Multiscale Data-Driven Energy Estimation and Generation

Tanguy Marchand, Misaki Ozawa, Giulio Biroli, Stéphane Mallat

Multiscale Data-Driven Energy Estimation and Generation

Tanguy Marchand,[1] Misaki Ozawa,[2,3] Giulio Biroli,[2] and Stéphane Mallat[4]
[1]*Département d'Informatique de l'Ecole Normale Supérieure, ENS, CNRS, Université PSL, Paris, France*
[2]*Laboratoire de Physique de l'Ecole Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Paris, France and Université Paris-Diderot, Sorbonne Paris Cité, Paris, France*
[3]*Université Grenoble Alpes, CNRS, LIPhy, 38000 Grenoble, France*
[4]*Collège de France, DI ENS,CNRS, Université PSL, Paris, France and CCM Flatiron Institute, New York, New York, USA*

(Received 18 July 2022; revised 2 October 2023; accepted 28 August 2023; published 30 November 2023)

We develop a multiscale approach to estimate high-dimensional probability distributions. Our approach applies to cases in which the energy function (or Hamiltonian) is not known from the start. Using data acquired from experiments or simulations we can estimate the energy function. We obtain a representation of the approximate probability distribution based on a multiscale cascade of conditional probabilities. This representation allows for fast sampling of many-body systems in various domains, from statistical physics to cosmology. Our method—the wavelet-conditional renormalization group (WCRG)—proceeds scale by scale, estimating models for the conditional probabilities of "fast degrees of freedom" conditioned by coarse-grained fields. These probability distributions are modeled by energy functions associated with scale interactions, and are represented in an orthogonal wavelet basis. The WCRG decomposes the microscopic energy function as a sum of interaction energies at all scales and can efficiently generate new samples by going from coarse to fine scales. Near phase transitions, this representation of the approximate probability distribution completely avoids the "critical slowing-down" of direct estimation and sampling algorithms. This is explained theoretically by combining results from RG and wavelet theories, and verified numerically for the Gaussian and φ^4-field theories. We show that multiscale WCRG energy-based models are more general than local potential models and can capture the physics of complex many-body interacting systems at all length scales. This is demonstrated for weak-gravitational-lensing fields reflecting dark-matter distributions in cosmology, which include long-range interactions with long-tail probability distributions. The WCRG has a large number of potential applications in nonequilibrium systems, where the underlying distribution is not known *a priori*. Finally, we discuss the connection between the WCRG and deep network architectures.

DOI: 10.1103/PhysRevX.13.041038

Subject Areas: Interdisciplinary Physics, Statistical Physics

I. INTRODUCTION

For a long time, physicists have worked in a setting in which the energy function (or Hamiltonian) is known from the start, and the challenge is to devise methods to explain the resulting physical properties and develop techniques to provide fast sampling of physical configurations. Nowadays, we are witnessing a change of paradigm. Thanks to a large amount of data available, it has become possible to estimate the energy function statistically, i.e., by inferring it from a large dataset of observations from experiments or simulations in cases in which the energy function is not known from the start. The majority of physical systems are characterized by a probability distribution. Whereas at equilibrium, this distribution has the Boltzmann-Gibbs form, there are no general principles to identify an out-of-equilibrium distribution. Devising approximate representations of such probability distributions that can be estimated efficiently and that enable fast sampling is now a major endeavor at the center of intense research activity.

This paradigmatic shift has been fueled by the upsurge of impressive results in machine learning (ML) combined with the availability of a massive amount of data in many field of physics, e.g., cosmology, condensed matter, statistical physics, and also the interest in nonequilibrium systems. For instance, generative models have been used in cosmology to constrain the values of the cosmological

constants [1], and in condensed matter theory and statistical physics to characterize phase transitions [2]. Beyond physics, estimating the probability distribution of a very large number of variables from datasets of examples is in fact also a major challenge, which is called "unsupervised learning" in data sciences. It has considerable applications in all sciences, including neuroscience [3] and protein sequence data analysis [4]. In these cases (and others), not only the probability distribution is of interest, but it is also instrumental in estimating the effective energy function governing the interactions between microscopic degrees of freedom. In image processing, such estimation problems were introduced in Ref. [5] to synthesize and discriminate image textures with maximum entropy models [6,7]. In recent years, generative models based on deep neural networks have obtained impressive results for images [8,9]. As exciting as it is, the theoretical understanding of all these new methods remains a completely open problem as those techniques are too complex to be analyzed. The major questions are the following: How faithful is their representation of the probability distribution? Why (and whether) can they be trained even in the presence of long-range correlations? Why can they produce new typical configurations very fast even in the presence of long-range correlations, e.g., at critical points?

