \sqrt{\frac{2}{\beta}} P_x C_N^{1/2} \end{aligned} \quad (\text{A35})$$

are Lipschitz continuous, which is a relatively straightforward calculation. There is an analogous argument in Sec. 3 of Ref. [9].

Following further the convergence results in Sec. 3 of Ref. [9], we can complete the proof of Theorem 1. The proof is similar to the proof of the stochastic Euler method. We will first prove a bound for the strong error

$$\tilde{e}(t) = \mathbb{E}[\|\tilde{s}(t) - \bar{\sigma}(t)\|^2] \quad (\text{A36})$$

at a fixed time  $t$ , where  $\bar{\sigma}(t)$  is the piecewise constant interpolation of the MH dynamics and  $\tilde{s}(t)$  is the solution to the SDE (21). Then  $\tilde{e}(t)$  and Doob's martingale inequality are used to obtain a uniform bound on

$$e(t) = \mathbb{E}\left[\sup_{0 \leq \tau \leq t} \|\tilde{s}(\tau) - \bar{\sigma}(\tau)\|^2\right]. \quad (\text{A37})$$

One must use the Itô isometry and Hölder's inequality to prove the following Grönwall inequality:

$$\tilde{e}(t) \leq (c_1 t + c_2) \int_0^t e(\tau) d\tau + c_3 \sqrt{\delta t}, \quad (\text{A38})$$

where  $c_1, c_2, c_3$  are functions of  $J, N, \beta, \text{Tr}(C_N)$ , and this gives the bound

$$\tilde{e}(t) \leq c_3 \sqrt{\delta t} e^{c_1 t + c_2}. \quad (\text{A39})$$

For a fixed  $t$ , take  $n = \lfloor \frac{t}{\delta t} \rfloor$ , then

$$\begin{aligned} \tilde{s}(t) - \bar{\sigma}(t) &= \int_0^{n\delta t} \{\mu[\tilde{s}(\tau)] - \mu[\bar{\sigma}(\tau)]\} d\tau \\ &\quad + \int_0^{n\delta t} \{B[\tilde{s}(\tau)] - B[\bar{\sigma}(\tau)]\} dW_\tau \\ &\quad + \int_{n\delta t}^t \mu[\tilde{s}(\tau)] d\tau \\ &\quad + \int_{n\delta t}^t B[\tilde{s}(\tau)] dW_\tau + \sum_{k=1}^n \tilde{r}^k + \sum_{k=1}^n \tilde{\zeta}^k, \end{aligned} \quad (\text{A40})$$

where the drift and diffusion coefficients,  $\mu(x)$  and  $B(x)$  are given in (A35) and the errors  $\tilde{r}$  and  $\tilde{\zeta}$  are bounded in (A7) and (A33). The remaining details are almost identical to those in [9], Section 3.4 and we refer the reader there for further details.

## APPENDIX B: INVARIANCE OF THE GIBBS DISTRIBUTION

In this Appendix, we present some direct calculations showing the invariance of the Gibbs distribution. In Sec. B 1, we present the well-known case of SDE (2) with additive noise. In Sec. B 2, we present the cases of the spin-system SDE (13) with white multiplicative noise using either the  $\sigma \times \cdot$  or the  $-\sigma \times (\sigma \times \cdot)$  projection, as well as the case of the SDE (22) with colored multiplicative noise, for which only the  $\sigma \times$  projection results in the invariance of the Gibbs distribution.

### 1. Additive noise

For the  $N$ -dimensional SDE (2) with constant matrix  $B$ , we show the invariance of the Gibbs distribution  $\rho(x) = e^{-\beta H(x)}$  by direct substitution into the Fokker-Planck equation (3). As  $\partial_j \rho = -\beta(\partial_j H)\rho$ , we have that

$$\begin{aligned} 0 &= \sum_i \sum_k \sum_j B_{ik} B_{jk} [(\partial_i \partial_j H)\rho - \beta(\partial_j H)(\partial_i H)\rho] \\ &\quad + \beta^{-1} \sum_i \sum_j \sum_k B_{ik} B_{jk} \partial_i [\partial_j (-\beta H)\rho] \\ &= \sum_i \sum_k \sum_j B_{ik} B_{jk} [(\partial_i \partial_j H)\rho - \beta(\partial_j H)(\partial_i H)\rho] \\ &\quad - \sum_i \sum_j \sum_k B_{ik} B_{jk} [(\partial_i \partial_j H)\rho - \beta(\partial_j H)(\partial_i H)\rho] \end{aligned}$$

and the terms on the right-hand side clearly cancel.

