MULTILEVEL PLAQUETTE-SPACE SAMPLING FOR LATTICE GAUGE THEORIES WITH LOCAL CONSTRAINT SOLVES

Anonymous authors
Paper under double-blind review

ABSTRACT

Lattice gauge theories are an important class of physical models with high-dimensional structured distributions that underpin first-principles calculations in particle, nuclear, and condensed-matter physics. Recent advances in generative modeling have opened new avenues for sampling Boltzmann distributions in lattice gauge theories, offering the potential to alleviate limitations of traditional Monte Carlo methods, including critical slowing down and topological freezing. However, generative samplers for lattice gauge theories are typically constructed in link space, where correlations become increasingly long-ranged toward weak coupling, posing a challenge for learning. Plaquettes provide a more natural representation, as the action is local in these variables. However, exact Bianchi constraints restrict them to a lower-dimensional manifold, complicating direct generative modeling. We introduce a multilevel normalizing-flow construction that samples directly in plaquette space while satisfying these constraints exactly. The key idea is a coarse-to-fine factorization that transforms a globally coupled constraint problem into a sequence of local solves: at each refinement, every determined plaquette depends on at most four newly generated variables, while the size of the constraint problem remains independent of the lattice size. We validate the construction for $U(1)$ in two and four dimensions and for $SU(2)$ in two dimensions. Our multilevel Plaquette-Space Sampler (PSS) substantially outperforms link-space baselines, with the advantage increasing toward weak coupling, where the gauge coupling becomes small. This regime is particularly challenging for generative sampling and, in asymptotically free gauge theories, is relevant to continuum studies.

1 INTRODUCTION

Lattice gauge theory provides a first-principles framework for studying strongly coupled quantum gauge theories, where perturbative methods are no longer applicable (Wilson, 1974; Gattringer & Lang, 2010). Numerical lattice simulations have become an essential tool for quantitative predictions in particle and nuclear physics, ranging from hadron spectroscopy and matrix elements to finite-temperature phenomena and precision tests of the Standard Model. These calculations reduce physical observables to expectation values over a high-dimensional probability distribution, making efficient sampling of gauge-field configurations a central computational problem.

In practice, these expectations are estimated predominantly with Markov chain Monte Carlo (MCMC) methods. Their computational cost, however, grows rapidly as the lattice spacing is reduced toward the continuum limit and correlations extend over increasingly many lattice sites (Wolff, 1990; Schaefer et al., 2011); generative models have been applied directly to this critical slowing down (Albergo et al., 2019; Wu et al., 2019; Kanwar et al., 2020). The problem is particularly severe for topological observables, whose autocorrelation times can grow dramatically and eventually lead to effective freezing in a fixed topological sector (Del Debbio et al., 2004; Lüscher & Schaefer, 2011). Improving the scalability of lattice sampling is therefore important not only computationally, but also for extending reliable first-principles predictions into increasingly demanding physical regimes.

Generative models provide a promising alternative to conventional Markov-chain updates. A model with tractable density q can generate global, approximately independent proposals, while the mismatch

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

with the target distribution $p \propto e^{-S}$ can be corrected exactly through importance reweighting or an independence Metropolis step (Albergo et al., 2019; Nicoli et al., 2020). Recent progress has shown that generative samplers can be particularly effective at mitigating topological freezing in low-dimensional lattice gauge theories (Kanwar et al., 2020; Rothkopf, 2026; Singha et al., 2026b). These results highlight their potential, but also shift the central challenge toward *scalability*: applying the same ideas to larger lattices and higher-dimensional gauge theories requires generative models whose quality does not deteriorate rapidly with lattice size, correlation length, or weak coupling.

Most existing generative approaches sample configurations in *link space*, including gauge-equivariant flows (Kanwar et al., 2020; Boyda et al., 2021; Abbott et al., 2023b), continuous and stochastic flows (Bacchio et al., 2023; Caselle et al., 2022a; Bonanno et al., 2026b; Gerdes et al., 2023), and diffusion models (Wang et al., 2024; Zhu et al., 2025). Recent stochastic-flow approaches aim to mitigate topological freezing in four-dimensional gauge theories by learning transformations with long non-equilibrium Monte Carlo trajectories, so that much of the computational cost still comes from repeated MCMC updates (Bonanno et al., 2026b). However, the Wilson gauge action, which provides the standard discretization of the gauge-field action, depends on the link variables only through the plaquettes—the ordered products of links around elementary squares.[1] Plaquettes are therefore a natural representation of the physical degrees of freedom: the action is local in them and gauge redundancy is removed up to one global transformation. In link space, by contrast, every plaquette is a derived quantity involving several links, while each link participates in several plaquettes. As the coupling weakens, the plaquettes become increasingly concentrated near the identity and correlations extend over longer distances, requiring increasingly precise coordination among the link variables. Consistent with this picture, link-space flows have been observed to degrade toward weak coupling (Abbott et al., 2024a).

This motivates generating plaquettes directly. Indeed, a recent study reported successful results using plaquette-space sampling for two-dimensional $U(1)$ theory (Singha et al., 2026b). The present work aims to generalize this approach to higher-dimensional theories and larger gauge groups. The difficulty is that a general plaquette configuration does not necessarily correspond to a valid link configuration: a valid plaquette configuration must satisfy geometric consistency constraints, known as the Bianchi identities (Batrouni, 1982), which restrict physically admissible plaquette fields to a lower-dimensional constraint manifold. While these constraints are simple in two dimensions, generating plaquette configurations that satisfy all Bianchi identities is non-trivial. Furthermore, in higher-dimensional theories, the size of the linear system grows rapidly with the lattice size, making it increasingly challenging to solve.

We tackle this challenge using multilevel generative sampling. At the coarsest level, a small number of plaquette variables, together with additional degrees of freedom called holonomies, are generated such that a subset of the Bianchi identities is satisfied. At each subsequent finer level, each coarse plaquette is decomposed into finer plaquettes while preserving its original value as the value of the corresponding Wilson loop at the finer level. This Wilson loop encloses multiple fine plaquettes, and its value is given by the ordered product of the plaquette variables contained within it. Consequently, all Bianchi identities satisfied at coarser levels are automatically inherited by the finer levels, and hence remain satisfied at the finest level. This transforms the globally coupled constraint problem into a sequence of local constraint problems: each determined plaquette depends on at most four newly generated variables, and the size of each local problem is independent of the lattice size. Furthermore, multilevel sampling naturally captures long-range correlations through the coarse-level samplers and has been shown to be advantageous (Singha et al., 2025; Bauer et al., 2025; Singha et al., 2026a; Hasenfratz et al., 2026b;a) for studying physical systems, particularly near criticality and towards the continuum limit—regimes of particular physical relevance.

We instantiate the method for $U(1)$ gauge theory in two and four dimensions and for $SU(2)$ in two dimensions. Numerical experiments show that our multilevel Plaquette-Space Sampler (PSS) substantially outperforms both single-level and multilevel link-space baselines. In two-dimensional compact $U(1)$, PSS scales more favorably with lattice size than the corresponding baselines, while in four-dimensional $U(1)$ and two-dimensional $SU(2)$, it becomes increasingly advantageous toward weak coupling.

[1] More general improved gauge actions may additionally depend on larger Wilson loops such as rectangular loops, to reduce discretization errors while retaining locality.

The main contributions of this paper are:

- We generalize plaquette-space sampling to four dimensions and to $SU(2)$ gauge theory.
- We propose a multilevel sampler for gauge theories that uses simple local procedures to satisfy the Bianchi identities while better capturing long-range correlations.
- We empirically demonstrate the benefits of both plaquette-space sampling and the multilevel scheme in physically relevant regimes where long-range correlations are pronounced.

2 Background

2.1 Lattice gauge theory

The lattice and its variables. We consider a periodic hypercubic lattice in d dimensions with V sites, each denoted by x, and unit vectors $\hat{\mu}$, $\mu = 0, \dots, d-1$. Gauge fields live on the lattice edges: the oriented link from x to $x + \hat{\mu}$ carries an element $U_\mu(x)$ of a compact group G, while the opposite orientation, i.e., from $x + \hat{\mu}$ to x, is $U_\mu(x)^\dagger$ (Fig. 1a). The choice of G specifies the gauge theory. When G is $U(1)$, $U = e^{i\theta}$ with $\theta \in (-\pi, \pi]$; and when G is $SU(2)$, $U = a_0 + i\,\boldsymbol{a} \cdot \boldsymbol{\sigma}$, with $(a_0, \boldsymbol{a})$ a unit vector in $\mathbb{R}^4$ and $\boldsymbol{\sigma}$ are 2×2 Pauli matrices. A lattice configuration is thus a collection of link variables $U_\mu(x)$, i.e. a point in G^{dV}.

Plaquettes and distribution. The elementary gauge-covariant object is the plaquette, the ordered product of links around a unit square in the $\mu\nu$ plane (Fig. 1a),

$$P_{\mu\nu}(x) = U_\mu(x)\, U_\nu(x + \hat{\mu})\, U_\mu(x + \hat{\nu})^\dagger\, U_\nu(x)^\dagger. \tag{1}$$

The Wilson action is a sum of single-plaquette terms, and the target distribution is

$$S(U) = -\beta \textstyle\sum_{x,\,\mu<\nu} \frac{1}{N} \operatorname{Re}\operatorname{tr} P_{\mu\nu}(x), \qquad p(U)\,\mathrm{d}U = \frac{1}{Z}\, e^{-S(U)} \prod_{x,\mu} \mathrm{d}U_\mu(x), \tag{2}$$

where N is the dimension of the matrices, $\mathrm{d}U$ the normalised Haar measure of G, and Z the partition function. The coupling β is the only parameter of the model.

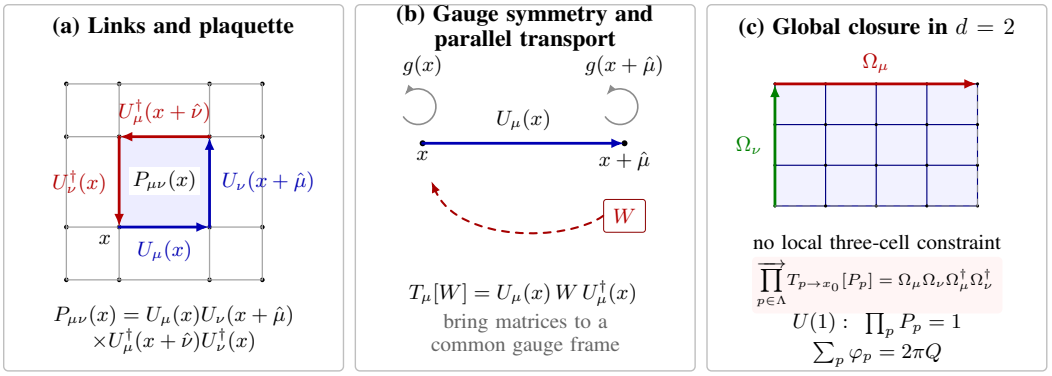


Figure 1: Links, gauge symmetry and the plaquette constraints. **(a)** Oriented links and the plaquette $P_{\mu\nu}(x)$ they form. **(b)** Independent gauge transformations at neighbouring sites, and the parallel transport of W from $x + \hat{\mu}$ to the frame at x. **(c)** In $d = 2$ the only constraint is the global closure through the holonomies Ω_μ and Ω_ν.

Gauge symmetry and gauge fixing. The link variables carry a local redundancy. A gauge transformation assigns an element $g(x) \in G$ to every site and acts as

$$U_\mu(x) \mapsto g(x)\, U_\mu(x)\, g(x + \hat{\mu})^\dagger. \tag{3}$$

A plaquette is a closed loop based at x, so along it the factors g cancel in neighbouring pairs and only the endpoint survives: $P_{\mu\nu}(x) \mapsto g(x)\, P_{\mu\nu}(x)\, g(x)^\dagger$. Its trace is therefore unchanged and so

is the action (2), while the Haar measure is invariant under multiplication by a fixed group element. Consequently, the whole distribution (2) is gauge invariant. The redundancy can be removed by fixing the $V-1$ links of a maximal spanning tree to the identity, leaving one global transformation at the root (Appendix B.1, Fig. 6). Gauge-invariant observables can be computed directly from the gauge-fixed samples. For gauge-dependent observables, the gauge degrees of freedom can be sampled independently from the uniform distribution.

Gauge covariance also dictates how objects attached to different sites may be combined. A matrix W attached to the site $x+\hat{\mu}$ carries the gauge freedom of that site, transforming as $W \mapsto g(x+\hat{\mu})\, W\, g(x+\hat{\mu})^\dagger$, so it cannot simply be multiplied with a matrix attached to x. Conjugating it with the link that joins the two sites, $T_\mu[W] = U_\mu(x)\, W\, U_\mu(x)^\dagger$, gives an object that transforms at x instead, because the factors $g(x+\hat{\mu})$ carried by the link (3) cancel against those of W. This is called parallel transport (Fig. 1b): the link is what carries an object to its neighbouring site. For $U(1)$ it has no effect, the group being Abelian; for $SU(2)$ it is what makes a product of plaquettes based at different sites well defined, as in the identity below (See Appendix C for more details).

The Bianchi identity. Any configuration of link variables produces plaquettes that satisfy a fixed set of relations exactly and automatically, because each link occurs in them once with each orientation. Accordingly, a sampler that generates plaquettes directly must satisfy all such constraints, collectively expressed by the lattice Bianchi identities. For any contractible closed surface Σ, the Bianchi identities imply

$$\overrightarrow{\prod}_{p\in\Sigma} T_{p\to x_0}\left[P_p^{\epsilon_{\Sigma p}}\right] = \mathbf{1}, \tag{4}$$

where the arrow indicates that the plaquettes are multiplied in the prescribed surface order, $T_{p\to x_0}[\cdot]$ transports each plaquette to a common base point x_0, and $\epsilon_{\Sigma p} = \pm 1$ specifies its orientation relative to Σ. For any closed non-contractible surface Λ arising from the toroidal topology induced by periodic boundary conditions, the Bianchi identities instead imply

$$\overrightarrow{\prod}_{p\in\Lambda} T_{p\to x_0}[P_p] = \Omega_\mu\, \Omega_\nu\, \Omega_\mu^\dagger\, \Omega_\nu^\dagger, \tag{5}$$

where Ω_μ and Ω_ν are the holonomies wrapping the μ- and ν-directions, which span Λ, respectively (Fig. 1c). See Appendix D for derivation, and Appendix D.2 for the explicit forms of the Bianchi identities for $U(1)$ and $SU(2)$ theories.

2.2 Related Work: Sampling with generative models

Expectation values under (2) are conventionally estimated using MCMC methods. As correlations grow, however, local updates become increasingly correlated and the cost of obtaining effectively independent configurations rises (Duane et al., 1987; Wolff, 1990; Schaefer et al., 2011). Generative models provide a different route: a model with tractable density q can propose independent configurations, while deviations from the target distribution are corrected statistically (Albergo et al., 2019; Nicoli et al., 2021).

Generative sampling for lattice field theory. Normalizing-flow samplers were first introduced for scalar lattice field theories (Albergo et al., 2019) and subsequently extended to gauge theories through exactly gauge-equivariant coupling layers, first for $U(1)$ (Kanwar et al., 2020), then for $SU(N)$ (Boyda et al., 2021), and later to arbitrary space-time dimension (Abbott et al., 2023b); see Albergo et al. (2021a) and Cranmer et al. (2023) for reviews. The framework has since been applied to fermionic theories (Albergo et al., 2021b; 2022), the Hubbard model (Schuh et al., 2026; Kreit et al., 2026), entanglement observables (Bulgarelli et al., 2025a), and correlated ensemble generation for QCD (Abbott et al., 2024b; 2026). Related approaches include stochastic normalizing flows based on Jarzynski's equality (Caselle et al., 2016; 2022a; Bulgarelli et al., 2025b), learned continuous maps (Bacchio et al., 2023; Gerdes et al., 2025), autoregressive models (Wu et al., 2019), and diffusion models for gauge theories (Wang et al., 2024; Zhu et al., 2025; Vega et al., 2026; Aarts et al., 2026; Kanwar & Vega, 2025). Very recently, Singha et al. (2026b) applied plaquette-space sampling to two dimensional $U(1)$ theory, and proposed a mixture-model approach to mitigate topological freezing. All these previous works, except Singha et al. (2026b), generate link variables directly. Our work generalizes the plaquette-space sampling to higher dimensions and larger gauge groups.

Hierarchical and constrained generative models. Coarse-to-fine sampling has a long history, including multigrid Monte Carlo (Goodman & Sokal, 1986; 1989), renormalization-group-guided updates (Schmidt, 1983; Faas & Hilhorst, 1986), multiscale equilibration (Endres et al., 2015), and multilevel variance-reduction methods (Cè et al., 2016; Giles, 2015). The learned hierarchical constructions have directly connected generative modeling with the renormalization group (Li & Wang, 2018; Koch-Janusz & Ringel, 2018; Hu et al., 2022). For discrete spin systems, hierarchical generative approaches include HAN (Białas et al., 2022) and RiGCS (Singha et al., 2025), the latter constructing configurations from coarse to fine by conditioning each refinement level on the preceding one. For continuous lattice field theories, closely related coarse-to-fine generative sampling has been developed through super-resolving normalizing flows (Bauer et al., 2025) and multilevel conditional-flow constructions (Singha et al., 2026a). In lattice gauge theory, Abbott et al. (2024a) proposed a multiscale normalizing-flow construction that progressively generates gauge fields from coarse to fine in link space, which we use as a baseline in our experiments. For an extended discussion, see Appendix A.

3 Proposed Method: Plaquette-Space Sampler (PSS)

Our approach is to construct a generative sampler that generates valid plaquette configurations satisfying all Bianchi constraints exactly, thereby ensuring the existence of corresponding link configurations. Although a single-level plaquette sampler can in principle be constructed by solving all Bianchi constraints simultaneously (see Appendix E), it does not scale to large lattice sizes in higher-dimensional theories, as the the resulting linear system grows rapidly, potentially making the computation challenging. Accordingly, we employ a multilevel sampling scheme.

In a d-dimensional lattice with V sites, the numbers of plaquette variables and independent Bianchi constraints are given by $M = \frac{d(d-1)}{2}V$ and $B = \frac{(d-1)(d-2)}{2}V + d - 1$, respectively (see Appendix B). Let $\mathcal{V}^{(i)} (\subset G^{M^{(i)}})$ be the set of valid plaquette fields. At the coarsest level $i = 0$, we generate $N^{(0)} = M^{(0)} - B^{(0)}$ free plaquette variables $F^{(0)} \sim \widetilde{q}_0(\cdot)$ and determine the remaining $B^{(0)}$ plaquettes from the Bianchi identities (Fig. 2, left). We also generate d holonomies $\Omega \sim q_\Omega$ without imposing any constraints. We denote this coarsest map by $R_0 : G^{N^{(0)}} \times G^d \mapsto \mathcal{V}^{(0)}$, which plays the role of R_i below, with the holonomies in place of the coarse field, so that $P^{(0)} = R_0(F^{(0)}; \Omega)$.

At each subsequent level, we generate the finer plaquettes $P^{(i)} \in \mathcal{V}^{(i)}$ conditional on the coarse plaquettes $P^{(i-1)}$. Here, unlike the approach of Bauer et al. (2025) and Abbott et al. (2024a), we preserve the coarse plaquette variables by imposing local blocking constraints that fix the corresponding Wilson loops at the finer level to the same values (Fig. 2, centre and right). This makes the number of free plaquette variables generated at level i be $N^{(i)} = \big(M^{(i)} - B^{(i)}\big) - \big(M^{(i-1)} - B^{(i-1)}\big)$. Let $R_i : G^{N^{(i)}} \times \mathcal{V}^{(i-1)} \mapsto \mathcal{V}^{(i)}$ be the fixed, bijective, and volume-preserving *refinement* map from the free plaquette variables to the valid plaquette configurations. Then, the sampling process is given as

$$F^{(i)} \sim \widetilde{q}_i(\,\cdot \,|\, P^{(i-1)}), \qquad P^{(i)} = R_i(F^{(i)}; P^{(i-1)}). \tag{6}$$

Construction of the refinement map R_i depends strongly on the lattice dimension and the gauge group, whose explicit forms are detailed in Fig. 2.