Our work addresses these issues from a complementary angle compared to direct studies of standard ML algorithms. We introduce the wavelet-condition renormalization group (WCRG). As we explain below, this method is different in scope from standard RG, but it relies on the same key premise. For this method, and focusing on configurations (samples) that come from physics, we answer all previous questions by providing both a theoretical explanation and numerical confirmations. Our aim is to obtain a generative model that can be fully analyzed and that provides an estimation of the effective energy function. Our method presents some of the features that make neural-network-based generative models very useful, but it is the advantage of being interpretable. Hence, it can provide important new physical insights for physics, in particular, for nonequilibrium systems. As we see, the trade-off (which arguably is an advantage, too) is that it requires physical insights.

The problem illustrated in Fig. 1 can be described in a nutshell as follows. One has a dataset of microscopic many-body configurations represented as spatial fields φ_0 provided by either multiple experimental observations or numerical simulations. One can think of φ_0 as the magnetization field in a model of ferromagnetism, e.g., the φ^4 model [10], or as a density field in a fluid, or in a dark-matter distribution in cosmology. Thus, φ_0 can be either in thermal equilibrium (e.g., the φ^4 model) or inherently nonequilibrium (e.g., dark matter). The underlying probability distribution $p_0(\varphi_0) = e^{-E_0(\varphi_0)}/Z_0$ is unknown *a priori* and must be estimated from the dataset to generate new

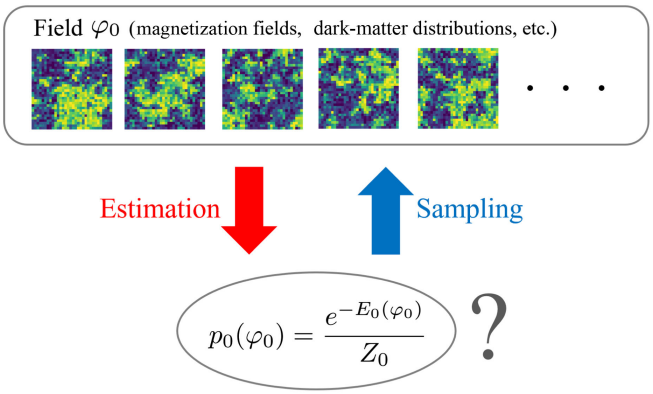

FIG. 1. A major challenge in physics and machine learning is to estimate the probability distribution p_0 of a field φ_0 and its microscopic energy function E_0, from examples of fields or configurations. New fields can then be generated by sampling this probability distribution. The WCRG is a fast multiscale approach which completely eliminates "critical slowing-down" phenomena near phase transitions.

samples efficiently and determine the energy function $E_0(\varphi_0)$ [we set $\beta = (k_B T)^{-1} = 1$; i.e., we measure the energy in units of $k_B T$]. For nonequilibrium systems, $E_0(\varphi_0)$ can be considered as an effective Hamiltonian that allows us to represent the high-dimensional (and nontrivial) probability distribution in a compact way. The tasks mentioned above are usually very difficult for two reasons. First, general high-dimensional distributions can be estimated only if the number of examples in the dataset grows exponentially with the system size [11]—an impossible requirement in practice. This curse of dimensionality can be avoided with prior information, which specifies suitable families of models for the probability distribution. Second, for systems that have fluctuations on a wide range of length scales, particularly near critical points, the estimation of model parameters is usually badly conditioned. It requires a large number of iterations, which grows as a power law of the system size and leads to large estimation errors. Furthermore, estimating the parameters and generating new samples requires performing Monte Carlo simulations [12]. Close to or at a critical point, Monte Carlo computations are hampered by a critical slowing-down [13,14], which produces a strong divergence of the mixing or decorrelation timescales when the system size increases.

By addressing these two problems, we develop a new and interdisciplinary approach to estimating an energy function and generating new samples. It combines and leverages ideas from the theory of RG, on the one hand, and from the wavelet theories, on the other hand, to define new energy-based models of probability distributions, which ensure that the estimation is well conditioned.