### 2. Multiplicative noise

Consider the following Stratonovich SDE with multiplicative noise

$$dX = -B(X)B^T(X)\nabla H(X) + \sqrt{2\beta^{-1}}B(X) \circ dW, \quad (\text{B1})$$

where  $B$  could be, for example, the block projection matrices  $P_1$  or  $P_2$ . It could also be the combination of  $PC^{1/2}$ . For the Fokker-Planck equation

$$\begin{aligned} \partial_t \rho(x, t) &= \sum_i \partial_i \{ (BB^T \nabla H)_i \rho(x, t) \} \\ &+ \beta^{-1} \sum_k \sum_i \partial_i \left\{ B_{ik} \sum_j \partial_j [B_{jk} \rho(x, t)] \right\} \\ &= \sum_i \partial_i \left\{ \sum_k \sum_j B_{ik} B_{jk} (\nabla H)_j \rho(x, t) \right\} \\ &+ \beta^{-1} \sum_k \sum_i \partial_i \left\{ B_{ik} \sum_j \partial_j [B_{jk} \rho(x, t)] \right\}, \quad (\text{B2}) \end{aligned}$$

$$\begin{aligned} \partial_i [B_{ik} (\partial_j B_{jk}) \rho] + \partial_i [B_{ik} B_{jk} \partial_j \rho] &= (\partial_i B_{ik}) (\partial_j B_{jk}) \rho + B_{ik} (\partial_i \partial_j B_{jk}) \rho + B_{ik} (\partial_j B_{jk}) (\partial_i \rho) + \partial_i (B_{ik} B_{jk}) (\partial_j \rho) + B_{ik} B_{jk} (\partial_i \partial_j \rho) \\ &= (\partial_i B_{ik}) (\partial_j B_{jk}) \rho + B_{ik} (\partial_i \partial_j B_{jk}) \rho - \beta B_{ik} (\partial_j B_{jk}) (\partial_i H) \rho \\ &\quad - \beta \partial_i (B_{ik} B_{jk}) (\partial_j H) \rho - \beta B_{ik} B_{jk} (\partial_i \partial_j H) \rho + \beta^2 B_{ik} B_{jk} (\partial_i H) (\partial_j H) \rho \end{aligned} \quad (\text{B4})$$

multiplied by  $\beta^{-1}$  and summed over  $i, j$  and  $k$ . Combining the terms in (B3) and (B4), the following terms, summed over  $i, j$ , and  $k$  are left over:

$$\beta^{-1} \{ (\partial_i B_{ik}) (\partial_j B_{jk}) \rho + B_{ik} (\partial_i \partial_j B_{jk}) \rho \} - B_{ik} (\partial_j B_{jk}) (\partial_i H) \rho. \quad (\text{B5})$$

For a generic multiplicative noise in the SDE of the form (B1), the Gibbs distribution is not guaranteed to be an invariant measure. Next, we consider specific cases for the matrix  $B$ .