Let $q_i(\,\cdot \,|\, P^{(i-1)}) = [R_i(\cdot; P^{(i-1)})]_{\#}\widetilde{q}_i\left(\cdot \middle| P^{(i-1)}\right)$ be the push-forward of $\widetilde{q}_i$ under R_i. Then, our complete multilevel sampler is given as

$$q(P) = q_0(P^{(0)}) \prod_{i=1}^{n} q_i(P^{(i)} | P^{(i-1)}) \tag{7}$$

with all densities with respect to the intrinsic measure on the Bianchi-constrained plaquette space. We model all conditionals $\widetilde{q}_i$ by normalizing flows with the exact sampling density given by $q(P) = q_0(P^{(0)}) \prod_{i=1}^{n} q_i(R_i^{-1}(P^{(i)}; P^{(i-1)}) | P^{(i-1)})$. We train the multilevel Plaquette-Space Sampler (PSS) by minimizing the reverse Kullback–Leibler divergence

$$\mathcal{L}(\theta) = \mathbb{E}_{P \sim q_\theta}\big[\log q_\theta(P) + S(P)\big] = \mathrm{KL}(q_\theta \| p) - \log Z. \tag{8}$$

Architectures and training schedules are given in Appendices I and J, and the sampling procedure is summarized in Algorithm 1.

The key advantage of the multilevel construction is that it combines exact constraint satisfaction with local refinement while retaining a tractable sampling density. Rather than solving the Bianchi

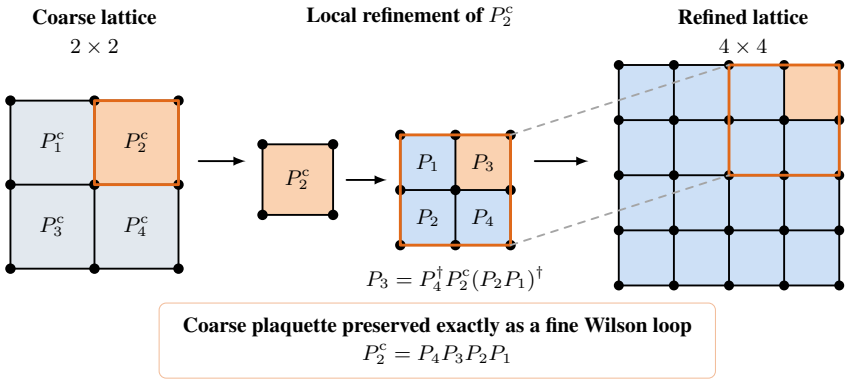


Figure 2: One refinement in plaquette space, for $d = 2$. **Left:** the coarse lattice; three plaquettes are free and P_2^c (orange) is fixed by the global closure. **Centre:** its top-right block, refined with P_1, P_2, P_4 (blue) generated and P_3 determined by $P_3 = P_4^\dagger P_2^c (P_2 P_1)^\dagger$. **Right:** the refined lattice; the orange outline marks the fine Wilson loop whose value is P_2^c.

Algorithm 1 One multilevel sample and its importance weight

$F^{(0)} \sim \widetilde{q}_0(\cdot), \quad \Omega \sim q_\Omega, \quad P^{(0)} \leftarrow R_0(F^{(0)}; \Omega), \quad \ell \leftarrow \log \widetilde{q}_0(F^{(0)})$ coarsest lattice
for $i = 1, \dots, n$ **do**
  $F^{(i)} \sim \widetilde{q}_i(\cdot \mid P^{(i-1)})$ learned conditional on $G^{N^{(i)}}$
  $P^{(i)} \leftarrow R_i(F^{(i)}; P^{(i-1)})$ exact deterministic refinement
  $\ell \leftarrow \ell + \log \widetilde{q}_i(F^{(i)} \mid P^{(i-1)})$
end for
$P \leftarrow P^{(n)}$; $S(P)$ from (2)
return $P, \quad \log q(P) = \ell, \quad \log w = -S(P) - \ell$

identities simultaneously, we enforce them level by level: every generated configuration lies on the constraint manifold by construction, while each determined plaquette depends on at most four newly generated variables. This avoids the scaling problem of the linear solve in higher-dimensional theories, as the matrix size of the linear systems remain small, independent of the lattice size. Because the refinement maps are bijective and Haar-volume preserving, the resulting multilevel density can be evaluated exactly and used for importance reweighting. At the same time, coarse plaquettes are preserved exactly as Wilson loops of the refined field, allowing long-range structure to be carried by the coarse levels while finer levels learn only the additional short-distance degrees of freedom. This is an important distinction from multiscale link-space flows (Abbott et al., 2024a), which use the hierarchy as a prior, and do not preserve the coarse-level link values as strict constraints on the final samples. These properties are reflected in the experiments of Section 4: the method scales more favourably with lattice size in two-dimensional $U(1)$, remains efficient as correlations grow in two-dimensional $SU(2)$, and shows an increasing advantage over link-space baselines toward weak coupling in four-dimensional $U(1)$.

4 Numerical Experiments

We evaluate our multilevel PSS in three complementary settings. Two-dimensional $U(1)$ provides a controlled test of scaling with lattice size along a line of constant physics; four-dimensional $U(1)$ probes its behaviour toward weak coupling, where link-space flows are known to deteriorate; and two-dimensional $SU(2)$ tests whether the advantage persists for a non-Abelian theory as the correlation length grows.

Evaluation Measures. For N generated samples $P_i \sim q$, we evaluate the KL divergence, the relative accuracy, and the effective sample size:

$$\mathrm{KL}(q\|p) = \log Z - \tfrac{1}{N}\textstyle\sum_i \log w_i, \qquad \mathrm{RA} = \hat{\mathcal{O}}/\mathcal{O}^{\mathrm{ref}} - 1, \qquad \mathrm{ESS} = \frac{(\sum_i w_i)^2}{N \sum_i w_i^2}, \tag{9}$$

where $w_i = \exp(-S(P_i))/q(P_i)$ is an unnormalized importance weight and $\hat{\mathcal{O}} = \frac{\sum_i w_i \mathcal{O}(P_i)}{\sum_i w_i}$ is the self-normalized importance sampling estimator for the observable $\mathcal{O}$.Unless stated otherwise $\mathcal{O}$ is the *mean plaquette*, the quantity summed by the action (2), averaged over lattice configurations. It is written $\langle \cos \varphi \rangle$ for $U(1)$ and $\langle \frac{1}{2} \operatorname{tr} U_p \rangle$ for $SU(2)$, and is what panel (b) of Figs. 3 to 5 reports. The true free energy $\log Z$ and reference values $\mathcal{O}^{\text{ref}}$ for the observables are computed independently by sufficiently accurate methods (see Appendices K and L). The KL divergence $\mathrm{KL}(q\|p)$ measures the discrepancy between the sampler density q and the target density p, with the corresponding values for some baseline methods reported in their original papers. RA allows us to assess both the bias and variance of the estimator, while ESS measures the sampling efficiency. Crucially, ESS is reliable only when the estimator is consistent with the reference value, i.e., when RA shows a small bias. Otherwise, ESS can be misleadingly large; for example, ESS $= 1$ when the sampler repeatedly generates the same configuration, regardless of the target distribution.

Baseline Methods. For each theory, we compare our multilevel sampler with a single-level plaquette sampler and with link-space flow baselines. The single-level plaquette model uses the same representation but solves all Bianchi constraints simultaneously (Appendix E). For two-dimensional $U(1)$, we implement two link-space baselines ourselves: a multiscale model (Abbott et al., 2024a) using the gauge-equivariant coupling layers (Kanwar et al., 2020) with rational-quadratic splines, and a single-scale version of the same flow applied directly to the full link lattice with a Haar base. For four-dimensional $U(1)$, we compare our sampler with the multiscale and single-scale link-space samplers of Abbott et al. (2024a). Since their code is not publicly available, we quote the KL-divergence values from Fig. 4 of Abbott et al. (2024a). For two-dimensional $SU(2)$, we train a continuous flow (Gerdes et al., 2025) using the released `bijx` library. Since neither training script nor pretrained weights are provided for our benchmark, we performed training and evaluation independently. Our implementation closely reproduces their reported results, yielding ESS $= 0.85$ and 0.79 at $\beta = 2.2$ and 2.7, respectively, compared with their reported ESS $= 0.87$ and 0.68.

Evaluation Protocol. Each trained model is evaluated once from a retained checkpoint using a fixed random seed, with all reported quantities (9)—KL divergence, RA, and ESS—computed from the same set of independent samples. Uncertainties are estimated by delete-block jackknife with 100 blocks over the same samples. All plaquette-space samplers and link-space baselines are evaluated with $N = 500{,}000$, except the two-dimensional $SU(2)$ link-space baseline at $N \leq 10^4$, where the continuous flow (Gerdes et al., 2025) generation is too costly—roughly one second per configuration. The smaller evaluation samples can bias finite-sample ESS estimates upward, so these choices do not artificially favour our plaquette-space models. Full estimator definitions and uncertainty procedures are given in Appendix L.

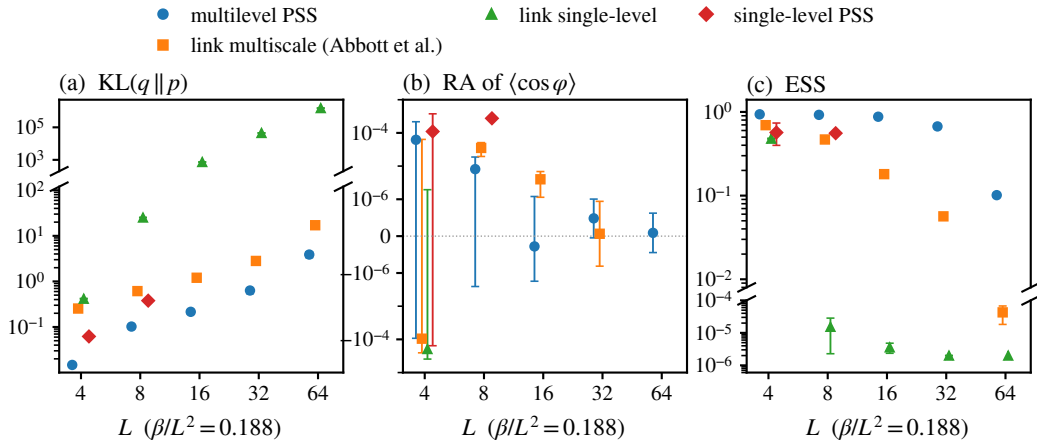


Figure 3: Two-dimensional $U(1)$ along the line of constant physics $\beta = 0.188L^2$. **(a)** $\mathrm{KL}(q\|p)$. **(b)** RA of the reweighted plaquette against the exact finite-volume value. **(c)** ESS. Panels (a) and (c) use logarithmic axes with one break; points across a break are not directly comparable.

4.1 RESULTS

Two-dimensional $U(1)$. We begin with two-dimensional $U(1)$ theory, which provides a direct test of scaling with lattice size. Along a line of constant physics, $\beta = cL^2$, increasing L keeps the physical theory fixed while resolving it on progressively finer lattices, so the correlation length in lattice units grows. In Fig. 3 (a)-(c), which show the three evaluation measures in Eq.(9), respectively, both single-scale (Kanwar et al., 2020) and multiscale (Abbott et al., 2024a) link-space baselines deteriorate rapidly with L at $c = 0.188$, yielding large KL divergences and large variances in observable estimates, as well as small ESS at $L = 32$ and 64. In contrast, our multilevel PSS yields significantly lower KL divergence, compatible observable estimates, and larger ESS. The advantage grows from $2.0\times$ at $L = 8$ to $4.9\times$ at $L = 16$ and $11.9\times$ at $L = 32$. At $L = 64$, the weights of the link-space flow are severely degenerate: $\mathrm{ESS} = 4 \times 10^{-5}$, with a single configuration carrying 16% of the total weight, compared with $\mathrm{ESS} = 0.10$ for our multilevel PSS.

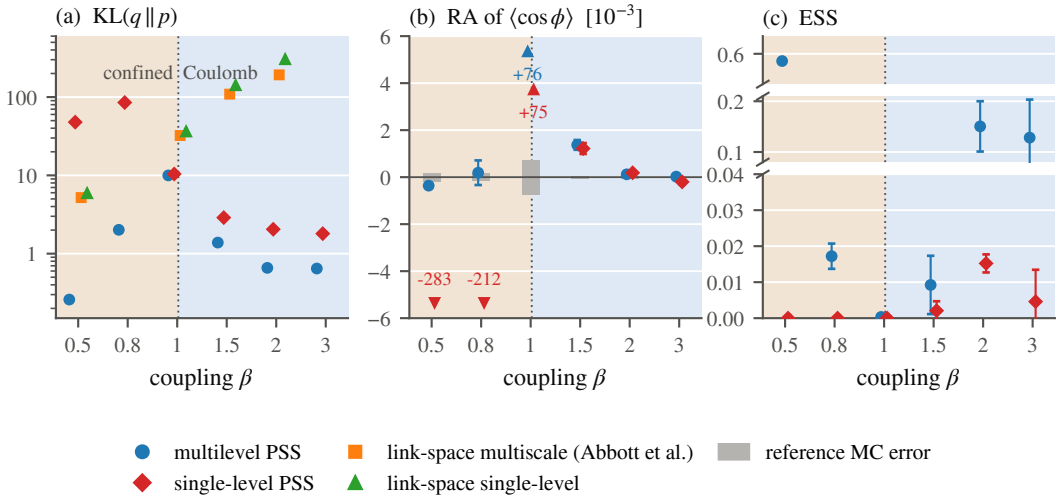


Figure 4: Four-dimensional $U(1)$. **(a)** $\mathrm{KL}(q\|p)$ as a function of β. **(b)** Relative Accuracy (RA) of the reweighted $\langle\cos\varphi\rangle$ against the heatbath reference, with the grey band indicating the reference's own uncertainty. **(c)** ESS, shown on a linear scale with two axis breaks. Values outside the displayed range are indicated at the panel boundary. Background shading denotes the confined and Coulomb phases, separated at $\beta_c \simeq 1.011$.

Four-dimensional $U(1)$. Next we test samplers on the $2^4 \to 4^4$ benchmark of Abbott et al. (2024a). As shown in Figure 4, our multilevel PSS achieves smaller KL divergences than both the single-level and multilevel link-space baseline samplers. Around the first-order transition at $\beta \simeq 1.011$, indicated by the background shading, all samplers perform poorly. Addressing this regime is beyond the scope of this work, as our multilevel PSS is designed to mitigate the issues that arise in the weak-coupling regime at larger β, where the system is in the massless Coulomb phase and correlations are long-ranged, with their effective extent ultimately limited by the finite lattice size. This is precisely where the multiscale link-space construction deteriorates: its KL divergence increases with β, whereas our PSSs improve, leading to a rapidly widening performance gap. As observed in Figure 4 (a), the gap in terms of the KL divergence exceeds two orders of magnitude at larger β.

Two-dimensional $SU(2)$. Two-dimensional $SU(2)$ tests the same construction in a non-Abelian theory as correlations become increasingly long-ranged. On the 16^2 lattice, the correlation length grows from roughly one lattice spacing to 6.6, so at the largest coupling it spans about 40% of the lattice extent. The link-space flow is competitive only when correlations are short and rapidly loses efficiency as they grow. The single-level plaquette sampler is more robust, remaining close to the multilevel sampler at $\xi/a = 2.2$, but it too deteriorates at longer correlation lengths and becomes effectively inefficient by $\xi/a = 6.6$. Our multilevel plaquette-space sampler remains efficient throughout this regime and substantially outperforms both baselines; to our knowledge, no multilevel link-space sampler has been proposed for two-dimensional $SU(2)$.

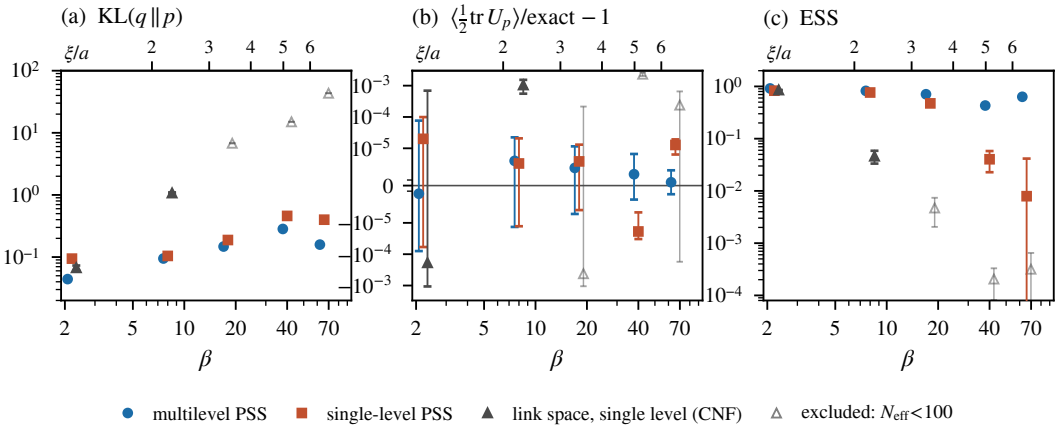


Figure 5: Two-dimensional $SU(2)$, $L = 16$. **(a)** $\mathrm{KL}(q\|p)$ in nats. **(b)** Reweighted $\langle \frac{1}{2} \operatorname{Tr} U_p \rangle$ relative to the exact finite-volume value, signed. **(c)** ESS. Three samplers: the multilevel and single-level plaquette flows and a link-space continuous flow. The upper axis is the correlation length ξ/a. Faded symbols mark $n_{\text{eff}} < 100$.

Taken together, the three experiments show a consistent pattern. In two-dimensional $U(1)$ our multilevel PSS scales substantially better with lattice size than the link-space baselines, while in four-dimensional $U(1)$ and two-dimensional $SU(2)$, its relative advantage grows toward weak coupling and longer correlations. Although the single-level PSS performs comparably to the multilevel counterpart for small lattice sizes, its performance degrades as the lattice size increases, and scaling it further to larger lattices in higher-dimensional theories could become computationally challenging, as discussed in Appendix G.1.

5 Conclusion

We introduced a multilevel plaquette-space sampler (PSS) that satisfies the lattice Bianchi identities exactly while preserving a tractable density. The construction combines two complementary advantages: plaquette variables make the action local and remove the local gauge redundancy, while the coarse-to-fine hierarchy keeps the remaining constraint structure local—avoiding the scaling issues of solving the Bianchi-constrained problem—and carries long-range information across scales.

The numerical results show that this combination becomes increasingly beneficial in the difficult regimes considered here: as we move toward the continuum limit in two-dimensional $U(1)$, toward weak coupling in four-dimensional $U(1)$, and as correlations lengthen in two-dimensional $SU(2)$. These results motivate extending the method to larger four-dimensional lattices and non-Abelian gauge theories, where exact local constraint handling and multilevel scale separation may offer further advantages. More broadly, the same strategy may be useful for other high-dimensional generative-sampling problems with structured constraints.

6 Limitations

Our current experiments are restricted to relatively small lattices and to a single training run per model, and the extension to large four-dimensional non-Abelian systems remains to be demonstrated. In addition, although the Bianchi constraints are satisfied exactly by construction, sampling quality can still deteriorate in regimes with rare structures that are difficult for the continuous flow to represent rare defect configurations, such as monopoles at weak coupling in four dimensions, are under-sampled, which is beyond the scope of this work. Addressing these effects will require larger-scale studies and, potentially, more expressive models that explicitly capture such discrete or topological degrees of freedom.