The WCRG proceeds as standard RG by building effective theories iteratively across many different scales by separating "fast, short-wavelength degrees of freedom" from "slow, long-wavelength degrees of freedom" at each

scale. The key concept from the RG theoretical framework [15] that is also central to our work is the property of the fast, short-wavelength degrees of freedom of being noncritical at each scale. From the mathematical point of view, this means that their probability distribution conditioned on the configuration at the coarser scale is nonsingular. Hence, it can be approximated and estimated efficiently. We use orthogonal wavelet theory to decompose a field in its coarser and fast components. Representing the fast degrees of freedom by wavelet fields provides localized representations of large classes of energy functions, which leads to an efficient calculation of marginal integrations. From the conditional probability at each scale and going progressively from the coarsest to the finest scales, one can obtain a novel multiscale representation of the probability distribution of φ_0. The WCRG is based on this RG representation and estimates the probability distribution $p_0(\varphi_0)$ by the product of conditional probability models.

There are a few but essential differences from the classical RG approach. The microscopic energy function (or Hamiltonian) is not known from the start. Providing an accurate estimate of that microscopic energy function is in fact one of the main objectives of our approach. Classical RG starts from an accurate knowledge of the microscopic energy and progressively loses microscopic information by coarse-graining. Instead, the WCRG progressively gains new information at each scale from data by estimating the probability distribution of the fast degrees of freedom. Our iteration starts from the longest scale at which the original data are obtained or available and progresses inversely from coarse to fine until reaching to the shortest (microscopic) scale of interest. Technically, our iterative procedure does not take the form of typical RG flow equations because the central entity in the iteration is the conditional probability distribution.

The WCRG is based on models of conditional probabilities across scales by introducing energy functions associated with interactions across scales that we call "scale interactions." This leads to models with higher expressibility than the ones based on estimating the energy function directly at the microscopic scale. Remarkably, our approach also guarantees that the method remains well conditioned even at phase transitions. Thanks to this property, it requires only a relatively small number of data, and it can generate new samples very fast. For instance, it allows generating new samples on times of order one, even at the critical point of statistical physics models. Our numerical applications to two-dimensional Gaussian and φ^4-field theories confirm that, indeed, the WCRG circumvents the problem of critical slowing-down completely. Its successful application to weak-gravitational-lensing fields in cosmology shows that the WCRG is a powerful approach to model complex many-body interactions. This is a highly nontrivial case study because the system is inherently out of equilibrium and is characterized by long-range interactions due to gravity (see Sec. VI. C).

The use of RG techniques and ideas in data sciences and machine learning has attracted a lot of attention over the years. All of these interesting applications have elements in common with our approach but also differ in aims and content. In particular, RG has been used recently as an inspiring analogy to study image generative models with deep neural networks in Refs. [16–19]. In Ref. [20], a conditional probability representation associated with wavelet transform similar to ours was studied within the normalizing flow framework with neural networks. However, the complexity of these architectures does not yet allow any mathematical analysis of these models, and thus do not provide explicit representations of the energy. Several works [21,22] also demonstrated close relations between deep convolutional network representations and iterated wavelet transforms with nonlinearities. They capture scale interactions but without conditional probabilities. As we discuss in the Conclusion, the WCRG provides a framework to analyze convolutional neural networks by modeling scale interactions through conditional probabilities, contrary to simpler methods that are based on energy models at a given scale.

The problem we address has also important connections with the topic of sampling in scientific computing. In this research field, many important techniques were developed over the years [multiscale Markov chain Monte Carlo (MCMC), multigrid, cluster methods, and umbrella sampling]. Recently, there has also been a surge of activity in relation to machine learning; see, e.g., Refs. [23,24]. Note, however, that the setting is completely different from ours. In this case, the energy function is known from the start, one has no data available, and the aim is to develop a fast method to obtain samples from the known probability distribution associated with the energy function at hand. Our case is quite the opposite: We have data at our disposal but no knowledge of the energy function. The WCRG approximate representation of the probability distribution allows for fast sampling, but it needs to be estimated. Data are the key ingredients that make this possible.