We first consider the case of a single spin,  $X \in \mathbb{S}^2$ , and the matrix  $B$  as either the  $3 \times 3$  projection matrix  $P_1$  in Eq. (14) or  $P_2$  in Eq. (15). For the case of  $P_1$ , the terms in (B5) are all zero,

$$\begin{aligned} \sum_{i,j,k} (\partial_i B_{ik}) (\partial_j B_{jk}) &= 0, \\ \sum_{i,j,k} B_{ik} (\partial_i \partial_j B_{jk}) &= 0, \\ \sum_{i,j,k} B_{ik} (\partial_j B_{jk}) (\partial_i H) &= 0, \end{aligned}$$

and we conclude that the Gibbs distribution is invariant. For the case of  $P_2$ , the terms in (B5) are

$$\begin{aligned} \sum_{i,j,k} (\partial_i B_{ik}) (\partial_j B_{jk}) &= 4(\sigma_x^2 + \sigma_y^2 + \sigma_z^2) = 4, \\ \sum_{i,j,k} B_{ik} (\partial_i \partial_j B_{jk}) &= -4(\sigma_x^2 + \sigma_y^2 + \sigma_z^2) = -4, \\ \sum_{i,j,k} B_{ik} (\partial_j B_{jk}) (\partial_i H) &= 0. \end{aligned}$$

The order  $\beta^{-1}$  terms cancel while the order one term is zero, thus for the case of  $P_2$ , the Gibbs distribution is also invariant. This direct calculation easily extends to the case of  $N$  spins

we consider the invariance of the Gibbs distribution  $\rho(x) = e^{-\beta H(x)}$  by direct substitution. The third line (drift terms) of the above equation leads to the terms

$$\begin{aligned} \partial_i (B_{ik} B_{jk}) (\partial_j H) \rho + B_{ik} B_{jk} (\partial_i \partial_j H) \rho + B_{ik} B_{jk} (\partial_j H) (\partial_i \rho) \\ = \partial_i (B_{ik} B_{jk}) (\partial_j H) \rho + B_{ik} B_{jk} (\partial_i \partial_j H) \rho \\ - \beta B_{ik} B_{jk} (\partial_j H) (\partial_i H) \rho \end{aligned} \quad (\text{B3})$$

that are summed over  $i, j$ , and  $k$ . The fourth line (diffusion terms) of Eq. (B2) leads to the terms

with white noise, and we conclude that with either projection in the SDE (13), the Gibbs distribution (1) is invariant.

We now consider the case of the colored noise SDE (22). Taking  $B = P_1 C_N^{1/2}$  and indexing the vector  $X$  as  $X_{ix} = x_i$ ,  $X_{iy} = y_i$  and  $X_{iz} = z_i$ , for  $i = 1, \dots, N$  with the  $i$ th spin vector being denoted as  $\langle x_i, y_i, z_i \rangle$ , we have that

$$\begin{aligned} B_{ix,jx} &= 0, & B_{ix,jy} &= -z_i \lambda_j \phi_{ji}, & B_{ix,jz} &= y_i \lambda_j \phi_{ji}, \\ B_{iy,jx} &= -z_i \lambda_j \phi_{ji}, & B_{iy,jy} &= 0, & B_{iy,jz} &= x_i \lambda_j \phi_{ji}, \\ B_{iz,jx} &= y_i \lambda_j \phi_{ji}, & B_{iz,jy} &= -x_i \lambda_j \phi_{ji}, & B_{iz,jz} &= 0, \end{aligned}$$

where  $\sum_{k=1}^N \bar{C}_{ik} \phi_{kj} = \lambda_j^2 \phi_{ij}$  for  $i, j = 1, \dots, N$ . By inspection we see that

$$\partial_{iq} B_{iq,jp} = 0$$

for all  $p, q \in x, y, z$ , and therefore all the terms in (B5) are zero. For the case of colored noise and the  $\sigma \times$  projection matrix  $P_1$ , the Gibbs distribution is an invariant measure for the SDE (22).

Taking  $B = P_2 C_N^{1/2}$  with  $P_2$  given by (15),

$$\begin{aligned} B_{ix,jx} &= (1 - x_i^2) \lambda_j \phi_{ji}, & B_{ix,jy} &= -x_i y_i \lambda_j \phi_{ji}, \\ B_{ix,jz} &= -x_i z_i \lambda_j \phi_{ji}, & B_{iy,jx} &= -x_i y_i \lambda_j \phi_{ji}, \\ B_{iy,jy} &= (1 - y_i^2) \lambda_j \phi_{ji}, & B_{iy,jz} &= -y_i z_i \lambda_j \phi_{ji}, \\ B_{iz,jx} &= -x_i z_i \lambda_j \phi_{ji}, & B_{iz,jy} &= -y_i z_i \lambda_j \phi_{ji}, \\ B_{iz,jz} &= (1 - z_i^2) \lambda_j \phi_{ji} \end{aligned}$$