AI USE STATEMENT

Generative AI tools were used to assist in language editing, literature discovery, feedback on research methodology and experimental design, and aspects of code development, organization, and finalization. They were not used to generate simulation data or report numerical results. All simulations, training, and evaluation runs, and quantitative results reported in this work, were produced by the authors using the procedures described in the paper. All AI-assisted text, references, and code were reviewed and verified by the authors. Scientific ideas, methodological choices, implementation decisions, interpretation of results, and final analysis remain the responsibility of the authors. The authors assume full responsibility for the content of the article and the accompanying research artifacts.

REPRODUCIBILITY STATEMENT

All code, stored evaluation results and launch scripts are provided as anonymous supplementary material, organised as four packages, one for each theory ($U(1)$ in two and four dimensions, $SU(2)$ in two dimensions) and one for the solve-support and conditioning analysis, each at three levels. Level 1 regenerates every figure and table of the paper from the stored evaluation results on a CPU in seconds, so every quoted number can be checked without training or a GPU. Level 2 re-evaluates the trained models with the protocol of Appendix L, $N = 500{,}000$ samples, one evaluation per model at a fixed seed, delete-block jackknife uncertainties; the checkpoints are withheld at review to preserve anonymity and will be deposited in a public archive at camera-ready, and each package records their SHA-256 hashes. Level 3 gives the exact launch command of every model with the architectures and schedules of Appendices I and J, resuming optimizer and scheduler state exactly across interrupted runs. Reference values are closed-form in two dimensions and come from heatbath and thermodynamic integration in four (Appendix K); the numerical gates of Appendix H and the conditioning scan of Appendix G.1 run from a self-contained reference implementation in the analysis package. Evaluations are bit-identical only on the same GPU model and driver; on other hardware they reproduce to within the quoted uncertainties.

REFERENCES

Gert Aarts, Diaa E. Habibi, Andreas Ipp, David I. Müller, Thomas R. Ranner, Lingxiao Wang, Wei Wang, and Qianteng Zhu. Generalizable equivariant diffusion models for non-abelian lattice gauge theory, 2026. URL https://arxiv.org/abs/2601.19552.

Ryan Abbott, Michael S. Albergo, Denis Boyda, Daniel C. Hackett, Gurtej Kanwar, Fernando Romero-López, Phiala E. Shanahan, and Julian M. Urban. Multiscale normalizing flows for gauge theories. In *Proceedings of the 40th International Symposium on Lattice Field Theory*, volume LATTICE2023 of *PoS*, pp. 035, 2024a. doi: 10.22323/1.453.0035.

Ryan Abbott, Aleksandar Botev, Denis Boyda, Daniel C. Hackett, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, Fernando Romero-López, Phiala E. Shanahan, and Julian M. Urban. Applications of flow models to the generation of correlated lattice QCD ensembles. *Phys. Rev. D*, 109(9):094514, 2024b. doi: 10.1103/PhysRevD.109.094514.

Ryan Abbott, Denis Boyda, Yang Fu, Daniel C. Hackett, Gurtej Kanwar, Fernando Romero-López, Phiala E. Shanahan, and Julian M. Urban. Variance reduction in lattice QCD observables via normalizing flows, 3 2026. URL https://arxiv.org/abs/2603.02984. arXiv preprint.

Ryan Abbott et al. Aspects of scaling and scalability for flow-based sampling of lattice QCD. *Eur. Phys. J. A*, 59(11):257, 2023a. doi: 10.1140/epja/s10050-023-01154-w.

Ryan Abbott et al. Normalizing flows for lattice gauge theory in arbitrary space-time dimension. *arXiv:2305.02402*, 2023b.

David Albandea, Pilar Hernández, Alberto Ramos, and Fernando Romero-López. Topological sampling through windings. *Eur. Phys. J. C*, 81(10):873, 2021. doi: 10.1140/epjc/s10052-021-09677-6.

Michael S. Albergo, Gurtej Kanwar, and Phiala E. Shanahan. Flow-based generative models for markov chain monte carlo in lattice field theory. *Phys. Rev. D*, 100:034515, 2019. doi: 10.1103/PhysRevD.100.034515.

Michael S. Albergo, Denis Boyda, Daniel C. Hackett, Gurtej Kanwar, Kyle Cranmer, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan. Introduction to normalizing flows for lattice field theory, 2021a. arXiv:2101.08176.

Michael S. Albergo, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, Julian M. Urban, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, and Phiala E. Shanahan. Flow-based sampling for fermionic lattice field theories. *Phys. Rev. D*, 104(11):114507, 2021b. doi: 10.1103/PhysRevD.104.114507.

Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, Fernando Romero-López, Phiala E. Shanahan, and Julian M. Urban. Flow-based sampling in the lattice Schwinger model at criticality. *Phys. Rev. D*, 106(1):014514, 2022. doi: 10.1103/PhysRevD.106.014514.

Simone Bacchio, Pan Kessel, Stefan Schaefer, and Lorenz Vaitl. Learning trivializing gradient flows for lattice gauge theories. *Phys. Rev. D*, 107(5):L051504, 2023. doi: 10.1103/PhysRevD.107.L051504.

G. G. Batrouni and M. B. Halpern. String, corner, and plaquette formulation of finite lattice gauge theory. *Physical Review D*, 30:1782–1790, 1984. doi: 10.1103/PhysRevD.30.1782.

Ghassan G. Batrouni. Plaquette formulation and the Bianchi identity for lattice gauge theories. *Nuclear Physics B*, 208:467–483, 1982. doi: 10.1016/0550-3213(82)90231-0.

Marc Bauer, Renzo Kapust, Jan M Pawlowski, and Finn Leon Temmen. Super-resolving normalising flows for lattice field theories. *SciPost Physics*, 19(3):077, 2025.

Piotr Białas, Piotr Korcyl, and Tomasz Stebel. Hierarchical autoregressive neural networks for statistical systems. *Computer Physics Communications*, 281:108502, 2022. doi: 10.1016/j.cpc.2022.108502.

Piotr Białas, Vaibhav Chahar, Piotr Korcyl, Tomasz Stebel, Mateusz Winiarski, and Dawid Zapolski. Hierarchical autoregressive neural networks in three-dimensional statistical system. *Comput. Phys. Commun.*, 318:109892, 2026. doi: 10.1016/j.cpc.2025.109892.

Claudio Bonanno, Claudio Bonati, and Massimo D'Elia. Large-N $SU(N)$ yang–mills theories with milder topological freezing. *JHEP*, 03:111, 2021. doi: 10.1007/JHEP03(2021)111.

Claudio Bonanno, Andrea Bulgarelli, Elia Cellini, Alessandro Nada, Dario Panfalone, Davide Vadacchino, and Lorenzo Verzichelli. A scalable flow-based approach to mitigate topological freezing. *arXiv e-prints*, 2026a. arXiv:2601.20708.

Claudio Bonanno, Andrea Bulgarelli, Elia Cellini, Alessandro Nada, Dario Panfalone, Davide Vadacchino, and Lorenzo Verzichelli. Scaling flow-based approaches for topology sampling in $\mathrm{SU}(3)$ gauge theory. *JHEP*, 04:051, 2026b. doi: 10.1007/JHEP04(2026)051. arXiv:2510.25704.

Denis Boyda, Gurtej Kanwar, Sébastien Racanière, Danilo Jimenez Rezende, Michael S. Albergo, Kyle Cranmer, Daniel C. Hackett, and Phiala E. Shanahan. Sampling using $SU(N)$ gauge equivariant flows. *Phys. Rev. D*, 103:074504, 2021. doi: 10.1103/PhysRevD.103.074504.

Andrea Bulgarelli, Elia Cellini, Karl Jansen, Stefan Kühn, Alessandro Nada, Shinichi Nakajima, Kim A. Nicoli, and Marco Panero. Flow-Based Sampling for Entanglement Entropy and the Machine Learning of Defects. *Phys. Rev. Lett.*, 134(15):151601, 2025a. doi: 10.1103/PhysRevLett.134.151601.

Andrea Bulgarelli, Elia Cellini, and Alessandro Nada. Scaling of stochastic normalizing flows in SU(3) lattice gauge theory. *Phys. Rev. D*, 111(7):074517, 2025b. doi: 10.1103/PhysRevD.111.074517.

Michele Caselle, Gianluca Costagliola, Alessandro Nada, Marco Panero, et al. Jarzynski's theorem for lattice gauge theory. *Phys. Rev. D*, 94:034503, 2016. doi: 10.1103/PhysRevD.94.034503.

Michele Caselle, Elia Cellini, Alessandro Nada, and Marco Panero. Stochastic normalizing flows as non-equilibrium transformations. *JHEP*, 07:015, 2022a. doi: 10.1007/JHEP07(2022)015.

Michele Caselle, Elia Cellini, Alessandro Nada, and Marco Panero. Stochastic normalizing flows for lattice field theory. *PoS*, LATTICE2022:005, 2022b. doi: 10.22323/1.430.0005. arXiv:2210.03139.

Michele Caselle, Elia Cellini, and Alessandro Nada. Numerical determination of the width and shape of the effective string using Stochastic Normalizing Flows. *JHEP*, 02:090, 2025. doi: 10.1007/JHEP02(2025)090.

Marco Cè, Leonardo Giusti, and Stefan Schaefer. Domain decomposition, multilevel integration, and exponential noise reduction in lattice qcd. *Phys. Rev. D*, 93:094507, 2016. doi: 10.1103/PhysRevD.93.094507.

Sourav Chatterjee and Persi Diaconis. The sample size required in importance sampling. *Annals of Applied Probability*, 28(2):1099–1135, 2018. doi: 10.1214/17-AAP1326.

Kyle Cranmer, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, and Phiala E. Shanahan. Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics. *Nature Rev. Phys.*, 5(9):526–535, 2023. doi: 10.1038/s42254-023-00616-w.

Luigi Del Debbio, Gian Mario Manca, and Ettore Vicari. Critical slowing down of topological modes. *Phys. Lett. B*, 594:315–323, 2004. doi: 10.1016/j.physletb.2004.05.038.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. In *5th International Conference on Learning Representations (ICLR 2017)*, Toulon, France, 2017.

Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987. doi: 10.1016/0370-2693(87)91197-X.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In *Advances in Neural Information Processing Systems*, volume 32, pp. 7509–7520, 2019.

Stavros Efthymiou, Matthew J. S. Beach, and Roger G. Melko. Super-resolving the ising model with convolutional neural networks. *Phys. Rev. B*, 99:075113, 2019. doi: 10.1103/PhysRevB.99.075113.

Michael G. Endres, Richard C. Brower, William Detmold, Kostas Orginos, and Andrew V. Pochinsky. Multiscale Monte Carlo equilibration: Pure Yang-Mills theory. *Phys. Rev. D*, 92(11):114516, 2015. doi: 10.1103/PhysRevD.92.114516.

M. Faas and H.J. Hilhorst. Hierarchical monte carlo simulation of the ising model. *Physica A: Statistical Mechanics and its Applications*, 135(2):571–590, 1986. ISSN 0378-4371. doi: 10.1016/0378-4371(86)90161-5. URL https://www.sciencedirect.com/science/article/pii/0378437186901615.

Ali Faraz, Ankur Singha, Dipankar Chakrabarti, Shinichi Nakajima, and Vipul Arora. Improvement of Heatbath Algorithm in LFT using Generative models. *PoS*, LATTICE2024:036, 2025. doi: 10.22323/1.466.0036. arXiv:2308.08615.

Jacob Finkenrath. Fine grinding localized updates via gauge equivariant flows in the 2d schwinger model. *PoS*, LATTICE2023:022, 2024. doi: 10.22323/1.453.0022.

Christof Gattringer and Christian B. Lang. *Quantum chromodynamics on the lattice*, volume 788. Springer, Berlin, 2010. ISBN 978-3-642-01849-7, 978-3-642-01850-3. doi: 10.1007/978-3-642-01850-3.

Mathis Gerdes, Pim de Haan, Corrado Rainone, Roberto Bondesan, et al. Learning lattice quantum field theories with equivariant continuous flows. *SciPost Physics*, 15:238, 2023.

Mathis Gerdes, Pim de Haan, Roberto Bondesan, and Miranda C. N. Cheng. Nonperturbative trivializing flows for lattice gauge theories. *Phys. Rev. D*, 112(9):094516, 2025. doi: 10.1103/31d5-hvp6.

Michael B. Giles. Multilevel monte carlo methods. *Acta Numerica*, 24:259–328, 2015. doi: 10.1017/S096249291500001X.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pp. 2672–2680, 2014.

J. Goodman and A. D. Sokal. Multigrid Monte Carlo Method for Lattice Field Theories. *Phys. Rev. Lett.*, 56:1015–1018, 1986. doi: 10.1103/PhysRevLett.56.1015.

J. Goodman and A. D. Sokal. Multigrid Monte Carlo method: conceptual foundations. *Phys. Rev. D*, 40:2035–2071, 1989. doi: 10.1103/PhysRevD.40.2035.

Martin Grabenstein and Klaus Pinn. Multigrid Monte Carlo algorithms for SU(2) lattice gauge theory: Two versus four dimensions. *Phys. Rev. D*, 50(11):6998–7010, 1994. doi: 10.1103/PhysRevD.50.6998.

Daniel C. Hackett, Chung-Chun Hsieh, Sahil Pontula, Michael S. Albergo, Denis Boyda, Jiunn-Wei Chen, Kai-Feng Chen, Kyle Cranmer, Gurtej Kanwar, and Phiala E. Shanahan. Flow-based sampling for multimodal and extended-mode distributions in lattice field theory. *arXiv e-prints*, 7 2021. arXiv:2107.00734.

Martin Hasenbusch. Fighting topological freezing in the two-dimensional CP^{N-1} model. *Phys. Rev. D*, 96:054504, 2017. doi: 10.1103/PhysRevD.96.054504.

Anna Hasenfratz, Ethan T. Neil, Letizia Parato, and Noah Schwartz. Renormalization-guided cascade upscaling for lattice field generation. *arXiv preprint*, Aug 2026a. arXiv:2608.28581 [hep-lat].

Anna Hasenfratz, Ethan T. Neil, Letizia Parato, and Noah Schwartz. Renormalization-guided inverse blocking for lattice field generation: construction and validation. *arXiv preprint*, Aug 2026b. arXiv:2608.28580 [hep-lat].

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Hong-Ye Hu, Dian Wu, Yi-Zhuang You, Bruno Olshausen, et al. Rg-flow: a hierarchical and explainable flow model based on renormalization group and sparse prior. *Machine Learning: Science and Technology*, 3(3):035009, 2022. doi: 10.1088/2632-2153/ac8393.

Gurtej Kanwar and Octavio Vega. Spectral Diffusion for Sampling on $\mathrm{SU}(N)$. In *42nd International Symposium on Lattice Field Theory*, 12 2025. arXiv:2512.19877.

Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan. Equivariant flow-based sampling for lattice gauge theory. *Physical Review Letters*, 125(12), Sep 2020. ISSN 1079-7114. doi: 10.1103/physrevlett.125.121601. URL `http://dx.doi.org/10.1103/PhysRevLett.125.121601`.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations (ICLR 2014)*, Banff, Canada, 2014.

Maciej Koch-Janusz and Zohar Ringel. Mutual information, neural networks and the renormalization group. *Nature Physics*, 14(6):578–582, 2018. doi: 10.1038/s41567-018-0081-4.

Janik Kreit, Andrea Bulgarelli, Lena Funcke, Thomas Luu, Dominic Schuh, Simran Singh, and Lorenzo Verzichelli. Toward Scalable Normalizing Flows for the Hubbard Model. In *42nd International Symposium on Lattice Field Theory*, 1 2026. arXiv:2601.18273.

Alessandro Laio, Guido Martinelli, and Francesco Sanfilippo. Metadynamics surfing on topology barriers: the CP^{N-1} case. *JHEP*, 07:089, 2016. doi: 10.1007/JHEP07(2016)089.

Shuo-Hui Li and Lei Wang. Neural network renormalization group. *Phys. Rev. Lett.*, 121:260601, 2018. doi: 10.1103/PhysRevLett.121.260601.

Martin Lüscher and Stefan Schaefer. Lattice QCD without topology barriers. *JHEP*, 07:036, 2011. doi: 10.1007/JHEP07(2011)036.

Nobuyuki Matsumoto, Richard C. Brower, and Taku Izubuchi. Decimation map in 2d for accelerating hmc. *PoS*, LATTICE2023:033, 2023. doi: 10.22323/1.453.0033.

Kim A. Nicoli, Shinichi Nakajima, Nils Strodthoff, Wojciech Samek, et al. Asymptotically unbiased estimation of physical observables with neural samplers. *Phys. Rev. E*, 101:023304, 2020. doi: 10.1103/PhysRevE.101.023304.

Kim A. Nicoli, Christopher J. Anders, Lena Funcke, Tobias Hartung, et al. Estimation of thermodynamic observables in lattice field theories with deep generative models. *Phys. Rev. Lett.*, 126: 032001, 2021. doi: 10.1103/PhysRevLett.126.032001.

Kim A. Nicoli, Christopher J. Anders, Tobias Hartung, Karl Jansen, et al. Detecting and mitigating mode-collapse for flow-based sampling of lattice field theories. *Phys. Rev. D*, 108:114501, 2023. doi: 10.1103/PhysRevD.108.114501.

Jan M Pawlowski and Julian M Urban. Reducing autocorrelation times in lattice simulations with generative adversarial networks. *Machine Learning: Science and Technology*, 1(4):045011, oct 2020. doi: 10.1088/2632-2153/abae73. URL https://doi.org/10.1088/2632-2153/abae73.

Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1530–1538. PMLR, 2015.

Alexander Rothkopf. Jump-Diffusion Stochastic Quantization for Euclidean Lattice Field Theories. *preprint arXiv*, 8 2026. arXiv:2608.16451.

Stefan Schaefer, Rainer Sommer, and Francesco Virotta. Critical slowing down and error analysis in lattice QCD simulations. *Nucl. Phys. B*, 845:93–119, 2011.

K. E. Schmidt. Using renormalization-group ideas in monte carlo sampling. *Phys. Rev. Lett.*, 51: 2175–2178, Dec 1983. doi: 10.1103/PhysRevLett.51.2175. URL https://link.aps.org/doi/10.1103/PhysRevLett.51.2175.

Dominic Schuh, Lena Funcke, Janik Kreit, Thomas Luu, and Simran Singh. Tackling the Sign Problem in the Doped Hubbard Model with Normalizing Flows. *arXiv*, March 2026. arXiv:2603.18205.

A Singha, J Kauffmann, E Cellini, K Jansen, and S Nakajima. Scalable generative sampling and multilevel estimation for lattice field theories near criticality. *arXiv preprint arXiv:2604.10209*, 2026a.

Ankur Singha, Dipankar Chakrabarti, and Vipul Arora. Generative learning for the problem of critical slowing down in lattice Gross-Neveu model. *SciPost Phys. Core*, 5:052, 2022. doi: 10.21468/SciPostPhysCore.5.4.052. URL https://scipost.org/10.21468/SciPostPhysCore.5.4.052.

Ankur Singha, Dipankar Chakrabarti, and Vipul Arora. Conditional normalizing flow for Markov chain Monte Carlo sampling in the critical region of lattice field theory. *Phys. Rev. D*, 107:014512, 2023a. doi: 10.1103/PhysRevD.107.014512.

Ankur Singha, Dipankar Chakrabarti, and Vipul Arora. Sampling u(1) gauge theory using a retrainable conditional flow-based model. *Phys. Rev. D*, 108(7):074518, 2023b. doi: 10.1103/PhysRevD.108.074518.