We close this introduction by suggesting potential applications of the WCRG. We envision two main directions. First, it can be used for nonequilibrium systems, e.g., for active and driven out-of-equilibrium systems, to estimate the relevant interactions between degrees of freedom at each scale and assessing the out-of-equilibrium nature of the problem at hand (for instance, testing if very long-range effective interactions are generated). Second, it can be used to perform directly RG from data, and hence estimating the critical theory and critical exponents, thus providing an alternative to standard RG methods. Its main advantage is to be data driven and versatile in the choice of the energy function, and hence directly applicable to challenging nonequilibrium problems.

A. A brief introduction to the WCRG

One of the key elements of RG is coarse-graining. From an input field φ_0 in d spatial dimensions, which may be an image ($d = 2$), one computes coarse-grained versions φ_j at each length scale $2^j a$, where a is the unit of length that is set to 1. The coarse-grained fields φ_j are illustrated in Fig. 2, together with wavelet fields $\bar{\varphi}_j$ corresponding to fast degrees of freedom discussed below. The probability distribution $p_j(\varphi_j)$ is defined at each scale 2^j by an energy function $E_j(\varphi_j)$. This energy is parametrized by a vector of coupling parameters θ_j, with $E_j(\varphi_j) = \theta_j^T U_j(\varphi_j)$, where U_j is the basis functions (or "operators" in the quantum-field-theory jargon). RG procedures specify the evolution of $p_j(\varphi_j)$ across scales through the RG flow of coupling parameters θ_j. In an RG step, one computes $p_j(\varphi_j)$ and θ_j from the finer scale $p_{j-1}(\varphi_{j-1})$ and θ_{j-1}, by integrating over the fast degrees of freedom of φ_{j-1} which are eliminated in φ_j. These degrees of freedom represent the short scale or highest wave-vector variations of φ_{j-1}. The theories describing large-scale critical fluctuations at a phase transition correspond to a nontrivial fixed point of the RG transformation. To compute this fixed point for the φ^4 model, Wilson originally represented the fast degrees of freedom over a basis of functions that are well localized in the Fourier basis [25]. He did so by defining what was later called a Shannon wavelet basis [25]. A considerable body of work has then been devoted to RG calculations in Fourier bases [15], and also in wavelet bases [21,26,27], particularly with perturbative methods in dimension $d = d_U - \epsilon$ (where d_U is the upper critical dimension, which is equal to 4 for the φ^4 model).

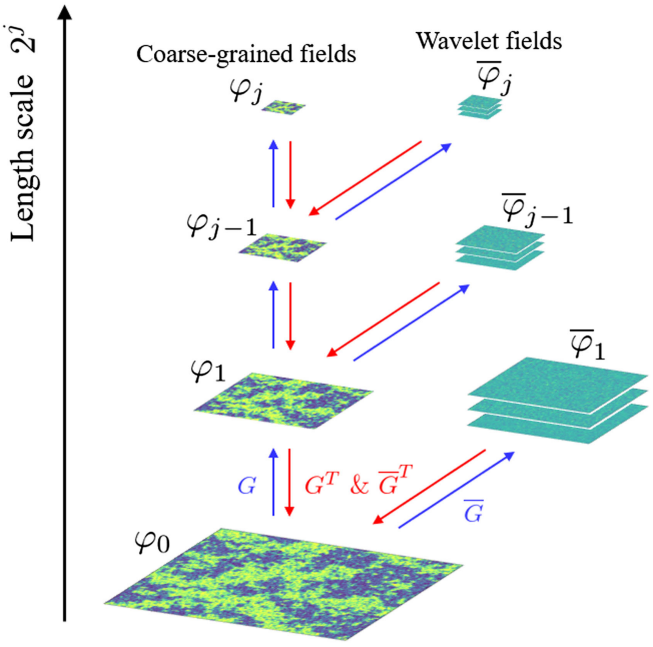

FIG. 2. Coarse-grained fields φ_j and wavelet fields $\bar{\varphi}_j$ at length scales 2^j are iteratively computed from a field φ_{j-1} at a finer scale 2^{j-1}. It is implemented with orthogonal convolutional and subsampling operators G and $\bar{G}$. A wavelet field $\bar{\varphi}_j$ represents the "fast degrees of freedom" of φ_{j-1} which have disappeared in φ_j. It contains three subfields in dimension $d = 2$ corresponding to spatial fluctuations along different orientations. The inverse wavelet transform reconstructs φ_{j-1} from φ_j and $\bar{\varphi}_j$ with the adjoint operators G^T and $\bar{G}^T$. Near the phase transition, a coarse-grained field φ_j has long-range spatial correlations, whereas $\bar{\varphi}_j$ has short-range correlations.