and noting that

$$\begin{aligned} \partial_{iq} B_{iq,jq} &= -2q_i \lambda_j \phi_{ji}, \\ \partial_{iq} B_{iq,jp} &= -p_i \lambda_j \phi_{ji}, \\ \partial_{jq} H &= -N(q_{j+1} - 2q_j + q_{j-1}), \end{aligned}$$

further algebra leads to the conclusion that the third term in (B5),

$$\sum_{i,j,k=1}^N \sum_{p,q,r \in \{x,y,z\}} B_{ip,kr} (\partial_{jq} B_{jq,kr}) \partial_{ip} H \neq 0.$$

Since this term can never cancel with the order  $\beta^{-1}$  terms for arbitrary  $\beta$ , we conclude that the Gibbs distribution is not an invariant measure for the colored noise SDE (22) with the  $-\sigma \times (\sigma \times \cdot)$  projection matrix  $P_2$ .

### APPENDIX C: WELL-POSEDNESS OF THE COLORED NOISE SPDE MODEL FOR TRACE CLASS MULTIPLIERS IN SOBOLEV SPACES

To establish the local well-posedness of (45) result in the case that  $M_\kappa(D)$  is trace class ( $\kappa$  sufficiently large), we may write the SPDE in the Itô formulation

$$d\sigma = \left[ -\sigma \times [M_\kappa(D)](\sigma \times \Delta\sigma) + \int K_\kappa(x, y) dy \sigma(x) \right] dt + \sigma \times [\mathcal{F}^{-1}m(|k|)]^{-\kappa} dW(k). \quad (C1)$$

Note that the lack of any geometric projection in the Itô correction term follows from a cancellation that arises from direct computation very similar to that in Sec. A 1 c. To prove local well-posedness of (C1), we establish some baseline energy estimates by first mollifying the equation, then proving a priori bounds in a high enough Sobolev space.

To proceed, let us record a few useful facts. First,  $H^s(\mathbb{T}^d)$  is an algebra for  $s > d/2$ , namely,  $\|uv\|_{H^s} \leq \|u\|_{H^s} \|v\|_{H^s}$  for  $s$  sufficiently large. Second,  $[\partial_x, (I - \Delta)^{\kappa/2}] = 0$ . Last, we observe that

$$\begin{aligned} & \int \Delta\sigma \cdot \{-\sigma \times [M_\alpha(D)](\sigma \times \Delta\sigma)\} dx \\ &= \int (\sigma \times \Delta\sigma) \cdot [M_\alpha(D)](\sigma \times \Delta\sigma) dx > 0, \end{aligned} \quad (C2)$$

which allows us to generate a priori bounds on  $\|\sigma\|_{H^s}$  uniformly bounded in  $\epsilon$  for  $s > \frac{d}{2}$ . Note, this also shows that a classical solution will have decaying  $H^1$  norm, as would be expected from the structure of the Gibbs measure.

*Remark 3.* We can easily see that were we using the cross-cross-projection on the drift term, then we could write the deterministic flow generated by the drift as

$$\partial_t \sigma = M_\kappa(D) \Delta \sigma + M_\kappa(D) (|\nabla \sigma|^2 \sigma)$$

suggesting that the dynamics of this nonlocal PDE should have a different diffusion time scale roughly given by  $M_\alpha(\lambda_1) \lambda_1^2$ . Numerically we observe a similar diffusion scaling for the cross-projection, which is explored in Sec. V.

The outline of the local well-posedness argument proof as in Ref. [41], Chap. 15, proceeds as follows:

(1) Define  $\sigma_\epsilon = \chi(D/\epsilon)\sigma$ , a frequency cutoff version of the equation.