Ankur Singha, Elia Cellini, Kim A. Nicoli, Karl Jansen, Stefan Kühn, and Shinichi Nakajima. Multilevel generative samplers for investigating critical phenomena. In *The Thirteenth International Conference on Learning Representations (ICLR 2025)*, 2025. URL https://openreview.net/forum?id=YcUV5apdlq.

Ankur Singha, Jacob Kauffmann, Karl Jansen, Jacob Finkenrath, Vipul Arora, and Shinichi Nakajima. Sector-Resolved Flow Sampling for Topologically Frozen Lattice Gauge Theories. *preprint arXiv*, 9 2026b. arXiv:2609.07781.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265. PMLR, 2015.

Octavio Vega, Javad Komijani, Aida El-Khadra, and Marina Marinkovic. Group-equivariant diffusion models for lattice field theory. *JHEP*, 07:213, 2026. doi: 10.1007/JHEP07(2026)213. arXiv:2510.26081.

Lingxiao Wang, Gert Aarts, and Kai Zhou. Diffusion models as stochastic quantization in lattice field theory. *JHEP*, 05:060, 2024. doi: 10.1007/JHEP05(2024)060.

Sebastian J. Wetzel. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Physical Review E*, 96(2):022140, 2017. doi: 10.1103/PhysRevE.96.022140.

Kenneth G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, 1974. doi: 10.1103/PhysRevD.10.2445.

Ulli Wolff. Critical slowing down. *Nuclear Physics B - Proceedings Supplements*, 17:93–102, 1990. doi: 10.1016/0920-5632(90)90224-I.

Dian Wu, Lei Wang, and Pan Zhang. Solving statistical mechanics using variational autoregressive networks. *Phys. Rev. Lett.*, 122:080602, Feb 2019. doi: 10.1103/PhysRevLett.122.080602. URL `https://link.aps.org/doi/10.1103/PhysRevLett.122.080602`.

Qianteng Zhu, Gert Aarts, Wei Wang, Kai Zhou, and Lingxiao Wang. Physics-conditioned diffusion models for lattice gauge theory. *arXiv*, 2025. arXiv:2502.05504.

A Extended related work

This appendix expands Section 2.2: how flow samplers degrade with volume, the hierarchical samplers our construction follows, the precise relation to the closest one, and the literature on topological freezing that bounds what the method currently achieves.

Generative sampling for lattice field theory. Modern generative modelling includes several major families, including variational autoencoders (VAEs) (Kingma & Welling, 2014), generative adversarial networks (GANs) (Goodfellow et al., 2014), normalizing flows (Rezende & Mohamed, 2015; Dinh et al., 2017), and diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020). These have been adapted to statistical physics and lattice field theory, where a generative model can represent a high-dimensional configuration distribution and serve as an alternative sampling algorithm. Flow-based sampling has received particular attention because its tractable density allows generated configurations to be corrected by importance reweighting or Metropolis acceptance. Normalizing flows for scalar lattice field theory were introduced by Albergo et al. (2019) and extended to gauge theories (Kanwar et al., 2020; Boyda et al., 2021), to stochastic flow constructions (Caselle et al., 2022a;b; 2025), and to conditional sampling between theory parameters (Singha et al., 2023a; Faraz et al., 2025; Singha et al., 2023b). Other paradigms have also been explored: GAN-based samplers reduce autocorrelation times in scalar theory (Pawlowski & Urban, 2020) and address critical slowing in the Gross–Neveu model (Singha et al., 2022), variational autoencoders identify phase structure (Wetzel, 2017), variational autoregressive networks give normalized samplers for spin systems (Wu et al., 2019), and diffusion models have been connected to stochastic quantization (Wang et al., 2024) and extended to gauge theories (Zhu et al., 2025).

Scaling with volume. Flow quality is known to deteriorate with volume (Abbott et al., 2023a), and reverse-KL training can collapse onto only part of a multimodal target (Hackett et al., 2021; Nicoli et al., 2023). For local theories the extensivity of KL, together with sample-complexity bounds for importance sampling (Chatterjee & Diaconis, 2018), makes this particularly severe at large volume. Exact reweighting remains possible whenever the proposal has sufficient support (Nicoli et al., 2020; 2021), so the practical question is not correctness but the cost of an accurate proposal. We address this

through the representation and the constraint structure rather than by introducing a new density model: the constraint layer of Appendix E.2 can be combined with different tractable generative models, and we use spline couplings (Durkan et al., 2019) as a convenient choice for circular variables.

Hierarchical and multiscale samplers. Coarse-to-fine sampling has a long history in lattice field theory: multigrid Monte Carlo (Goodman & Sokal, 1986; 1989; Grabenstein & Pinn, 1994), renormalization-group-guided updates (Schmidt, 1983; Faas & Hilhorst, 1986), multiscale thermalization (Endres et al., 2015), decimation maps (Matsumoto et al., 2023), and multilevel variance reduction (Cè et al., 2016; Giles, 2015). Learned versions followed. In lattice gauge theory, renormalization-group schemes have been combined with normalizing flows for $U(1)$ (Finkenrath, 2024) and, in the construction we use as a baseline, for $SU(3)$ (Abbott et al., 2024a). For discrete spin systems, hierarchical autoregressive networks (Białas et al., 2022) sample regions of a configuration in parallel with a shared autoregressive network, following a recursive domain decomposition (Cè et al., 2016), and have since been carried to three dimensions (Białas et al., 2026). Closest in spirit is RiGCS (Singha et al., 2025), which builds a configuration recursively from coarse to fine and conditions each level on the one below, using conditional flows of the kind introduced by Singha et al. (2023a) and carried to $U(1)$ gauge theory by Singha et al. (2023b); related constructions extend the idea to continuous lattice field theories as multilevel conditional flows (Singha et al., 2026a) and as super-resolution of configurations (Bauer et al., 2025; Efthymiou et al., 2019). Our refinement levels are conditional in exactly this sense. What is new is that the refined variables are *constrained*, so every level must additionally solve the Bianchi identity exactly, and the contribution is that this solve can be kept local.

The closest precedent, and generative modelling under constraints. The closest precedent in lattice gauge theory is the multiscale flow of Abbott et al. (2024a), which also refines one lattice direction at a time, but in *link space*. There the hierarchy serves as a prior for a separate fine-lattice flow of 48 gauge-equivariant coupling layers on every link, which the authors state is required because the hierarchy captures only a subset of the correlations; that flow is unconstrained by the hierarchy and carries the final density. Here no such flow follows. Where a single-direction refinement leaves some plaquette orientations last generated at a coarser resolution (four-dimensional $U(1)$ at $\beta = 0.5, 0.8, 2, 3$), we add one further pass of the same refinement per direction at the finest lattice: the field is re-parametrised as its blocking along μ and the free variables of that refinement, only the free variables are transformed, conditioned on the blocked field, and the constraint solve is unchanged (Table 6). Each pass fixes the blocking along its own direction, and the density remains the product of exact conditionals (7). More importantly, working in plaquette space introduces exact Bianchi constraints. The central issue is therefore not anisotropic refinement itself, but how that refinement changes the structure of the constraint solve: a global single-level parametrisation becomes nonlocal and increasingly ill conditioned, while the multilevel construction keeps the solve local and its size and conditioning independent of lattice extent (Appendix G.1). This notion of *solve locality* is the main distinction from existing hierarchical gauge-field samplers. It also places the work within generative modelling under exact constraints. Gauge redundancy is commonly handled through equivariance (Kanwar et al., 2020; Boyda et al., 2021); moving to gauge-invariant variables removes that redundancy but introduces equality constraints between the variables. Soft penalties do not enforce these exactly, while projection generally changes the induced density and complicates exact likelihood evaluation. We instead parametrise the constraint manifold directly, and its two-dimensional realization (Appendix F.4) shows that the mechanism is not specific to any one dimension.

Topology and defects. The severe slowing down of topological modes toward the continuum limit (Del Debbio et al., 2004; Schaefer et al., 2011; Lüscher & Schaefer, 2011) has motivated dedicated algorithms, including open boundary conditions (Lüscher & Schaefer, 2011), metadynamics (Laio et al., 2016), tempering (Hasenbusch, 2017; Bonanno et al., 2021), and winding updates (Albandea et al., 2021). Generative approaches are more recent (Bonanno et al., 2026a;b; Rothkopf, 2026; Singha et al., 2026b) and show that global learned proposals can substantially improve transitions between weakly connected sectors, whether by resolving the sectors explicitly (Singha et al., 2026b) or by letting the stochastic dynamics jump between them (Rothkopf, 2026). In four-dimensional compact $U(1)$ the relevant defects are magnetic monopoles together with the fluxes through the six two-tori. Our parametrisation spans the full link configuration space and represents monopoles

exactly through integer lifts of the determined plaquettes (Appendix M); at weak coupling nonzero global flux sectors are suppressed by $e^{-2\pi^2\beta}$, so the trivial-sector heatbath reference reproduces the full theory to exponential accuracy (Appendix K). Sampling such rare defect configurations at the correct rate is a separate problem, beyond the scope of this work; the natural route is to treat the discrete defect variables explicitly, in the same spirit that winding updates introduce dedicated topology-changing moves for Markov chains. In two dimensions, where the sectors can be sampled directly, the multilevel construction keeps the topological autocorrelation time near unity up to $L = 32$, at a coupling where the link-space baselines have already lost a resolved ESS (Appendix N).

B Degrees of freedom on T^d

B.1 Tree gauge, holonomies and the link–plaquette pairing

Tree gauge. Because of local gauge symmetry, the link variables contain redundant degrees of freedom. A *tree gauge* removes this redundancy by choosing a maximal spanning tree $\mathcal{T}$ of $V-1$ links connecting all V sites without forming a loop. Starting from a root, the gauge transformation (3) can be used successively to set every tree link to $\mathbf{1}$. This fixes the gauge everywhere up to the transformation at the root and leaves $dV-(V-1)=(d-1)V+1$ independent links. In Fig. 6 we use a comb-shaped tree: the bottom row without its wrap-around link together with all vertical links outside the top row, $(L-1)+L(L-1)=V-1$.

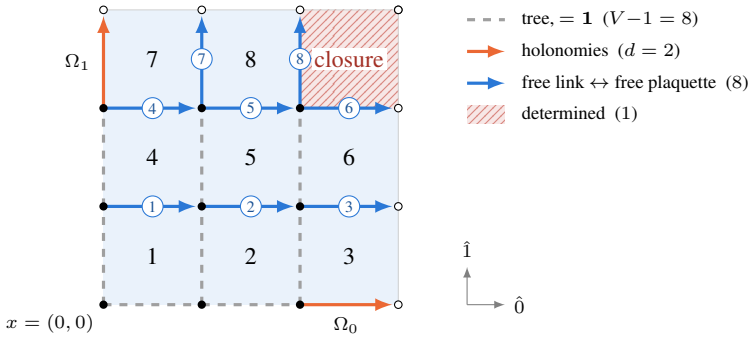


Figure 6: Degrees of freedom on a periodic 3×3 lattice in $d=2$ ($V=9$), shown in comb gauge. Dashed grey links form the maximal tree and are fixed to $\mathbf{1}$. Orange links are the holonomies Ω_0, Ω_1. The eight numbered blue links pair one-to-one with the eight free plaquettes, while the hatched corner plaquette is fixed by the global closure (5). Hollow sites denote periodic images.

Holonomies. Among the remaining links, exactly d close the non-contractible cycles of the torus, one in each direction. Since every other link on such a cycle belongs to the tree and equals $\mathbf{1}$, the ordered product around the cycle reduces to the closing link itself: this is the Polyakov loop Ω_μ. The residual gauge transformation at the root acts by conjugation, $\Omega_\mu \mapsto g\,\Omega_\mu\,g^\dagger$, so its conjugacy class is gauge invariant. Unlike plaquettes and their products, these loops are non-contractible and therefore cannot be reconstructed from plaquettes alone.

Plaquettes and links. A plaquette sampler must distinguish the plaquettes that can be chosen freely from those fixed by the lattice constraints. Of the $\binom{d}{2}V$ plaquettes, the Bianchi identities for $d\geq 3$ and the global closures determine a subset D, leaving a free set F. The same counting appears from the link side: after removing the d holonomies from the $(d-1)V+1$ gauge-fixed links, $(d-1)(V-1)$ degrees of freedom remain. Thus

$$\underbrace{(d-1)V+1}_{\text{gauge-fixed links}} = \underbrace{d}_{\text{holonomies}} + \underbrace{(d-1)(V-1)}_{\text{free plaquettes}}, \qquad \underbrace{\tbinom{d}{2}V}_{\text{plaquettes}} = \underbrace{(d-1)(V-1)}_{\text{free}} + \underbrace{|D|}_{\text{determined}}. \tag{10}$$

This agreement is explicit in tree gauge: the $(d-1)(V-1)$ non-tree links pair one-to-one with the free plaquettes. Adding them back sequentially, each closes a plaquette whose other links are already known, so (1) determines the link from the plaquette, and vice versa. The remaining plaquettes are

then fixed by the Bianchi identities and global closures. Free plaquettes together with the holonomies therefore provide coordinates for gauge-fixed link fields. For $d = 2$, $V = 9$, $18 = 8 + 2 + 8$ links and $9 = 8 + 1$ plaquettes; for $d = 4$, $V = 256$, $1024 = 255 + 4 + 765$ links and $1536 = 765 + 771$ plaquettes, giving a determined fraction of 50.2%.

The multilevel sampler does not choose its tree: the links each refinement sets to $\mathbf{1}$ form one, and their union over levels is again maximal (Appendix F.2). No level traverses it, each closing its constraint in plaquette space. Should links be wanted — for an observable that winds around the lattice, and so is not a product of plaquettes — the pairing above recovers them for any group in any dimension, by running it forwards: choose a maximal tree, draw the holonomies, and take each remaining link from the plaquette it closes.

B.2 The same count without a gauge choice

The count also follows without a gauge choice, from the cochain complex of the periodic lattice (Batrouni, 1982; Batrouni & Halpern, 1984). Let C^k be the k-cochains — sites, links and plaquettes for $k = 0, 1, 2$ — and $\mathrm{d} : C^k \to C^{k+1}$ the coboundary, $\mathrm{d}^2 = 0$. With V sites in d dimensions,

$$n_k \equiv \dim C^k = \tbinom{d}{k} V, \qquad b_k \equiv \dim H^k = \tbinom{d}{k}, \tag{11}$$

the second because T^d carries $\binom{d}{k}$ independent k-cycles. Write $Z^k = \ker \mathrm{d}$ and $B^k = \operatorname{im} \mathrm{d}$. The lattice is connected, so the constant 0-cochains are the one-dimensional kernel of $\mathrm{d} : C^0 \to C^1$, giving $\dim B^1 = V - 1$ and, with $b_1 = d$, $\dim Z^1 = (V - 1) + d$. Rank–nullity for $\mathrm{d} : C^1 \to C^2$ then gives

$$\dim B^2 = dV - \dim Z^1 = (d - 1)(V - 1), \tag{12}$$

the number of independent continuous degrees of freedom in a link-generated plaquette field.

Local and global constraints enter separately. With $b_2 = \binom{d}{2}$, so $\dim Z^2 = (d - 1)(V - 1) + \binom{d}{2}$, the local Bianchi identities restrict an arbitrary plaquette field in C^2 to the closed subspace Z^2 and therefore number

$$n_2 - \dim Z^2 = \tbinom{d}{2} V - (d - 1)(V - 1) - \tbinom{d}{2} = \Big[\tbinom{d}{2} - (d - 1)\Big](V - 1). \tag{13}$$

The remaining $\binom{d}{2}$ directions of $Z^2/B^2 \simeq H^2(T^d)$ are the fluxes through the non-contractible two-tori, and the global closure relations select the link-generated fields B^2. For compact $U(1)$ the integer flux sectors are distinct branches of those conditions; being discrete labels, they leave (12) unchanged. Hence $|F| = \dim B^2 = (d - 1)(V - 1)$, $|D| = \binom{d}{2} V - (d - 1)(V - 1)$, and

$$f_d = \frac{|D|}{\binom{d}{2} V} = 1 - \frac{(d - 1)(V - 1)}{\binom{d}{2} V} \xrightarrow[V \to \infty]{} 1 - \frac{2}{d}, \tag{14}$$

so $f_2 = 1/V \to 0$ and $f_4 = \frac{1}{2} + \frac{1}{2V} \to \frac{1}{2}$, as in Appendix E.1.

On the link side, modding out the $V - 1$ non-trivial gauge transformations leaves $dV - (V - 1) = \dim B^2 + b_1$ continuous degrees of freedom: $\dim B^2$ carried by the plaquettes and $d = b_1$ by the holonomies, which for $U(1)$ are flat directions of the Wilson action. Plaquettes and holonomies therefore suffice to reconstruct a gauge-fixed link field, as Appendix B.1 does explicitly.

The argument is linear, so it applies to $U(1)$ directly and otherwise componentwise in the Lie algebra: for non-Abelian G every continuous count gains a factor $\dim G$, while the Bianchi and closure relations become non-linear and involve transport and the holonomies without changing the local dimension count. On the finest four-dimensional lattice used here, $L = 4$,

$$n_2 = 1536, \qquad \dim B^2 = 3(V - 1) = 3 \cdot 255 = 765, \tag{15}$$

the number of free plaquette variables in every construction in this work.

C Parallel transport and local gauge frames

For a non-Abelian gauge theory, matrices based at different lattice sites are expressed in different local gauge frames. Associate to each site x a colour space $V_x \simeq \mathbb{C}^N$. A local gauge transformation

acts as an independent change of basis, $\psi(x) \mapsto g(x)\psi(x)$, and a plaquette or Wilson loop based at x transforms as

$$W(x) \mapsto g(x)W(x)g(x)^\dagger. \tag{16}$$

A matrix $W(y)$ based at $y = x+\hat{\mu}$ transforms instead with $g(y)$. Hence the direct product of matrices at different sites is not gauge covariant,

$$W(x)W(y) \mapsto g(x)W(x)g(x)^\dagger g(y)W(y)g(y)^\dagger, \tag{17}$$

because $g(x)^\dagger g(y) \neq \mathbf{1}$ in general.

The link $U_\mu(x)$ identifies the neighbouring frames, transforming as (3), so a matrix based at $x+\hat{\mu}$ can be transported to the frame at x as

$$W^{(x)}(x+\hat{\mu}) \equiv U_\mu(x)W(x+\hat{\mu})U_\mu(x)^\dagger. \tag{18}$$

Under a gauge transformation,

$$W^{(x)}(x+\hat{\mu}) \mapsto g(x)W^{(x)}(x+\hat{\mu})g(x)^\dagger, \tag{19}$$

because the transformations at $x+\hat{\mu}$ cancel. Thus $W(x)W^{(x)}(x+\hat{\mu})$ is a well-defined covariant product in the common frame at x.

Therefore, before multiplying plaquettes with different base points—in particular in a non-Abelian Bianchi identity—all factors must first be parallel transported to a common base point. For $U(1)$, conjugation is trivial, $UWU^\dagger = W$, so this transport is invisible; for $SU(2)$ and $SU(3)$ it must be treated explicitly.