The wavelet-conditional renormalization group does not focus on the probability distribution $p_j(\varphi_j)$, or on $E_j(\varphi_j)$ at a given scale, or on θ_j. It is instead based on a different representation of microscopic probability distributions. It relies on conditional probabilities, which specify interactions across scales. These are functions of the wavelet fields $\bar{\varphi}_j$ representing the fast degrees of freedom or high-wave-number fluctuations. The importance of scale interactions is well known for multiscale physics phenomena such as fluid turbulence, producing energy cascades across scales [28–30]. Wavelet transform has been used to analyze these interactions [31] and to introduce specific parametric models such as multifractal random walks [32,33]. Low-dimensional wavelet scale interaction models have been shown to be sufficient to synthesize complex turbulent fluids and mass density fields in astrophysics [34–36]. However, such approaches do not provide a general numerical and theoretical framework that enables us to recover the energy functions of probability distributions. Energy functions are a very useful output of our method as, besides their fundamental character, they provide a compact and interpretable representation of high-dimensional probability distributions.

Within the WCRG, the connection between scales is studied as follows: At each scale 2^j, a wavelet transform decomposes a field φ_{j-1} into wavelet fields $\bar{\varphi}_j$, which capture its high-wave-vector variations, and a coarse-grained field φ_j [37]. In spatial dimension $d = 2$, $\bar{\varphi}_j$ has three wavelet subfields corresponding to different spatial orientations that we call channels. This is shown in Fig. 2 for the two-dimensional φ^4-field theory just above the critical point. In spatial dimension d there are $2^d - 1$ channels. The WCRG focuses on the conditional probability $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ of each wavelet field $\bar{\varphi}_j$ given the coarse-grained field φ_j. The microscopic probability $p_0(\varphi_0)$ is factorized as a product of these conditional probabilities up to a maximum scale which (without loss of generality) is $L = 2^J$:

$$p_0(\varphi_0) = \alpha p_J(\varphi_J) \prod_{j=1}^{J} \bar{p}_j(\bar{\varphi}_j|\varphi_j),$$

where α is a Jacobian constant. Each conditional probability $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ is represented by a Gibbs energy function

$\bar{E}_j$ which specifies scale interactions. The energy $\bar{E}_j = \bar{\theta}_j^T \bar{U}_j$ is represented as a linear combination of local operators $\bar{U}_j$ weighted by conditional coupling parameters $\bar{\theta}_j$. The microscopic probability $p_0(\varphi_0)$ is specified by the WCRG coupling parameters of all $\bar{p}_j$ and of the maximum scale p_J:

$$\{\theta_J, \bar{\theta}_j\}_{1 \le j \le J}.$$

Using this representation, the WCRG estimates from data the probability distribution, the energy function, and samples new configurations. Note that we introduce the overbar notation to denote quantities which are across scales, e.g., conditional probabilities, energies, and coupling associated with scale interactions. The coupling parameters $\{\theta_J, \bar{\theta}_j\}_{1 \le j \le J}$ are obtained from the training dataset by maximum-likelihood estimation, which is equivalent to minimize a Kullback-Leibler divergence. They are used to generate new samples φ_0 of the microscopic field distribution $p_0(\varphi_0)$. The theories describing large-scale fluctuations at critical points correspond to fixed points $\bar{\theta}_j = \bar{\theta}_{j-1}$. The WCRG parametrization remains stable even at phase transitions because it relies on a parametrization $\bar{\theta}_j$ of interaction energies $\bar{E}_j$ instead of a parametrization of E_j. Indeed, although each φ_j has long-range correlations and a singular covariance matrix at phase transitions, wavelet fields $\bar{\varphi}_j$ still have short-range correlations, as proved by the theorem in Sec. IV. This property is illustrated by the decomposition of φ^4 field near phase transition in Fig. 2. As a result, the WCRG coupling parameters do not suffer from any critical slowing-down, for Gaussian fields, φ^4 models, and simulated cosmological data, as we show numerically.