(2) By the (C2) adapted to this setting, this ODE system has global existence for each  $\epsilon$

(3) For  $s > \frac{d}{2} + 2$ , computing  $\partial_t \mathbb{E}[\|\sigma_\epsilon\|_{H^s}^2]$  in a similar fashion to (C2) gives a signed quantity on the highest derivatives and using the algebra property gives uniform a uniform existence time bound using a simple Gronwall inequality argument in  $\epsilon$  on  $\mathbb{E}[\|\sigma_\epsilon\|_{H^s}]$  provided  $s$  is sufficiently large provided the initial data are sufficiently regular. We use the Itô formula

$$\begin{aligned} \mathbb{E}[\|\sigma_\epsilon\|_{H^s}^2] &= \mathbb{E}[\|\sigma_\epsilon\|_{H^s}^2(0)] \\ &+ \mathbb{E}\left\{ \int_0^t \langle \sigma_\epsilon(r), \sigma_\epsilon \times [M_\kappa(D)](\sigma_\epsilon \times \Delta\sigma_\epsilon)(r) \rangle_{H^s} dr \right\} \\ &+ \mathbb{E}\left\{ \int_0^t \langle \sigma_\epsilon(r), \int K_\kappa(x, y) dy \sigma_\epsilon(x, r) \rangle_{H^s} dr \right\} \\ &+ \mathbb{E}\left\{ \int_0^t \|\sigma_\epsilon \times (\mathcal{F}^{-1}[m(|k|)]^{-\kappa})\|_{L_2(L^2, \dot{H}^s)}^2 dr \right\}, \end{aligned}$$

where  $L_2(H_1, H_2)$  is the space of linear Hilbert-Schmidt operators from Hilbert space  $H_1$  to Hilbert space  $H_2$ .

(4) Since balls in the  $H^s$  norm are compact, the process is tight, and hence taking the weak limit in  $\epsilon$  gives the solution  $\sigma \in H^s$ .

(5) This can be bootstrapped into supremum over time.

(6) Uniqueness and continuity follow from energy estimates once the classical solution has been constructed.

(7) Note, in all of this, there is an Itô correction term to be controlled in each energy estimate. However, as it carries no derivatives, it is lower order and can be absorbed in the energy estimate without much difficulty.

*Remark 4.* This rather crude treatment of (45) is by no means the state of the art. However, as we are more concerned here with the microscopic to macroscopic convergence to such an equation, it suffices for our purposes. The recent work of Ref. [5] handles the white noise case of (45) and gives an appropriate renormalization technique adapted to a general geometric setting. It would be interesting to extend this to the range of covariance matrices given here between identity (white) and  $\kappa > \frac{1}{2}$  (trace class), but this pursuit goes beyond the focus of the present work.

[1] R. V. Kohn, M. G. Reznikoff, and E. Vanden-Eijnden, *J. Nonlinear Sci.* **15**, 223 (2005).  
[2] A. de Bouard and A. Debussche, *Commun. Math. Phys.* **205**, 161 (1999).  
[3] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions* (Cambridge University Press, Cambridge, 2014).  
[4] M. Hairer, *Invent. Math.* **198**, 269 (2014).  
[5] Y. Bruned, F. Gabriel, M. Hairer, and L. Zambotti, *arXiv:1902.02884* (2019).  
[6] W. K. Hastings, *Biometrika* **57**, 97 (1970).

[7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).  
[8] A. W. C. Lau and T. C. Lubensky, *Phys. Rev. E* **76**, 011123 (2007).  
[9] Y. Gao, K. Kirkpatrick, J. Marzuola, J. Mattingly, and K. Newhall, *arXiv:1806.05282* [Comm. Math. Sci. (to be published)].  
[10] M. D. Ryser, N. Nigam, and P. F. Tupper, *J. Comput. Phys.* **231**, 2537 (2012).  
[11] M. Struwe *et al.*, *J. Diff. Geom.* **28**, 485 (1988).