D Lattice Bianchi identity

Consider an elementary cube based at x and spanning the directions μ, ν and ρ. Its faces are the plaquettes (1). For a non-Abelian group, plaquettes based at different vertices cannot be multiplied directly, so the three plaquettes on the far faces are transported back to the common base point x,

$$\begin{aligned}
\widetilde{P}^{(\mu)}_{\nu\rho}(x) &= U_\mu(x)\, P_{\nu\rho}(x+\hat{\mu})\, U_\mu(x)^\dagger, \\
\widetilde{P}^{(\nu)}_{\mu\rho}(x) &= U_\nu(x)\, P_{\mu\rho}(x+\hat{\nu})\, U_\nu(x)^\dagger, \\
\widetilde{P}^{(\rho)}_{\mu\nu}(x) &= U_\rho(x)\, P_{\mu\nu}(x+\hat{\rho})\, U_\rho(x)^\dagger.
\end{aligned} \tag{20}$$

With a consistent orientation of the six faces, the ordered product around the boundary of the cube is

$$\begin{aligned}
\mathcal{B}_{\mu\nu\rho}(x) = P_{\mu\rho}(x)\, \widetilde{P}^{(\rho)}_{\mu\nu}(x)\, P_{\nu\rho}(x)^\dagger \\
\times \left[\widetilde{P}^{(\nu)}_{\mu\rho}(x)\right]^\dagger\, P_{\mu\nu}(x)^\dagger\, \widetilde{P}^{(\mu)}_{\nu\rho}(x).
\end{aligned} \tag{21}$$

The cancellation is visible in the first two factors already,

$$\begin{aligned}
P_{\mu\rho}(x)\, \widetilde{P}^{(\rho)}_{\mu\nu}(x) = U_\mu(x)\, U_\rho(x+\hat{\mu})\, U_\mu(x+\hat{\rho})^\dagger\, U_\rho(x)^\dagger \\
\times U_\rho(x)\, U_\mu(x+\hat{\rho})\, P_{\mu\nu}(x+\hat{\rho})\, U_\mu(x+\hat{\rho})^\dagger\, U_\rho(x)^\dagger,
\end{aligned} \tag{22}$$

where $U_\rho(x)^\dagger U_\rho(x)$ and $U_\mu(x+\hat{\rho})^\dagger U_\mu(x+\hat{\rho})$ cancel. The remaining faces are arranged in the same way by the transports and the surface ordering, so that repeated use of $U^\dagger U = UU^\dagger = \mathbf{1}$ removes every link and gives

$$\mathcal{B}_{\mu\nu\rho}(x) = \mathbf{1}. \tag{23}$$

No commutativity is used: transport and surface ordering arrange the factors so that each link cancels against its inverse, which is why the identity holds for $SU(2)$ and $SU(3)$ exactly as for $U(1)$.

D.1 Which closed surfaces exist

The identity (4) holds for any closed surface Σ assembled from plaquettes: transporting its faces to a common base point and multiplying them in surface order makes every link of Σ occur once with

each orientation, and the product collapses to the identity. Which surfaces a periodic lattice carries therefore decides the form of the constraint, and there are two kinds.

For $d \geq 3$ the boundary of every elementary cube c is a contractible closed surface, and each one gives a *local* relation: (4) at $\Sigma = \partial c$,

$$\overrightarrow{\prod}_{p \subset \partial c} T_{p \to x_c}\big[P_p^{\epsilon_{cp}}\big] = \mathbf{1}, \tag{24}$$

with x_c any corner of the cube as the base point x_0 and $\epsilon_{cp} \equiv \epsilon_{\partial c\, p}$ its face orientations. It is (23) written for a general cube.

In every dimension each coordinate plane $\mu\nu$ in addition carries the two-torus Λ of all its plaquettes, a closed but *non-contractible* surface, and there the cancellation is not complete. Cutting the torus along its two cycles opens it into a square whose boundary is traversed as $a\, b\, a^{-1} b^{-1}$, with a and b the cycles in the μ and ν directions. The plaquettes of Λ tile that square, so transporting them to x_0 and multiplying in surface order cancels every interior link as before, while the four sides are left over and contribute the holonomies wrapping the plane in that order,

$$\overrightarrow{\prod}_{p \in \Lambda} T_{p \to x_0}[P_p] = \Omega_\mu\, \Omega_\nu\, \Omega_\mu^\dagger\, \Omega_\nu^\dagger, \tag{25}$$

one *global* relation per plane, $\binom{d}{2}$ in all. This is the closure stated as (5) in the main text; it collapses to the identity only when the group is Abelian, so that the two holonomies commute.

D.2 Explicit forms for the three instantiations

$U(1)$ **in two dimensions.** The links are the phases $U_\mu(x) = e^{i\theta_\mu(x)}$, so a plaquette is itself a phase, $P_{\mu\nu}(x) = e^{i\varphi_{\mu\nu}(x)}$, whose angle is the oriented sum of its four link angles,

$$\varphi_{\mu\nu}(x) = \theta_\mu(x) + \theta_\nu(x + \hat{\mu}) - \theta_\mu(x + \hat{\nu}) - \theta_\nu(x). \tag{26}$$

There are no cubes, so (5) is the only relation; the group is Abelian, the holonomies commute and the right-hand side is unity. Because the angles are defined modulo 2π, what remains is a single integer constraint on the whole lattice,

$$\prod_{p \in \Lambda} P_p = 1, \qquad \sum_{p \in \Lambda} \varphi_p = 2\pi Q, \qquad Q \in \mathbb{Z}, \tag{27}$$

with the angles on their principal branch; the integer Q is the topological charge.

$U(1)$ **in four dimensions.** The parametrisation is unchanged, but every elementary cube now contributes (24), which in the same way becomes an integer relation,

$$\sum_{p \subset \partial c} \epsilon_{cp}\, \varphi_p = 2\pi m_c, \qquad m_c \in \mathbb{Z}, \tag{28}$$

where m_c is a monopole current, the four-dimensional counterpart of a magnetic charge. There are in addition $\binom{4}{2} = 6$ global relations of the form (27), one per coordinate plane. The difference from two dimensions is one of kind: a single global condition there, an *extensive* set of local ones here, one per cube. The integers m_c and Q are not additional variables; they are read off from the angles, and they are what makes the constraint discrete rather than merely geometric.

$SU(2)$ **in two dimensions.** A plaquette is a group element, written in the same form as the links, $P_{\mu\nu} = \cos\omega_{\mu\nu} + i \sin\omega_{\mu\nu}\, \boldsymbol{n} \cdot \boldsymbol{\sigma}$, with $\boldsymbol{n}$ a unit vector in $\mathbb{R}^3$ and the angle $\omega \in [0, \pi]$ fixed by $\cos\omega = \frac{1}{2} \operatorname{tr} P$. As in two-dimensional $U(1)$ there are no cubes and (5) is the only relation, but the group is non-Abelian: the parallel transports and the surface ordering cannot be dropped, and the commutator on the right-hand side does not reduce to unity, so the two holonomies enter the constraint itself rather than cancelling out of it.

E Single-level plaquette sampling

E.1 Plaquettes as sampling variables

The target distribution (2) is constrained by the Bianchi identities and therefore lies on a lower-dimensional manifold $\mathcal{M}$ of the full plaquette space. Since $\mu(\mathcal{M}) = 0$ in the ambient measure, a

full-dimensional flow with density q_θ gives $\Pr_{q_\theta}(P \in \mathcal{M}) = \int \mathbf{1}_{\mathcal{M}}\, q_\theta \, d\mu = 0$. Penalty terms only suppress violations, while projection onto $\mathcal{M}$ generally destroys the tractable density needed for likelihood evaluation and importance weighting. We therefore enforce the constraints by construction, splitting the plaquettes into a free set F and a determined set D. Removing from the $\binom{d}{2}V$ plaquettes the $\big[\binom{d}{2} - (d-1)\big](V-1)$ independent cube identities (24) and the $\binom{d}{2}$ global closures (5), one per coordinate plane (Appendix B), leaves

$$|F| = (d-1)(V-1), \qquad f_d = \frac{|D|}{\binom{d}{2} V} = 1 - \frac{(d-1)(V-1)}{\binom{d}{2} V} \xrightarrow[V\to\infty]{} 1 - \frac{2}{d}, \tag{29}$$

so $f_2 \to 0$ while $f_4 \to \frac{1}{2}$. The free plaquettes together with the d holonomies form coordinates for gauge-fixed link fields (Appendix B, Fig. 6).

This count fixes how many plaquettes are determined, not how non-local the solve is. For $j \in D$, let k_j denote the number of free plaquettes on which it depends. Independent errors of size ϵ then propagate as $O(\sqrt{k_j}\,\epsilon)$ in j, so Eq. (29) alone gives no information about the solve support k_j.

E.2 The exact constraint layer

A flow generates the free plaquettes and a deterministic layer computes the determined ones, provided the plaquettes are ordered so that every determined one is fixed by a relation among plaquettes placed before it. The relations are the cube identities (24) in $d \geq 3$ and the single global closure (5) in $d = 2$. The layer holds to machine precision on every sample, trained or not, and everything it needs is computed once, offline: the orderings are properties of the lattice, not of the model or of β. The multilevel construction reuses the same solve level by level, with the blocking relations in place of the flux rows (Appendix F.3).

E.2.1 $U(1)$: an integer linear solve

The system. Let $\varphi \in \mathbb{R}^{n_2}$, $n_2 = \binom{d}{2}V$, be the vector of all plaquette angles, unwrapped (in $\mathbb{R}$ rather than $(-\pi, \pi]$). The relations satisfied by the plaquettes of a link field are linear in φ with integer coefficients, of two kinds. *Bianchi rows*, one per elementary cube c, the oriented sum of its six faces $\sum_{p\subset\partial c} \epsilon_{cp}\varphi_p = 0$, which is (28) with $m_c = 0$: $\binom{d}{3}V$ rows with entries $0, \pm 1$, of which $\big[\binom{d}{2} - (d-1)\big](V-1)$ are independent (Appendix B.2). *Flux rows*, one per plane $\mu\nu$,

$$\sum_{a,b} \varphi_{\mu\nu}\big|_{x_\mu = a,\, x_\nu = b,\, x_\perp = 0} = 2\pi n_{\mu\nu}, \tag{30}$$

the sum over the plaquettes of one two-torus, the unwrapped form of (27). Stacked,

$$\boldsymbol{A}\varphi = b, \qquad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{\mathrm{B}} \\ \boldsymbol{A}_{\mathrm{W}} \end{pmatrix}, \qquad b = \begin{pmatrix} 0 \\ 2\pi n \end{pmatrix}, \tag{31}$$

with $\boldsymbol{A}_{\mathrm{B}}$ the Bianchi rows and $\boldsymbol{A}_{\mathrm{W}}$ the $\binom{d}{2}$ flux rows. In two dimensions there are no cubes, $\boldsymbol{A}_{\mathrm{B}}$ is empty, and the system is the single flux row $\sum_p \varphi_p = 2\pi n$.

The solve. Split the columns into free and determined plaquettes, $F \sqcup D$, with $|D| = \operatorname{rank} \boldsymbol{A}$ chosen so that the square block $\boldsymbol{A}_D$ is invertible. Then $\boldsymbol{A}_F\varphi_F + \boldsymbol{A}_D\varphi_D = b$ gives the layer, an affine map on the unwrapped angles,

$$\varphi_D = -\boldsymbol{\Gamma}\,\varphi_F + \boldsymbol{A}_D^{-1} b, \qquad \boldsymbol{\Gamma} = \boldsymbol{A}_D^{-1}\boldsymbol{A}_F \in \{-1, 0, 1\}^{|D|\times|F|}, \tag{32}$$

in which $\boldsymbol{\Gamma}$ is the only object the model interacts with. In two dimensions D is one plaquette and $\boldsymbol{\Gamma}$ a row of ones, $\varphi_D = 2\pi n - \sum_F \varphi_F$; in four dimensions it fixes half the plaquettes from the other half. Setting $n = 0$ loses nothing: the unwrapped plane sum of $\mathrm{d}\theta$ telescopes to zero for every link field, so every configuration lies in that parametrisation, and a non-zero physical flux enters only through the integer lifts below, costing an action of order $2\pi^2\beta$ where it could matter (Appendix K). Without the flux rows the free space would contain the $\binom{d}{2}$ harmonic directions, which are the plaquette field of no link configuration; with them its dimension is $(d-1)(V-1)$ (Appendix B), or 765 at $L = 4$ in four dimensions.

Proposition 1 (Constant Jacobian). *For every φ_F the point $\varphi = (\varphi_F, \varphi_D)$ satisfies* (31) *exactly, and its density on the constraint set with respect to $|F|$-dimensional Hausdorff measure is*

$$\log q_{\mathcal{M}}(\varphi) = \log q_F(\varphi_F) - \tfrac{1}{2} \log\det\big(\boldsymbol{I}_{|F|} + \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma}\big), \tag{33}$$

whose second term is independent of φ_F and b.

Proof. The map $\varphi_F \mapsto \varphi$ is the affine embedding $\varphi = \varphi_0(b) + \boldsymbol{B}\varphi_F$ with $\boldsymbol{B} = (\boldsymbol{I}_{|F|}, -\boldsymbol{\Gamma}^\top)^\top$, whose volume element $\sqrt{\det \boldsymbol{B}^\top \boldsymbol{B}} = \sqrt{\det(\boldsymbol{I} + \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma})}$ is a constant. □

The constant cancels in self-normalised importance weights and shifts the reverse-KL objective by a constant, so it is never evaluated.

Lemma 1 (Integrality). *φ_D is well defined on $\mathbb{T}^{|D|}$ if and only if $\boldsymbol{\Gamma}$ is integer, which holds whenever $\boldsymbol{A}_D$ is unimodular, i.e. every elimination pivot is ± 1.*

Proof. Under $\varphi_k \mapsto \varphi_k + 2\pi$ the solve shifts φ_j by $-2\pi\boldsymbol{\Gamma}_{jk}$, a multiple of 2π if and only if $\boldsymbol{\Gamma}_{jk} \in \mathbb{Z}$. All inputs are integer, so a non-integer can arise only by division, and with every pivot of modulus one no division occurs. (Unimodularity is sufficient, not necessary: $\boldsymbol{A}_D = \boldsymbol{A}_F = (2)$ gives $\boldsymbol{\Gamma} = (1)$.) □

Wrapped angles, monopoles and charge. The solve produces unwrapped angles. The physical plaquette is $\bar{\varphi}_j = \mathrm{wrap}(\varphi_j) \in (-\pi, \pi]$, and the integer $k_j = (\varphi_j - \bar{\varphi}_j)/2\pi$ is the *lift* of the determined plaquette. Inserting the wrapped angles into the relations returns (28) and (27) with $m_c = -\sum_{p \subset \partial c} \epsilon_{cp} k_p$ and $Q = -\sum_p k_p$: the monopole numbers in $d \geq 3$ and the topological charge in $d = 2$ are sums of lifts, read off the sample rather than generated. The flow never sees an integer variable; the tails of its continuous density decide how often a determined plaquette leaves $(-\pi, \pi]$, which is how monopoles arise in the sampler (Appendix M).

Elimination and checks. The partition is computed in exact integer arithmetic, taking at each step a pivot from the row of minimum remaining support (Appendix G); the coarse 2^4 system of the chain is reduced with a fixed column priority. The elimination counts non-unit pivots rather than assuming there are none, and in every system built for this work, at every level of every chain, the count is zero and all coefficients of $\boldsymbol{\Gamma}$ lie in $\{-1, 0, +1\}$. After elimination the rows of $\boldsymbol{\Gamma}$ are sparse (Table 1: $O(1)$ non-zeros on a refinement level, a median support growing linearly with L for the single-level solve), so applying (32) is a sparse integer matrix–vector product. Whenever a system is built or a batch generated, three conditions are asserted: every pivot is ± 1, the free-variable count equals $(d-1)(V-1)$, and the configurations satisfy every relation to machine precision.

E.2.2 $SU(2)$: LINKS IN TREE ORDER

For a non-Abelian group the relations are ordered products of plaquettes based at different sites, so imposing them directly would mean deriving and solving an ordered identity for every relation. The flow still generates plaquettes. The layer takes them, together with the sampled d holonomies, and constructs the link field from which they come; the plaquettes of a link field satisfy every relation automatically, so none need be imposed. No transporter appears in the construction, which is what the tree buys: every site is reached from the root along identity links, so plaquettes based at different sites may be multiplied at no cost. The lattice is $L \times L$, $V = L^2$, and the pairing is that of Appendix B.1 and Fig. 6: in the comb gauge every link outside the tree is the last unknown factor of exactly one plaquette, the one it closes.

Generation. The flow produces $V - 1$ elements $P_p \in SU(2)$, one for every plaquette except the corner plaquette p_* at $x = (L-1, L-1)$, which the global relation closes (Fig. 6). The two holonomies Ω_0, Ω_1 are drawn from Haar measure.

Reconstruction. With the tree links equal to $\mathbf{1}$ and the holonomies in place, the links follow in the sweep order of Appendix B: each plaquette of row $j < L-1$ fixes the horizontal link above it, and each plaquette of the top row fixes the top vertical link to its right. In both cases (1) has one unknown factor and one solution; for the first, with $P_{01}(x) = U_0(x)\,U_1(x+\hat{0})\,U_0(x+\hat{1})^\dagger\,U_1(x)^\dagger$,

$$U_0(x+\hat{1}) = U_1(x)^\dagger\,P_{01}(x)^\dagger\,U_0(x)\,U_1(x+\hat{0}), \tag{34}$$

and analogously for the top row. One pass turns the $V-1$ generated plaquettes into the $V-1$ free links, with no Jacobian: at every step the unknown link is the generated plaquette multiplied left and right by elements already fixed, and Haar measure is invariant under both, so the density of the link field relative to Haar equals the flow density on the P_p times the Haar density of the holonomies.

The determined plaquette. After the pass every link is known, so P_{p_*} is read off them rather than solved for, which is (5) holding automatically:

$$T_{p_*\to x_0}\big[P_{p_*}\big] = \Big(\overrightarrow{\prod_{p\neq p_*}} T_{p\to x_0}[P_p]\Big)^\dagger\,\Omega_0\,\Omega_1\,\Omega_0^\dagger\,\Omega_1^\dagger, \tag{35}$$

the commutator of the holonomies divided by the transported product of all other plaquettes. It is the single determined plaquette of (29) in two dimensions, with solve support $V-1$, the non-Abelian counterpart of $\varphi_D = -\sum_F \varphi_F$. The action (2) is evaluated on all V plaquettes, including P_{p_*}, and the importance weight is $w = e^{-S}/q$.

Higher dimensions. The same reconstruction applies in any d, with a nested comb in place of the two-dimensional one: every link outside the tree is fixed by the plaquette it closes, and the plaquettes closing each cube are then determined by (24). The route is worth more there than in two dimensions, the cube identities being $O(V)$ relations rather than one. We have not trained a non-Abelian single-level model in $d \geq 3$: there every level is a refinement of Appendix F and no reconstruction is needed.

E.3 Why one level is hard to learn

Table 1: Solve support and conditioning of the $U(1)$ constraint layer. Single level: all plaquettes of the lattice generated at once (Appendix E.2). Refinement level: one doubling of the multilevel construction (Appendix F), whose values do not depend on the lattice extent (Appendix G).

	single level		refinement level	
lattice	support (median / max)	$\kappa(\boldsymbol{I}+\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma})$	support (max)	κ
2^3	5 / 7	20	4	5
8^3	17 / 159	877	4	5
2^4	5 / 11	45	4	7
4^4	13 / 47	391	4	7

The layer is exact, but it changes the problem the flow has to solve. Solved on the whole lattice at once, the constraint is global in the generated variables: a determined plaquette is fixed by a chain of cube relations that ends on free plaquettes, and on a periodic lattice such chains cross the lattice. Table 1 shows the consequence for $U(1)$, the solve support of a single-level layer growing with the lattice and with it the conditioning of the target (Appendix G). At weak coupling $S \simeq \frac{\beta}{2}\|\varphi\|^2$, so substituting (32),

$$S \;\simeq\; \tfrac{\beta}{2}\,\varphi_F^\top\big(\boldsymbol{I}+\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\big)\varphi_F + \text{linear}, \tag{36}$$

a Gaussian that couples every pair of free angles sharing a determined plaquette, with an anisotropy given by $\kappa(\boldsymbol{I}+\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma})$. A coupling flow transforms a few variables per layer and must build these long-range, badly scaled correlations implicitly; in two dimensions the single determined plaquette depends on all $V-1$ others. A constraint that is local in the physical variables has become global in the variables the model generates.

F THE REFINEMENT AND THE CHAIN

Notation follows Section 3: $\mathcal{V}^{(i)}$ denotes the set of valid plaquette fields at level i, $F^{(i)}$ the $N^{(i)}$ free variables generated at that level, R_i the refinement map of (6), and $\Omega \in G^d$ the holonomies. Coarse quantities carry a superscript c: P^{c} for a coarse plaquette and U^{c} for a coarse link. All statements below were verified numerically for $U(1)$ and $SU(2)$ in $d = 2, \dots, 5$ (Appendix H).

F.1 BLOCKING, REFINEMENT AND EXACTNESS

The coordinates. A blocking map $\mathcal{K}_i : \mathcal{V}^{(i)} \to \mathcal{V}^{(i-1)}$, together with the free variables $F^{(i)} \in G^{N^{(i)}}$, must define

$$\big(\mathcal{K}_i, F^{(i)}\big) : \ \mathcal{V}^{(i)} \longrightarrow \mathcal{V}^{(i-1)} \times G^{N^{(i)}} \qquad \text{bijective, with unit Haar Jacobian,} \tag{37}$$

with inverse R_i: a valid coarse field and arbitrary $F^{(i)}$ determine exactly one valid fine field. We refine one direction μ at a time, in the gauge $U_\mu(2X) = \mathbf{1}$, so that the two fine plaquettes produced by a split are based at the same point $2X$. Then

$$\mathcal{K}_i : \qquad P^{\mathrm{c}}_{\mu\nu}(X) = P_{\mu\nu}(2X + \hat{\mu})\, P_{\mu\nu}(2X), \qquad P^{\mathrm{c}}_{\nu\rho}(X) = P_{\nu\rho}(2X), \qquad (\nu, \rho \neq \mu), \tag{38}$$

and

$$R_i : \begin{aligned} &P_{\mu\nu}(2X) = F_\nu, \quad P_{\mu\nu}(2X + \hat{\mu}) = P^{\mathrm{c}}_{\mu\nu} F^\dagger_\nu, \\ &P_{\nu\rho}(2X) = P^{\mathrm{c}}_{\nu\rho}, \quad P_{\nu\rho}(2X + \hat{\mu}) = F_\nu\, T_\nu\big[F_\rho(X + \hat{\nu})\big]\, P^{\mathrm{c}}_{\nu\rho}\, T_\rho\big[F_\nu(X + \hat{\rho})\big]^\dagger F^\dagger_\rho, \end{aligned} \tag{39}$$

where T_ν, T_ρ denote transport by coarse links and are therefore functions of P^{c} and Ω.

The four classes and their supports. The four lines of (39) exhaust the possible fine plaquettes: generated, fixed by its blocking partner, copied, or determined by one cube identity. Their solve supports are respectively $1, 1, 0, 4$, independent of G, d, and the lattice extent. The first three follow directly from (43); the fourth is the cube identity (24) at $2X$ in the directions μ, ν, ρ. Its two faces transverse to μ are $P_{\nu\rho}(2X)$ and $P_{\nu\rho}(2X + \hat{\mu})$, while the four faces containing μ are generated elements at μ-even sites. There are $d - 1$ plaquettes of each class per coarse site, hence

$$N^{(i)} = (d - 1)V^{(i-1)}, \tag{40}$$

which is the level count used in Section 3. In $d = 2$ there is no third direction and hence no cube class; only the first line of (39) remains. For $U(1)$ the transporters drop out and (38) becomes linear in the angles,

$$\Phi_{\mu\nu}(X) = \varphi_{\mu\nu}(2X) + \varphi_{\mu\nu}(2X + \hat{\mu}), \qquad \Phi_{\nu\rho}(X) = \varphi_{\nu\rho}(2X), \tag{41}$$

with Φ coarse and φ fine. In these coordinates, the condition number of $\boldsymbol{I} + \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma}$ is $1, 5, 7, 9$ for $d = 2, 3, 4, 5$, respectively, at every extent tested (Appendix G, Table 5).

Exactness. We prove Eq. (37) at the link level. Decimation along μ and its inverse are

$$U^{\mathrm{c}}_\mu(X) = U_\mu(2X)\, U_\mu(2X + \hat{\mu}), \qquad U^{\mathrm{c}}_\nu(X) = U_\nu(2X) \quad (\nu \neq \mu), \tag{42}$$

and

$$U_\mu(2X) = \mathbf{1}, \quad U_\mu(2X + \hat{\mu}) = U^{\mathrm{c}}_\mu(X), \quad U_\nu(2X) = U^{\mathrm{c}}_\nu(X), \quad U_\nu(2X + \hat{\mu}) = F_\nu(X)\, U^{\mathrm{c}}_\nu(X), \tag{43}$$

respectively. With $U_\mu(2X) = \mathbf{1}$, (42) reproduces the blocked plaquettes in (38). The bijection and unit Jacobian then follow in three steps.

(i) Substituting (43) into (42) recovers U^{c}, while $P_{\mu\nu}(2X) = F_\nu(X)$ because the first and third factors of that plaquette are identities.

(ii) Any U' that decimates to U^{c} can be brought to the form (43) by the gauge transformation (3) with $g(2X) = \mathbf{1}$ and $g(2X + \hat{\mu}) = U'_\mu(2X)$. This leaves every decimated product unchanged and is the unique transformation supported on the new sites with this property, since $g(2X + \hat{\mu})$ is fixed by $g(2X)U'_\mu(2X)g(2X + \hat{\mu})^\dagger = \mathbf{1}$. The gauge choice is therefore necessary rather than merely

convenient: without it, one pair (U^{c}, F) would correspond to an entire orbit of fine fields. For $U(1)$ the plaquettes are gauge invariant, so no such choice is required.

(iii) For fixed U^{c}, the only links that depend on F are $U_\nu(2X+\hat{\mu}) = F_\nu(X)U^{\mathrm{c}}_\nu(X)$, one for each generated element. Each is a right translation by an element independent of F. Since Haar measure is right invariant, product Haar measure on the $F_\nu(X)$ is mapped to Haar measure on the fine links with unit Jacobian (numerically $|\log|\det J|| \le 2\times 10^{-14}$, Table 5).

F.2 The chain

We start from the one-site torus, whose d links are the holonomies Ω, and refine successively along $\mu = 0, \dots, d-1$, repeating this cycle n times. The links fixed to $\mathbf{1}$ form a maximal tree at every level. This follows by induction: the one-site torus fixes none, while a refinement along μ fixes the $V^{(i-1)}$ links $U_\mu(2X) = \mathbf{1}$. Each previously fixed coarse link remains fixed after refinement, appearing either as $U_\nu(2X)$ or as $U_\mu(2X+\hat{\mu})$, so the total becomes $(V^{(i-1)}-1) + V^{(i-1)} = V^{(i)} - 1$. These links contain no loop, since any such loop would project to a loop in the coarse tree. Summing (40) over the volumes $1, 2, 4, \dots, V/2$ gives

$$d + \sum_i N^{(i)} = d + (d-1)\big(1 + 2 + \cdots + \tfrac{V}{2}\big) = d + (d-1)(V-1), \tag{44}$$

which is precisely the count in Appendix E.1: the holonomies together with the generated elements give all degrees of freedom of a link field in tree gauge. Applying step (iii) at each level, the full map to non-tree links has unit Haar Jacobian, so the density factorises as (7) and the importance weight is e^{-S}/q.

F.3 Plaquette form for $U(1)$

For $U(1)$ the refinement can be written directly in terms of plaquette angles. Equation (41) is the plaquette-space image of (42): after substitution into (1), the interior link of a face containing μ appears twice with opposite signs, so the coarse face is the sum of the two fine faces it covers, while a face not containing μ is simply the fine face at even x_μ. Since the blocking map is a sum followed by restriction, and both commute with the coboundary, it intertwines the fine and coarse coboundaries,

$$\mathrm{d_c}\, \boldsymbol{A}_{\mathrm{K}} = \boldsymbol{A}'_{\mathrm{K}}\, \mathrm{d_f} \quad \Longrightarrow \quad \mathrm{d_f}\varphi = 0 \;\Rightarrow\; \mathrm{d_c}\Phi = 0, \tag{45}$$

so the coarse Bianchi identity is inherited rather than imposed again. The level system is

$$\boldsymbol{A}\varphi = b, \qquad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{\mathrm{B}} \\ \boldsymbol{A}_{\mathrm{K}} \end{pmatrix}, \qquad b = \begin{pmatrix} 0 \\ \Phi \end{pmatrix}, \qquad (\boldsymbol{A}_{\mathrm{K}}\varphi)_P = \sum_{p\in P} \varphi_p, \tag{46}$$

where $\boldsymbol{A}_{\mathrm{B}}$ contains the fine Bianchi rows and $\boldsymbol{A}_{\mathrm{K}}$ one blocking row per coarse plaquette P. The coarse field enters only through the constant b; no flux rows are needed, because the fluxes are fixed at the coarsest level and inherited through (45). The elimination and the two properties of Appendix E.2 then apply unchanged, and Appendix G compares these coordinates with those of (39).

The four-dimensional chain combines the first four refinements, $1^4 \to 2^4$, into a single joint flow on the $45 = 3\cdot 16 - 3$ angles of the 2^4 lattice, with fluxes fixed to zero. It then refines one direction at a time, generating $48, 96, 192, 384$ angles, copying the same numbers at support zero, and solving $96, 192, 384, 768$. Each step is a unimodular integer change of variables from the $F_\nu(X)$ of (43), so its Jacobian is one (Table 5). Refining all directions at once is also the image of a link decimation and therefore also preserves the Bianchi identity, but within a coarse cell it leaves fine faces that are not fixed by any single cube relation. The constraints then chain across the cell and the support grows with the extent (Appendix G). Thus refining one direction at a time is not required for exactness, but for locality.

F.4 Two-dimensional refinements

In two dimensions both directions are refined at once, equivalent to two successive refinements along $\mu = 0$ and then $\mu = 1$. Each 2×2 block is generated from its coarse plaquette using three elements, while the fourth is fixed by

$$P^{\mathrm{c}}(X) = P_4 P_3 P_2 P_1, \tag{47}$$

or equivalently

$$P_3 = P_4^\dagger \, P^{\mathrm{c}} \, (P_2 P_1)^\dagger. \tag{48}$$

Hence $N^{(i)} = 3V^{(i-1)}$, which is (40) applied twice. For $U(1)$ this is simply (41) applied twice in angle variables, and the layer is the $d = 2$ case of Appendix E.2: one blocking row per block and no cube rows. The holonomies do not enter the action and are drawn uniformly; the flux is fixed to zero at the coarsest level, and the topological charge of a sample is the sum of the lifts of its determined plaquettes (Appendix E.2).

For $SU(2)$ only the ordering changes, while (47) remains unchanged. Refining first along direction 0 generates $F(X)$ at each coarse site and produces the intermediate plaquettes $F(X)$ and $P^{\mathrm{c}}(X)F(X)^\dagger$. Refining along direction 1 then generates one element at each intermediate site. Labelling the block top-left, bottom-left, top-right, bottom-right (Fig. 2),

$$\begin{aligned} P_1 &= G_L^\dagger F, & P_2 &= G_L, \\ P_3 &= G_R^\dagger \, P^{\mathrm{c}} F^\dagger, & P_4 &= G_R, \end{aligned} \qquad F, G_L, G_R \in SU(2) \text{ generated}, \tag{49}$$

where G_L, G_R are the second-refinement elements at the two intermediate sites. The three links $U_0(2X)$, $U_1(2X)$ and $U_1(2X + \hat{0})$ inside the block are identities, so all four fine plaquettes are based at $2X$ and no transporter appears. Generating P_1, P_2, P_4 and determining P_3 through (48) gives $P_3 = G_R^\dagger P^{\mathrm{c}} F^\dagger$ with $F = P_2 P_1$. The two parametrisations therefore differ only by translations and inversions and have the same unit Haar Jacobian.

The implementation contains one transporter as a gauge artefact. The two-dimensional $SU(2)$ code inherits the comb gauge of Appendix B.1 from the coarsest level, where the link $a = U_0(2X)$ at the lower-left corner of a block need not be the identity. The flow therefore works with the transported variables $Q_4 = aP_4a^\dagger$ and $Q_3 = aP_3a^\dagger$, for which (48) again holds. This causes no complication: a depends only on P_1 and P_2, which are always generated, so no determined plaquette enters a transporter and all blocks can be solved in parallel. Conjugation preserves both Haar measure and the trace, and the transformation between the two descriptions is the identity at every block corner, leaving the coarse links unchanged and allowing the levels to compose.

The holonomies are those of the coarsest level. They are drawn from Haar measure and remain unchanged under blocking, since coarse links are products of fine ones. They enter the fine plaquettes only through the coarsest plaquette, the commutator $\Omega_0\Omega_1\Omega_0^\dagger\Omega_1^\dagger$, which is supplied to the conditioner at every level (Appendix I). Their density is Haar rather than learned: the n-th character of the exact marginal is suppressed as $\exp\big[-(n^2-1)(L/\xi)^2/3\big]$, giving a deviation of 8.5×10^{-3} at $(L/\xi)^2 = 5$ and 2.3×10^{-9} at 20, for $(L/\xi)^2$ between 5.9 and 196 over the couplings studied.

G SOLVE SUPPORT AND CONDITIONING

The layer is exact for every admissible partition (F, D), but not every partition is equally learnable. For $j \in D$ the *solve support*

$$k_j = \big|\{\, i \in F : \mathbf{\Gamma}_{ji} \neq 0 \,\}\big| \tag{50}$$

is the row sparsity of $\mathbf{\Gamma}$ in (32), $k_j = 0$ meaning a plaquette copied from the coarse field. Errors of scale ϵ in the generated angles reach a determined angle as $O(\sqrt{k_j}\,\epsilon)$ if independent and $O(k_j\epsilon)$ if correlated, and the divergence collects a contribution from every plaquette, determined ones included. The count of Appendix B fixes $|F|$ and $|D|$ but is blind to k: two eliminations of one system can share $|F|$ and differ entirely in support. The second quantity is $\kappa(\boldsymbol{I} + \mathbf{\Gamma}^\top \mathbf{\Gamma})$, the matrix of (33) and (36). Table 1 is extracted from the two tables here.

The anisotropic chain and the isotropic step generate the same $48 + 96 + 192 + 384 = 720$ angles, so the comparison isolates locality from capacity. The isotropic and single-level solves are long-ranged, median support 13 and maximum 47; the anisotropic chain is local and becomes more so along the chain, median $3, 2, 1, 1$. Determinism is not the difference: the last anisotropic level determines 75% of the plaquettes, the isotropic step only 53%.

At matched parameter count the multilevel chain reaches half the divergence of the single-level model, 1.00 against 2.05, with ESS 4.65% against 1.52% (Fig. 7); 4^4 is the smallest lattice on which the comparison can be made.

Table 2: Support of the constraint solve and condition number κ of $\boldsymbol{I} + \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma}$ for the systems used in the four-dimensional experiments (integer elimination, minimum-support pivoting). Support 0 means the plaquette is copied from the coarse field. These are the coordinates the trained models use. In the refinement coordinates of Appendix F, every determined plaquette depends on at most four generated variables and $\kappa = 7$ (Table 3); the two choices are related by a unimodular integer change of variables, and both are independent of the lattice size.

system (fine shape, direction)	plaquettes	generated	determined	median	mean	max	κ
coarse level 2^4	96	45	51	5	6.0	11	45
$2^4 \to 4\times 2^3$, $\mu = 0$	192	48	144	3	2.8	8	45
$\to 4^2\times 2^2$, $\mu = 1$	384	96	288	2	2.7	9	43
$\to 4^3\times 2$, $\mu = 2$	768	192	576	1	2.3	8	30
$\to 4^4$, $\mu = 3$	1536	384	1152	1	1.7	4	7
isotropic step $2^4 \to 4^4$	1536	720	816	13	13.7	47	—
single level 4^4 (Fig. 7)	1536	765	771	13	13.6	47	391

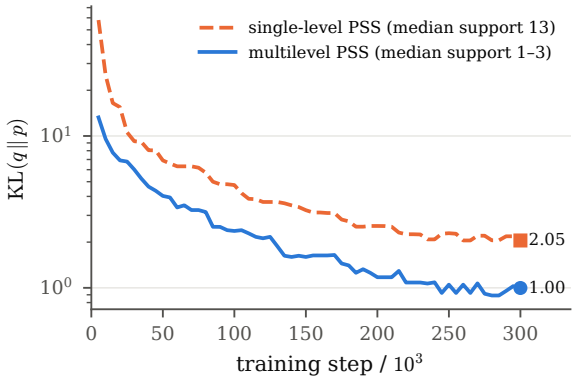


Figure 7: Four-dimensional $U(1)$ at $\beta = 2$: single-level against multilevel plaquette space at matched parameter count (3.91M against 4.06M), during training. Markers are the final evaluations of 500,000 samples, KL $= 2.05$ and 1.00. The two models also differ in width, depth, learning rate and seed, and the single-level model has the wider and deeper conditioner (Appendix I).

Table 3: Single-level solve of the whole lattice against one refinement level of the coordinates of (39), for $U(1)$. The refinement value was measured on extents 4 and 8 and is the same for both.

	single level					refinement level
d	lattice	generated	median support	max support	$\kappa(\boldsymbol{I} + \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma})$	κ
2	4^2, 8^2, 16^2	15, 63, 255	15, 63, 255	15, 63, 255	16, 64, 256	1
3	2^3, 4^3, 8^3	14, 126, 1022	5, 9, 17	7, 31, 159	20, 93, 877	5
4	2^4, 4^4	45, 765	5, 9	11, 47	45, 379	7
5	—	—	—	—	—	9

In Table 3 the single-level median support grows linearly with the extent and κ roughly with the volume, so such a model must represent a Gaussian whose widths span $\sqrt{\kappa} \approx 20$ on 4^4 and keep growing; on a refinement level κ is $1, 5, 7, 9$ in $d = 2, \ldots, 5$ at every extent. The single-level model of Fig. 7 uses a column priority that gives median support 13 and $\kappa = 391$ on 4^4 (Table 2), slightly above the minimum-support values of Table 3. The generated elements of (39) are more local than the eliminated coordinates our models use — median support 1, maximum 4 and $\kappa = 7$ on every level, against $45, 43, 30, 7$ — and the two are related by a unimodular integer matrix (Appendix H), so the first three refinements could be better conditioned by a change of coordinates alone.

G.1 SIZE AND CONDITIONING OF THE SINGLE-LEVEL SOLVE WITH EXTENT

The single-level layer solves (31) for the whole lattice at once. Table 4 gives the size of that system for the theories of this paper and for the four-dimensional non-Abelian theories one would want next, and Fig. 8 the condition number of (36) against it. For $SU(N)$ the constraint Jacobian at the identity is the $U(1)$ one tensored with $\boldsymbol{I}_{\dim G}$, so every count is $\dim G$ times the $U(1)$ count and κ is the same number.

Size. The solve has $\dim G\,|D|$ rows with $|D| = 1$ in $d = 2$ and $|D| = 3V + 3$ in $d = 4$: 3.1×10^6 for $U(1)$ and 2.5×10^7 for $SU(3)$ on 32^4. The elimination is done once, but the layer needs its result $\boldsymbol{\Gamma}$ explicitly, $|D| \times |F|$ with a row support that grows with the extent (Table 3, median $5, 9, \ldots$ at $L = 2, 4, \ldots$), so both the elimination and what it has to store grow faster than the lattice.