- [12] B. Guo and M.-C. Hong, *Calculus Var. Partial Differ. Equations* **1**, 311 (1993).
- [13] W. Wernsdorfer, E. B. Orozco, K. Hasselbach, A. Benoit, B. Barbara, N. Demoncy, A. Loiseau, H. Pascard, and D. Mailly, *Phys. Rev. Lett.* **78**, 1791 (1997).
- [14] W. T. Coffey and Y. P. Kalmykov, *J. Appl. Phys.* **112**, 121301 (2012).
- [15] L. Banas, Z. Brzezniak, M. Neklyudov, and A. Prohl, *Stochastic Ferromagnetism: Analysis and Numerics*, De Gruyter Studies in Mathematics (Walter de Gruyter, Berlin/Boston, 2014), Vol. 58.
- [16] O. Chugreeva and C. Melcher, *IMA J. Appl. Math.* **83**, 261 (2018).
- [17] M. Nishino and S. Miyashita, *Phys. Rev. B* **91**, 134411 (2015).
- [18] U. Atxitia, O. Chubykalo-Fesenko, R. W. Chantrell, U. Nowak, and A. Rebei, *Phys. Rev. Lett.* **102**, 057203 (2009).
- [19] R. F. L. Evans, D. Hinzke, U. Atxitia, U. Nowak, R. W. Chantrell, and O. Chubykalo-Fesenko, *Phys. Rev. B* **85**, 014433 (2012).
- [20] J. C. Mattingly *et al.*, *Ann. Appl. Probab.* **22**, 881 (2012).
- [21] J. Kuntz, M. Ottobre, and A. M. Stuart, *Ann. Inst. H. Poincaré Probab. Stat.* **55**, 1599 (2019).
- [22] D. Hochberg, C. Molina-París, J. Pérez-Mercader, and M. Visser, *Phys. Rev. E* **60**, 6343 (1999).
- [23] M. Newman and G. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, New York, 1999), Chaps. 1–4.
- [24] K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt, and S. McKay, *Comput. Phys.* **7**, 156 (1993).
- [25] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2014).
- [26] G. G. Batrouni, *Phys. Rev. B* **70**, 184517 (2004).
- [27] I. Maccari, A. Maiorano, E. Marinari, and J. J. Ruiz-Lorenzo, *Eur. Phys. J. B* **89**, 127 (2016).
- [28] A. M. Stuart, *Acta Numerica* **19**, 451 (2010).
- [29] M. Hairer, A. M. Stuart, and J. Voss, *Signal Processing Problems on Function Space: Bayesian Formulation, Stochastic PDEs and Effective MCMC Methods*, The Oxford Handbook of Nonlinear Filtering (Oxford University Press, Oxford, UK, 2011), pp. 833–873.
- [30] G. O. Roberts, A. Gelman, and W. R. Gilks, *Ann. Appl. Probab.* **7**, 110 (1997).
- [31] L. A. Breyer and G. O. Roberts, *Stoch. Proc. Appl.* **90**, 181 (2000).
- [32] B. Jourdain *et al.*, *Bernoulli* **20**, 1930 (2014).
- [33] B. Jourdain *et al.*, *Ann. Appl. Probab.* **25**, 2263 (2015).
- [34] M. Hairer *et al.*, *Commun. Math. Sci.* **3**, 587 (2005).
- [35] M. Hairer *et al.*, *Ann. Appl. Probab.* **17**, 1657 (2007).
- [36] J. C. Mattingly, A. M. Stuart, and M. V. Tretyakov, *SIAM J. Numer. Anal.* **48**, 552 (2010).
- [37] B. Øksendal, *Stochastic Differential Equations* (Springer, Berlin, Heidelberg, 2003).
- [38] L. Hörmander, *Acta Math.* **119**, 147 (1967).
- [39] L. Hörmander, *The Analysis of Linear Partial Differential Operators* (Springer, Berlin, Heidelberg, 2015), Vol. 4.
- [40] B. Gess and J. M. Tölle, *J. Diff. Eq.* **260**, 4973 (2016).
- [41] M. E. Taylor, *Partial Differential Equations III. Nonlinear Equations*, 2nd ed., Applied Mathematical Sciences (Springer, New York, 2011), Vol. 117.