Conditioning. In $d = 2$, $\boldsymbol{\Gamma}$ is a single row of ones, $\boldsymbol{I} + \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \boldsymbol{I} + \boldsymbol{1}\boldsymbol{1}^\top$ has eigenvalues V and 1, and $\kappa = V$ exactly for every group and extent; the measured values in Table 4 are this. In $d = 4$ the measured extents give $\kappa/V = 1.5,\ 1.6,\ 1.7,\ 1.8$ at $L = 4, 6, 8, 10$: κ grows linearly with the volume here too. The power law through $L \geq 4$ has exponent 1.05 and extrapolates to 1.3×10^5 on 16^4 and 2.4×10^6 on 32^4. By (36) this is the aspect ratio, squared, of the Gaussian the flow has to represent in its own coordinates at weak coupling. The one-off exact integer elimination itself grows faster than the system: 0.1, 4.7, 99 and 824 seconds on one CPU core at $L = 4, 6, 8, 10$ (1,030 to 40,006 rows), as the fill-in grows with the extent, so at 32^4 (3.1×10^6 rows for $U(1)$) it becomes challenging with this elimination strategy.

The refinement level. One level of the multilevel construction has $\kappa = 1$ in $d = 2$ and 7 in $d = 4$, support at most 4, measured on extents 4 and 8 and the same on both; the stencil is translation invariant, so neither can depend on L. Nothing of extent-dependent size is eliminated or stored: the same fixed relation is applied at every site. The count $|D|$ of determined plaquettes still grows with V, but as V independent local evaluations, not as one system.

Table 4: Size and conditioning of the single-level solve against lattice extent, for the theories of this paper and the four-dimensional non-Abelian theories one would want next. Counts are real dimensions: plaquette variables $\dim G \cdot M$, rows of the solve $\dim G \cdot |D|$ with $|D| = \operatorname{rank} \boldsymbol{A}$ in (31), generated variables $\dim G \cdot |F|$, the size of $\boldsymbol{I} + \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}$. κ is measured with minimum-support pivoting (Table 3); in two dimensions it is V exactly; † marks the power law of Fig. 8, $\kappa \propto |F|^{1.05}$, extended. Last column: one refinement level in the coordinates of (39), the same at every extent.

			single level, whole lattice				refinement level
theory	L	V	variables	solve rows	generated	κ	κ
2D $U(1)$	8	64	64	1	63	64	1
	16	256	256	1	255	256	1
	32	1,024	1,024	1	1,023	1,024	1
	64	4,096	4,096	1	4,095	4,096	1
2D $SU(2)$	16	256	768	3	765	256	1
4D $U(1)$	4	256	1,536	771	765	379	7
	8	4,096	24,576	12,291	12,285	7,088	7
	16	65,536	393,216	196,611	196,605	$\approx 1.3 \times 10^{5\dagger}$	7
	32	1,048,576	6,291,456	3,145,731	3,145,725	$\approx 2.4 \times 10^{6\dagger}$	7
4D $SU(2)$	16	65,536	1,179,648	589,833	589,815	$\approx 1.3 \times 10^{5\dagger}$	7
	32	1,048,576	18,874,368	9,437,193	9,437,175	$\approx 2.4 \times 10^{6\dagger}$	7
4D $SU(3)$	16	65,536	3,145,728	1,572,888	1,572,840	$\approx 1.3 \times 10^{5\dagger}$	7
	32	1,048,576	50,331,648	25,165,848	25,165,800	$\approx 2.4 \times 10^{6\dagger}$	7

What the determined half requires. 771 of the 1536 plaquettes are a deterministic function of the other 765, each a signed sum of a median of 13 of them, so a generator can reproduce every marginal it emits and still be wrong about the half it computes: that half is fixed by high-order correlations, not by marginals. A null with the correct marginals and no correlation, the 765 angles drawn independently from the exact one-plaquette distribution and passed through the same solve, produces 419.4 ± 0.3 wrapping defects per configuration at $\beta = 0.5$ and 149.0 ± 0.3 at $\beta = 2$. At $\beta = 0.5$ the single-level model *is* this null: 419.3 defects against a reference of 352.969 ± 0.043,

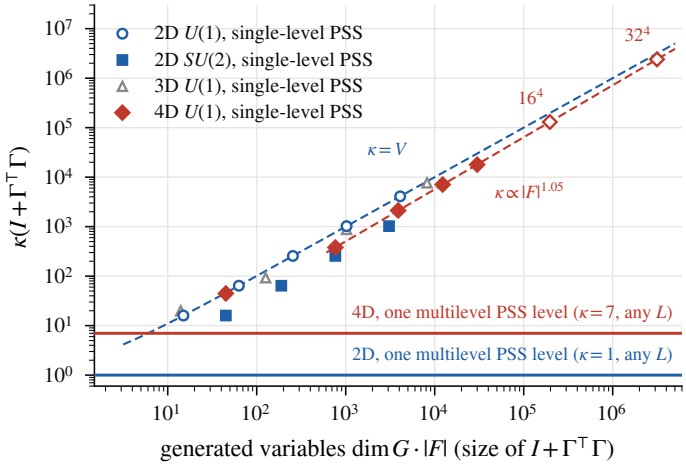


Figure 8: Condition number of the single-level solve against its size, for the three theories of this paper (3D $U(1)$ from Table 3 as a reference), and one refinement level of the multilevel construction, which is flat. Filled markers are measured; the open diamonds are the power law through the four-dimensional points extended to 16^4 and 32^4; the two-dimensional line is $\kappa = V$, exact. $SU(2)$ sits at three times the size of $U(1)$ with the same κ.

KL $= 47.5$, evidence recovery 8×10^{-7}. At $\beta = 2$ the same construction suppresses them to 0.0023, a factor 6.6×10^4 below the null, with KL $= 2.05$ and recovery 0.97. The correlation the solve requires is therefore learned in full at one coupling and not at all at the other, with no change of construction, constraint or objective, while the multilevel sampler learns it at both, reaching 352.990 ± 0.034 against 352.969 ± 0.043 at $\beta = 0.5$. The defects are magnetic monopoles, the coboundary of the integer lifts (Appendix M).

H VERIFICATION OF THE CONSTRUCTION

The construction makes structural claims — that a refinement is a bijection, that its Jacobian is one, that the solve supports are bounded — which are exact statements and can therefore be checked to machine precision rather than argued. Table 5 is that check. Every row was run in a reference implementation written for an arbitrary compact group and dimension, for $U(1)$ with angles over the reals, so the Abelian identities are tested exactly rather than modulo 2π, and for $SU(2)$ with unit quaternions, in double precision. No row involves training.

How the Jacobian is measured. Each generated element F is perturbed as $e^{\omega} F$ with ω in the Lie algebra, and each non-tree link U is read as $e^{\omega'} U$. Haar measure is bi-invariant, so its density in these coordinates is the same at every point, and the Haar Jacobian of the map is the determinant of $\partial\omega'/\partial\omega$ at $\omega = 0$, computed by automatic differentiation at two random points per lattice.

How the measure is tested. The importance-sampling rows generate every level from a conditional proposal and reweight by e^{-S}/q with (7), so they test the density bookkeeping and the measure, not only the geometry. The proposal draws each $F_\nu(X)$ from the law of its two electric terms in (39), von Mises for $U(1)$ and von Mises–Fisher on S^3 for $SU(2)$. On a two-dimensional torus the Haar-normalised partition function is known in closed form,

$$Z = \sum_{n \in \mathbb{Z}} I_n(\beta)^V \quad [U(1)], \qquad Z = \sum_{n \geq 1} \big(2 I_n(\beta)/\beta\big)^V \quad [SU(2)], \tag{51}$$

with $S = -\beta \sum \frac{1}{N} \operatorname{Re} \operatorname{tr} P$: at $L = 16$ the estimate is 210.9423 ± 0.0014 against 210.9424 for $U(1)$ at $\beta = 2$, and 248.0914 ± 0.0009 against 248.0905 for $SU(2)$ at $\beta = 3$. Where no closed form exists the same estimator is compared with importance sampling from Haar-distributed links, which uses no gauge fixing and none of the construction, and on 2^4 for $U(1)$ with the thermodynamic-integration $\log Z$ of Appendix K, which additionally tests the unimodularity of the coarse level. A row counts only if its estimate has at least $1{,}000$ effective samples.

Table 5: Numerical gates. Rows hold for both groups unless stated. "Exact": integer valued and hit exactly. Other entries are the largest deviation over the lattices listed. The last row is limited by that sampler's chart interpolant, not by the construction.

what is checked	lattices	worst
Structure of one refinement		
count $(d-1)(V-1)+d$; fixed links a spanning tree	$d = 2$–5, to 4^4	exact
refinement is a bijection	7 lattices, $d = 2$–5	$2.3{\cdot}10^{-14}$
generated element equals its plaquette	7 lattices, $d = 2$–5	$9{\cdot}10^{-16}$
cube identities hold on generated fields	7 lattices, $d \geq 3$	$5.3{\cdot}10^{-15}$
Haar Jacobian of the full map is one	$d = 2$–5, 17–387 vars	$2.2{\cdot}10^{-14}$
solve supports are $1, 1, 0, 4$	$d = 2$–5, extents $4, 8$	exact
condition number $1, 5, 7, 9$ for $d = 2$–5	$d = 2$–5, extents $4, 8$	extent independent
The measure, by importance sampling		
$\log Z$ and plaquette against the closed form	$4^2, 8^2, 16^2$, two β	pull ≤ 1.6
agreement with sampling from Haar links	$2^3,\ 4{\times}2^2,\ 2^4$	pull ≤ 1.7
$\log Z$ against thermodynamic integration	2^4, $\beta = 0.2,\ 0.5$	pull ≤ 0.7
The implementations used in the experiments		
each level is a unimodular change of variables	five levels of $2^4 \to 4^4$	exact
solves give back the coarse field and cube identities	five levels	$7{\cdot}10^{-15}$
two-dimensional block equals two refinements	$L_c = 2, 4, 8$	exact
$SU(2)$ block relation, inverse and closure	$L \leq 6$	$5{\cdot}10^{-14}$
$SU(2)$ reconstructed links are Haar	$L \leq 6$	$2.1{\cdot}10^{-3}$
$SU(2)$ block log-density against its Jacobian	$L \leq 6$	$7{\cdot}10^{-7}$

I ARCHITECTURE

The model is a stack of conditional flows. Level i draws its free variables from a learned conditional density and the constraint layer turns them into a valid configuration, $F^{(i)} \sim \widetilde{q}_i(\cdot \mid P^{(i-1)})$, $P^{(i)} = R_i(F^{(i)}; P^{(i-1)})$. Because R_i is a bijection of unit Haar Jacobian (37), the density transfers without a correction,

$$\log q_i\big(P^{(i)} \mid P^{(i-1)}\big) = \log \widetilde{q}_i\big(F^{(i)} \mid P^{(i-1)}\big), \qquad \log q(P) = \sum_{i=0}^{n} \log \widetilde{q}_i\big(F^{(i)} \mid P^{(i-1)}\big), \quad (52)$$

so the exact log-density needed by the objective (8) and by the importance weights costs no more than the per-level flow densities. Each $\widetilde{q}_i$ is a normalizing flow on $\mathbb{T}^{N^{(i)}}$ conditioned on the coarse configuration; the coarse level $i = 0$ is unconditional. Learning is therefore a conditional density-estimation problem per resolution, and the physics enters only through R_i, which has no parameters.

Coupling layers. $\widetilde{q}_i$ is a uniform base composed with n coupling layers. Layer ℓ partitions the components into active A_ℓ and passive $\bar{A}_\ell$ and maps

$$f_j \mapsto \mathrm{RQS}_K\big(f_j; \theta_j\big) \ \ (j \in A_\ell), \qquad \theta = c_\phi\big(\cos f_{\bar{A}_\ell}, \sin f_{\bar{A}_\ell}, P^{\mathrm{c}}\big) \in \mathbb{R}^{3K}, \quad (53)$$

with RQS_K the circular rational-quadratic spline (Durkan et al., 2019) on K bins, matched in value and derivative at $\pm\pi$, and c_ϕ the conditioner. Its last layer is zero-initialised, so each coupling starts at the identity and the stack starts at the base. For $SU(2)$, (53) acts on the angles of an equivariant parametrisation over a Haar base.

Why the split matters. A coupling layer represents the dependence of A_ℓ on $\bar{A}_\ell$ but none among A_ℓ, so the partitions decide what the stack can express, and two requirements act on them. Every component must be active in some layers and passive in others, which is what the alternation in Table 6 provides. And the constraint couples the free variables: a determined plaquette is a function of k_j of them, (32). If two of those inputs are active in the same layer, that layer cannot represent their joint effect on it, which is why the constrained splits admit at most one active input per determined plaquette. This is the only point at which the architecture is aware of the constraint at all.

Conditioner. c_ϕ is dense on the 2^4 coarse level and in $d = 2$, where it sees the coarse plaquette of the block and its neighbours, and for $SU(2)$ the holonomy commutator as well. On a $d = 4$ refinement it is a convolution on the fine lattice: free angles scattered to their plaquette positions, 36 input channels — $\cos f$, $\sin f$ on $\bar{A}_\ell$, active and passive indicators, and $\cos P^{\mathrm{c}}$, $\sin P^{\mathrm{c}}$ repeated along the refinement direction, six plaquette orientations each — a star stencil of radius 1 applied by periodic shifts, and SiLU activations. The shifts make c_ϕ exactly equivariant under lattice translations, so the symmetry of the target is built in rather than learned, and the parameter count is independent of the volume.

What the chain buys the model. The single-level model is the same machinery at $n = 0$: one unconditional flow over all $|F|$ free variables, followed by the global solve. The difference the chain makes is locality of the learning problem. On a refinement every determined plaquette is a function of at most four generated variables and the level's condition number is 7 at every extent (Appendix G), so a radius-one equivariant convolution conditioned on the coarse field can see every dependence the constraint creates, and one network serves every level and every volume. Solved on the whole lattice at once the median support is 13 with maximum 47 and grows with the extent, and $\kappa = 391$ on 4^4: no fixed stencil reaches the variables a determined plaquette couples, and the flow has to build those correlations implicitly through depth. The coarse configuration enters as conditioning input and carries the long-range structure, so each level has only to learn the degrees of freedom its own refinement adds. This is the architectural content of the construction, and Fig. 7 tests it at matched parameter count.

Table 6: Our models. Width/depth is that of c_ϕ in (53). Couplings run coarse level first, then each refinement; one entry means the same count at every level. Splits: *halves* of a random permutation; *parity* over sites times a per-layer orientation pattern; *constrained*, no determined plaquette with two active inputs in a coupling, whose models also pass c_ϕ each partially determined plaquette's running sum and distance to $\pm\pi$. At $\beta = 0.5, 0.8, 2.0, 3.0$ four further stacks of eight couplings act on the finished field, one per direction and conditioned on it blocked along that direction, an exact change of variables.

theory	levels	model	β	K	width/depth	couplings	split
4D $U(1)$	$2^4 \to 4^4$	multilevel	0.5, 0.8	8	64/3	24; 18, 18, 16, 12	constrained
	$2^4 \to 4^4$	multilevel	1.0	8	64/3	8	halves
	$2^4 \to 4^4$	multilevel	1.5	8	128/4	8	halves
	$2^4 \to 4^4$	multilevel	2.0, 3.0	8	128/4	8	parity
	4^4, single	single	all	8	128/4	8	halves
2D $U(1)$	$2^2 \to 32^2$, L=4	multilevel	3.0	8	64/3	6; 8 per level	halves
	$2^2 \to 64^2$, $L \geq 8$	multilevel	$0.1875L^2$	8	64/3	6; 8 per level	halves
	L^2, single, $L \leq 8$	single	$0.1875L^2$	8	8/2	48	halves
2D $SU(2)$	$2^2 \to 16^2$	multilevel	all	8	64/3	32; 8 base	halves
	16^2, single	single	all	8	64/3	32	halves

Table 7: Baselines. Link-space flows are trained against the same action as our models. The single-level plaquette flow is our own construction and is in Table 6.

theory	baseline	construction
2D $U(1)$	link multiscale (Abbott et al., 2024a)	coarse $L/2$ prior, then a fine flow
2D $U(1)$	link single level (Kanwar et al., 2020)	one flow on all links
4D $U(1)$	link multiscale (Abbott et al., 2024a)	as published, values from Fig. 4
2D $SU(2)$	link continuous flow (Gerdes et al., 2025)	ODE on links, `bijx`, 40 steps

The 2D rows use the refinement of Appendix F.4, with c_ϕ seeing the coarse plaquette of the block and its neighbours, and for $SU(2)$ the holonomy commutator. The $SU(2)$ levels apply their 32 flow steps over 16 group-element blocks with an interleave of 4, the base level 8 steps over 4; holonomies are drawn from the Haar base, not learned. The 4D coarse level is dense on its 45 free angles; the single-level model generates all 765 free angles of 4^4 through the solve of Appendix G and is the comparison in Fig. 7.

J TRAINING PROTOCOL

All models, ours and the link-space baselines we trained, minimise the reverse Kullback–Leibler divergence (8) with pathwise gradients through the flow and the constraint layer. No Monte Carlo configuration enters training; the ensembles of Appendix K are used only for validation. Adam, double precision. The 4D chain and the $SU(2)$ chain are trained as a whole, every level jointly from random initialisation. The 2D $U(1)$ chain is trained as a ladder along $\beta = cL^2$: one level at a time, each warm-started with the levels below it frozen, so the budget is per level. Schedules are set by the step target, and runs interrupted by cluster limits resume optimiser and scheduler state exactly.

Checkpoints are written during training together with a training-time estimate $\widehat{\mathrm{KL}}$ on fresh samples, every 5,000 steps on 65,536 samples in $d = 4$, or every 10,000 on 16,384 at $\beta = 0.5,\ 0.8,\ 3.0$; the same schedule is used for the baselines. One retained checkpoint per model is evaluated, and every number in the paper is recomputed from one fresh evaluation of 500,000 samples (Appendix L), never from the training-time estimate.

The settings of Table 8 are target dependent: width, couplings, clipping and schedule were chosen per coupling, and we claim no single recipe across the transition and no wall-clock advantage over Markov chain Monte Carlo (Appendix L).

K REFERENCE VALUES AND THE PARTITION FUNCTION

No Monte Carlo configuration enters training. The quantities here validate the trained samplers: reference expectation values for the reweighted observables, and the $\log Z$ that turns the estimated divergence into the true one.

Protocol. Reference ensembles are generated in link variables by heatbath, each link drawn exactly from its conditional von Mises distribution given its six staples, the two site-parity sublattices of a direction updated in turn so that every half-step is an exact block update. At each coupling eight chains of 30,000 sweeps are run, four from the ordered configuration and four from uniformly random links; the first 6,000 are discarded and every later sweep is measured. The quoted error is the larger of the autocorrelation-corrected error of the pooled mean and the standard error of the chain means, so neither an underestimated autocorrelation nor a confined chain can flatter it.

β	τ_{int} (sweeps)	$\langle \cos\varphi \rangle$	monopoles per configuration	$\log Z$
0.5	0.6	0.245790 ± 0.000043	352.969 ± 0.043	732.941 ± 0.009
0.8	1.1	0.402005 ± 0.000066	250.18 ± 0.06	420.204 ± 0.011
1.0	15	0.64944 ± 0.00048	48.5 ± 0.4	262.052 ± 0.019
1.5	0.7	0.813425 ± 0.000044	0.411 ± 0.004	75.942 ± 0.021
2.0	0.6	0.865396 ± 0.000020	0.0094 ± 0.0007	-45.114 ± 0.021
3.0	0.6	0.912952 ± 0.000011	none observed	-210.003 ± 0.021

Only $\beta = 1$, next to the transition, has a long autocorrelation. There ordered and disordered chains agree after 36 sweeps, two orders of magnitude inside the 6,000 discarded, and the plaquette distribution is a single broad peak of width 0.033 with monopole spread 28: no sign of two coexisting phases on this volume.

Flux sectors. Heatbath chains never change the flux through the six two-tori. A unit flux costs an action of about $2\pi^2\beta$, suppressed by e^{-30} at $\beta = 1.5$ and e^{-39} at $\beta = 2$, so a chain starting in one is a metastable artefact rather than an equilibrium sample, and is visibly displaced: $\langle \cos\varphi \rangle = 0.7998$ against 0.8134 at $\beta = 1.5$. The reference at $\beta \geq 1.5$ therefore uses trivial-sector chains only — five of eight at $\beta = 1.5$, seven of eight at $\beta = 2$, all eight at $\beta = 3$ — one of the three dropped at $\beta = 1.5$ going instead for a flux made ambiguous by a monopole, its mean agreeing with the rest. At $\beta \leq 1$ monopoles are dense and flux is not well defined, so all chains are used. The flow can generate every sector (Appendix E.2).

Cross-check in the sampler's own coordinates. To exclude an error in the target as the flow sees it, Metropolis updates were run on $p(f) \propto e^{-S(\varphi(f))}$ over the 765 free angles of the single-level

Table 8: Training settings for every model reported in Section 4. "exp" decays the learning rate exponentially from the first value to the second over the scheduled steps. Two-dimensional budgets for the $U(1)$ ladder are per level, in the order root, $L = 4, 8, 16, 32, 64$. Gradients are clipped at 1, except the 4D multilevel model at $\beta = 2$ at 10 and the baselines at 0.5. Each $SU(2)$ single-level run has converged or diverged well inside 25,000 steps, so the budget is not what limits it.

theory	model	steps	batch	learning rate	sched.
4D $U(1)$, β=0.5, 0.8	multilevel	260,000	192	$3{\cdot}10^{-4} \to 10^{-5}$	exp
4D $U(1)$, β=1.0, 1.5	multilevel	300,000	192	$3{\cdot}10^{-4} \to 5{\cdot}10^{-6}$	exp
4D $U(1)$, β=2.0	multilevel	300,000	192	$3{\cdot}10^{-4} \to 5{\cdot}10^{-6}$	exp
4D $U(1)$, β=3.0	multilevel	300,000	192	$3{\cdot}10^{-4} \to 5{\cdot}10^{-6}$	exp
4D $U(1)$, all β	single level	300,000	192	$3{\cdot}10^{-4} \to 5{\cdot}10^{-6}$	exp
2D $U(1)$, L=4–64	multilevel	20/50/30/30/50/45k	256	10^{-3}	ladder
2D $SU(2)$, all β	multilevel	100,000	128	$3{\cdot}10^{-4}$	const.
2D $SU(2)$, all β	single level	25,000, converged	128	$3{\cdot}10^{-4}$	const.
2D $U(1)$	link multiscale, baseline	80,000	512	10^{-3}	const.
2D $U(1)$	link single level, baseline	40,000	512	10^{-3}	const.
2D $U(1)$	plaquette single level	80,000	512	10^{-3}	const.

4^4 parametrisation at $\beta = 1$: sixteen chains of 3,000 sweeps from $f = 0$ and from uniform f give $\langle \cos\varphi \rangle = 0.6488$ and 0.6546 with 47.6 and 45.2 monopoles per configuration, each to about 0.008, both agreeing with the heatbath reference. The target and its coordinates are therefore correct, including the monopole sector.

Exact $\log Z$**.** Throughout this appendix and in the training code the action is written $S = \beta \sum_p (1 - \cos\varphi_p)$, which is (2) plus the constant $\beta \binom{d}{2} V$; the tabulated $\log Z$ are for this form and for the angle measure $\mathrm{d}f$ on the 765 free angles, whereas (2) uses normalised Haar measure, $\mathrm{d}f/2\pi$ per angle; they therefore differ from those of (2) by $-1536\,\beta + 765 \log 2\pi$, while every divergence and weight is unchanged. The flow's importance weights estimate $Z(\beta) = \int_{T^{765}} e^{-S(\varphi(f))}\,\mathrm{d}f$, whose logarithm follows from thermodynamic integration,

$$\log Z(0) = 765\,\log 2\pi, \qquad \frac{\mathrm{d}\log Z}{\mathrm{d}\beta} = -1536\,\big(1 - \langle \cos\varphi \rangle_\beta\big), \tag{54}$$

integrated by the trapezoid rule on a grid of spacing 0.02, refined to 0.005 across the transition. Each grid point uses sixty-four ordered-start chains contributing 1,600 measurements after thermalisation; the error, from the spread of chain means propagated with the trapezoid weights, is at most 0.021 in $\log Z$, hence at most that in the true divergence and 2% in the evidence recovery.

L Estimators and evaluation protocol

One evaluation per model. Each model is evaluated once, on $N = 500{,}000$ independent samples from its retained checkpoint with a fixed seed unless stated otherwise (the link-space continuous flow for $SU(2)$ at $N \le 10^4$, Section 4; the Wilson loops of Table 9 at 2×10^5), and every number reported for it — divergence, effective sample size, top weight, evidence recovery and every reweighted observable — comes from that one sample. Evaluations are never combined, training-batch statistics are never quoted, and any departure is stated where the number is used.

Weights, divergences and errors. For samples $f_i \sim q$ on the free angles, $\log \tilde{w}_i = -S(\varphi(f_i)) - \log q(f_i)$, the constant Jacobian of (33) cancelling throughout. Then

$$\mathrm{ESS} = \frac{\big(\sum_i \tilde{w}_i\big)^2}{N \sum_i \tilde{w}_i^2}, \qquad t = \frac{\max_i \tilde{w}_i}{\sum_i \tilde{w}_i}, \qquad \log \hat{Z} = \log \frac{1}{N} \sum_i \tilde{w}_i, \tag{55}$$

$$\widehat{\mathrm{KL}} = -\frac{1}{N} \sum_i \log \tilde{w}_i + \log \hat{Z}, \qquad \mathrm{KL} = -\frac{1}{N} \sum_i \log \tilde{w}_i + \log Z, \qquad R = e^{\log \hat{Z} - \log Z} = e^{\widehat{\mathrm{KL}} - \mathrm{KL}}, \tag{56}$$

with the $\log Z$ of Appendix K. Expectations are self-normalised, $\hat{O} = \sum_i \tilde{w}_i O_i / \sum_i \tilde{w}_i$; errors on ESS, KL and every reweighted observable are delete-block jackknife over the same N draws in 100 blocks, that on KL including the uncertainty of $\log Z$. The observable panels all report the relative accuracy $\hat{O}/O_{\mathrm{ref}} - 1$ of (9): against the exact finite-volume value in Figs. 3 and 5, and against the heatbath reference in Fig. 4, in units of 10^{-3} with the reference's own uncertainty as a band. The monopole number uses all N samples, since at $\beta = 2$ only about one configuration in a hundred carries one.

$\widehat{\mathrm{KL}}$ is the divergence available without $\log Z$, biased low by exactly $-\log R$ — a large gap when a model misses much of the target, $\widehat{\mathrm{KL}} = 3.40$ against $\mathrm{KL} = 10.0$ at $\beta = 1$. Every divergence in the paper is the true one. A flow has full support, so $\mathbb{E}_q[\tilde{w}] = Z$ and $R \to 1$ as $N \to \infty$ for every model; at finite N, R says how much of the evidence the drawn samples recovered, and may sit slightly above one, as at $\beta = 0.5$ ($R = 1.03$). With $\log Z$ known to 0.021, R is limited by the sample, not the reference.

Where the normalisation comes from. $\mathrm{KL}(q\|p) = \mathbb{E}_q[\log q + S] + \log Z$: the first term is a sample mean, the second a constant whose source decides what the reported number means. We take it from thermodynamic integration, so it never involves the model. The link-space baseline we compare against estimates it from the model instead, as $\log \widehat{Z} = \log \overline{\tilde{w}}$ (Abbott et al., 2024a, Fig. 4). The two differ by exactly the log of the evidence recovery,

$$\mathrm{KL}_{\widehat{Z}} = \mathrm{KL} - \log(1/R), \tag{57}$$

and by Jensen's inequality the model-based convention is a lower bound in expectation, short by $\log(1/R)$ nats, so it flatters a model precisely where the model covers the target least well. Figure 9 shows the effect on our own sampler: away from the transition the two agree to a few per cent, while at $\beta = 1$, where $R = 1.3 \times 10^{-3}$, the reported divergence falls from 10.0 to 3.40, suppressing the peak by a factor of three. The published link-space values cannot be corrected, their $\log \widehat{Z}$ not being reported; the direction of the effect makes our comparison against them conservative.

What the effective sample size shows, and the n_{eff} cut. Since $\sum_i \tilde{w}_i^2 \geq \max_i \tilde{w}_i^2$, the heaviest single sample caps it, $N \cdot \mathrm{ESS} \leq 1/t^2$: at $\beta = 1$, $t = 0.043$ caps the effective number of samples near 540 and the measured value is about 170. When t is far above $1/N$ the effective sample size rests on a handful of samples and is itself noisy, so t is reported next to it and an estimate is treated as a measurement only when no sample dominates. More importantly, it measures how evenly the weight is spread over the region the model reaches, not whether the model reaches the whole target — a model confined to part of it can have nearly uniform weights there — so every result is reported with its true divergence, evidence recovery and comparison against the reference, and the effective sample size is never used on its own as evidence of correctness. Points with $n_{\mathrm{eff}} = N \cdot \mathrm{ESS} < 100$ are drawn faded, one cut applied to every panel of a figure so that the same set of points appears throughout. The cut is the right criterion for the weighted estimators, which a single dominant weight invalidates, but not for the divergence: $\mathrm{KL}(q\|p) = \log Z - \overline{\log w}$ is an unweighted mean over draws from q, unbiased at any n_{eff}, with uncertainty $\mathrm{Var}(\log w)/N$ that the sample sizes used here resolve, so faded divergences are valid measurements.

M MONOPOLES

For $U(1)$ the solve (32) is exact over $\mathbb{R}$, while the physical plaquette is $\mathrm{wrap}\,\varphi \in (-\pi, \pi]$. With the integer lifts $k := (\varphi - \mathrm{wrap}\,\varphi)/2\pi$ and $\mathrm{d}\varphi = 0$,

$$m := \frac{1}{2\pi}\,\mathrm{d}\big(\mathrm{wrap}\,\varphi\big) = -\,\mathrm{d}k, \qquad Q := \frac{1}{2\pi}\sum_x \mathrm{wrap}\,\varphi_{01}(x) = -\sum_x k_{01}(x) \quad (d = 2), \tag{58}$$

so the monopole charge of a three-cube is the coboundary of the integer lifts of its faces, and in two dimensions the topological charge is their sum. Generated angles lie in $(-\pi, \pi]$ and have $k = 0$, so a monopole is produced exactly when the real-valued solve of a determined plaquette leaves $(-\pi, \pi]$, crossing the seam at $\pm\pi$. The sector is therefore inside the parametrised set rather than excluded by it: uniformly random free angles give 476 monopoles per configuration on 4^4, more than the reference at any coupling studied. How closely the trained flow reproduces their density at weak coupling is a question of sampling rare defect configurations, which is beyond the scope of this work (Section 6).

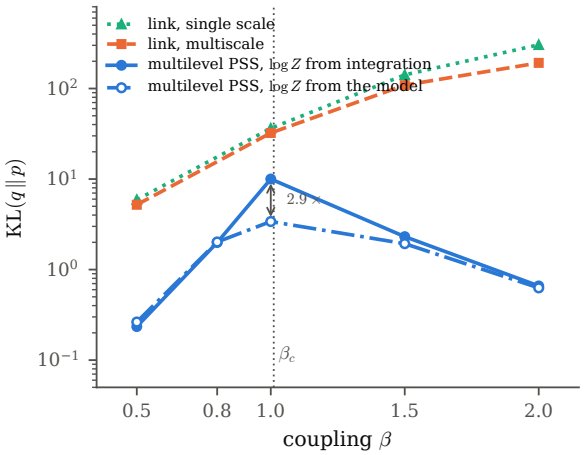


Figure 9: The same four-dimensional models and evaluations with the normalisation taken two ways: filled markers use $\log Z$ from thermodynamic integration, open markers the model's own $\log \widehat{Z}$. The vertical gap is $\log(1/R)$ and nothing else, so it closes where the sampler covers the target and opens where it does not. A missed mode contributes nothing to the model-based quantity, which is how a transition peak can be flattened by the convention alone.

N TWO DIMENSIONS

The same constraint layer, with the block relation of Appendix F.4 as the linear system, gives a multilevel sampler for two-dimensional compact $U(1)$: the coupling layers of Appendix I and the objective of Appendix J, an unconditional base on a 2×2 lattice, then power-of-two doublings to L, each new level warm-started with the lower levels frozen. Every L is trained along a line of constant physics, $\beta = cL^2$, so a ladder describes one continuum theory at successive resolutions.

Topology is not what this construction addresses. The difficulty of the flow is set by how active the topology is, the spread of Q, rather than by β directly: at small c, where $\langle Q^2 \rangle$ is large, the target is strongly multimodal in Q and the mode-seeking reverse-KL flow loses ESS, while where the topology is mild ($\langle Q^2 \rangle < 1$ at every level) ESS is high and flat across scales. This is a coverage problem in the topological variable, not a defect of the constraint layer, whose exactness is independent of c. We do not address it here; methods that resolve the sectors explicitly, such as the sector-resolved flow sampling of Singha et al. (2026b), are complementary to the construction of this paper and could be combined with it.

Topological susceptibility. The plaquette is strongly self-averaging and therefore a weak test of a sampler: even weights dominated by a few configurations reproduce it to parts in 10^5. The susceptibility is the sensitive observable, and on the torus it is known exactly from the character expansion at every (β, V), so no reference simulation is needed. Figure 10 shows the reweighted χ_{top} against that value along $\beta = 0.188\,L^2$. The multilevel sampler agrees within 0.7σ at every size, the largest deviation being $+0.5\%$ at $L = 32$. Points whose weights are degenerate are omitted, their errors being too wide to test; the single-level plaquette flow at $L = 8$ is drawn at -100%, its χ_{top} identically zero with every configuration at $Q = 0$, which is the clearest statement that imposing the closure constraint does not by itself generate topology. That this is topological rather than a property of the global solve is settled by $SU(2)$: two dimensions give it the same single closure row and hence the same $\kappa = V$, since $\mathbf{\Gamma}_{SU(N)} = \mathbf{\Gamma}_{U(1)} \otimes \boldsymbol{I}_{\dim G}$, but it is simply connected and has no sectors to collapse into, and its single-level sampler is efficient at $L = 16$ (Fig. 5), degrading only at long correlation length and through a weight tail rather than a bias.

Protocol, and the check against the exact value. Every point of Fig. 3 is one evaluation with seed 0, and every size is a stage of the same ladder, with no averaging over seeds and no selection among draws. Used as an independence-Metropolis proposal the sampler has a topological autocorrelation time near unity, 0.69–1.03 for $L \leq 32$. At the largest extent of the main-text ladder, $L = 64$ ($\beta = 768$), the multilevel sampler reaches ESS $= 0.101 \pm 0.006$ after 45,000 steps at the finest level, with plaquette and susceptibility within 0.2σ of the exact values and $\mathrm{KL}(q\|p) = 3.9$ nats against 17 for the link-space multiscale flow.

$SU(2)$ Wilson loops. Table 9 evaluates both plaquette-space samplers on Wilson loops of area 1, 4 and 9. The action is the 1×1 loop, so $\langle \frac{1}{2} \mathrm{tr}\, U_p \rangle$ is essentially the training target, whereas the 2×2 and 3×3 loops are observables neither sampler was trained to reproduce and which two-dimensional

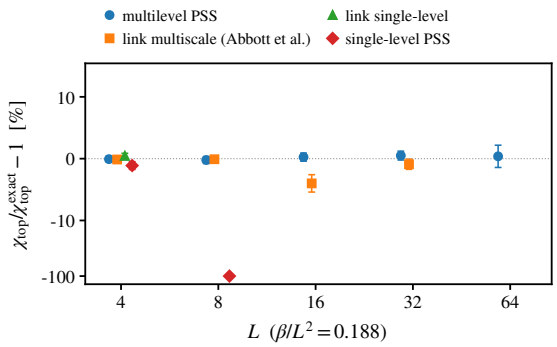


Figure 10: Reweighted topological susceptibility against the exact character-expansion value along $\beta = 0.188\,L^2$, as a relative deviation. Samplers whose weights are degenerate are omitted; the single-level plaquette flow at $L = 8$ has $\chi_{\rm top} \equiv 0$ and is drawn at -100%.

$SU(2)$ supplies in closed form. All thirty measurements agree with the exact values to within 2.2σ, so the reweighted estimates are consistent with the exact values, within uncertainties, beyond the quantity the flow optimises. The failure mode at weak coupling is a heavy weight tail rather than a displaced bulk: at $\xi/a = 6.6$ the largest single weight carries 1.5% of the single-level estimator and 48% of the link-space one, against at most 4.4×10^{-4} for the multilevel sampler at any coupling. Effective sample sizes and reweighted observables are draw-dependent at these volumes, so each plaquette-space number quoted in Fig. 5 is one evaluation at $N = 500{,}000$ at a fixed seed (the link-space flow at $N \le 10^4$, Section 4), and where we suspected residual bias we repeated it with independent seeds; at $\beta = 40$ this rejected an earlier 12,000-step model whose deviation reproduced at -1.2×10^{-5} over five draws (13σ) and confirmed the model reported here at $+1.5 \times 10^{-6}$ over four draws (0.7σ).

Table 9: Wilson loops for the two plaquette-space samplers on 16^2 $SU(2)$, reweighted at $N = 2\times10^5$ and compared with the exact values. Entries are $10^5(\langle W\rangle - W_{\rm exact})$, with the deviation in standard deviations in parentheses; $W_{\rm exact}$ is listed because the loop decays with area. The action is the 1×1 loop, so 2×2 and 3×3 are observables neither sampler was trained to reproduce.

		multilevel PSS			single-level PSS		
β	ξ/a	1×1	2×2	3×3	1×1	2×2	3×3
2.2	1.1	$-11.8\,(2.2)$	$-9.7\,(1.2)$	$+0.4\,(0.1)$	$-9.0\,(1.6)$	$-15.6\,(1.8)$	$+6.8\,(0.9)$
8	2.2	$-0.6\,(0.3)$	$-1.8\,(0.3)$	$-3.5\,(0.3)$	$-1.9\,(0.9)$	$-4.5\,(0.5)$	$+4.4\,(0.3)$
18	3.4	$+0.2\,(0.2)$	$+6.1\,(1.1)$	$+8.4\,(0.7)$	$-1.1\,(1.0)$	$-7.6\,(1.2)$	$-32.0\,(1.9)$
40	5.1	$-0.6\,(0.8)$	$-0.4\,(0.1)$	$+2.7\,(0.3)$	$-1.1\,(0.5)$	$-14.8\,(1.5)$	$-45.2\,(1.9)$
65.9135	6.6	$-0.2\,(0.5)$	$-0.0\,(0.0)$	$+2.0\,(0.4)$	$+0.7\,(0.8)$	$+3.0\,(0.6)$	$+4.4\,(0.4)$

$W_{\rm exact}$ at 1×1, 2×2, 3×3:
$\beta = 2.2$: 0.4645, 0.0465, 0.0010; $\beta = 8$: 0.8192, 0.4505, 0.1662
$\beta = 18$: 0.9179, 0.7099, 0.4625; $\beta = 40$: 0.9627, 0.8591, 0.7105
$\beta = 65.9135$: 0.9773, 0.9124, 0.8136