To recover the microscopic energy function $E_0(\varphi_0)$, we regress a parametrized model of the free energy $\bar{F}_j(\varphi_j)$ associated with each interaction energy $\bar{E}_j$. It is the normalization factor of the conditional probability $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$. The microscopic energy is then obtained as a sum of all interaction energies parametrized by $\bar{\theta}_j$, together with their free energies. As opposed to microscopic energies E_0 with local potentials, such multiscale models can also capture long-range interactions.

The WCRG sampling proceeds from coarse to fine scales by iteratively sampling $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$. It computes a new field $\bar{\varphi}_j$ given φ_j from which we recover the finer scale field φ_{j-1} with an inverse wavelet transform. It does not require us to compute the free energies $\bar{F}_j$, which are normalization factors. As opposed to inverse RG algorithms [38,39], which also proceed from coarse to fine scales, it does not sample directly the distributions p_j to recover each φ_j. Indeed, sampling fast degrees of freedom $\bar{\varphi}_j$ avoids critical slowing-down phenomena, because they have a short-range correlation.

In summary, the WCRG estimates and samples conditional probability distributions of wavelet fields connecting fluctuations across scales. It relies on the mathematics and physics of RG, which focuses on interactions across scales and fast degrees of freedom [15]. The manuscript is organized as follows. Section II reviews the properties of wavelet transforms and their connection with the RG formalism. Section III introduces the wavelet-conditional renormalization-group representation of probabilities and energy functions. Section IV presents maximum-likelihood estimations of the WCRG coupling parameters. It shows that the resulting algorithms are well conditioned and have a fast convergence that is not affected by critical slowing-down. Section V relates the WCRG coupling parameters to microscopic energy functions through free-energy calculations, which are performed with thermodynamic integrations. Section VI gives numerical applications of the WCRG to two-dimensional systems, Gaussian theory, φ^4-field theory, and weak-lensing cosmological data, demonstrating the fast convergence of estimation and sampling algorithms even at critical points. Section VII provides a conclusion with a discussion and perspectives.

II. RENORMALIZATION GROUP AND WAVELETS

The pioneering works on RG by Kadanoff *et al.* [40], Wilson [25], Wilson and Fisher [41], and others [15,42] rely on scale separation. They characterize the evolution of the energy function from fine to coarse scales by progressively integrating out fast degrees of freedom. Critical phenomena such as phase transitions are identified as fixed points of the renormalization group. In the following, we set up the formalism needed to perform the renormalization-group transformation over orthogonal wavelet coordinates. This provides the basic tools to then introduce a conditional renormalization based on conditional probabilities across scales.

A. RG in a wavelet orthogonal basis

We focus on d-dimensional scalar field theories, but our framework can be straightforwardly extended to vector fields. We consider theories that are translational invariant. In order to provide a concrete example of our theory, we focus on the φ^4-field theory, which is a central model in the theory of second-order phase transitions. Henceforth, we denote by φ_0 a many-body configuration field. The site index i in $\varphi_0(i)$ belongs to a d-dimensional lattice of lattice spacing $a = 1$, linear size L, and hence with L^d sites, e.g., $d = 1$ (linear), $d = 2$ (square), or $d = 3$ (cube). The probability distribution of φ_0 is given by the Boltzmann law $p_0(\varphi_0) = Z_0^{-1} e^{-E_0(\varphi_0)}$, where E_0 is the (configurational) energy function of the field φ_0. We set $\beta = (k_B T)^{-1} = 1$; i.e., we measure the energy in units of $k_B T$. The notation $\langle \cdot \rangle_p$ is used to denote statistical averages over the probability

distribution p. Throughout the paper, we consider the microscopic scalar field φ_0, which is normalized such that $\langle\varphi_0(i)\rangle_{p_0} = 0$ and $\langle|\varphi_0(i)|^2\rangle_{p_0} = 1$. Note also that we do not use the field theory terminology "connected correlation function" but the more standard one in probability theory, "covariance."

1. Coarse-graining and renormalization

Let us denote φ_j the coarse-grained version of φ_0 at a length scale 2^j defined over a coarser grid (lattice) with intervals 2^j and hence $(L2^{-j})^d$ sites. In this coarse-graining procedure, the lattice spacing is each time reduced by 2. The coarser field φ_j is iteratively computed from φ_{j-1} by applying a coarse-graining operator, which acts as a scaling filter G eliminating high-wave-vector contributions and subsamples the field [43]:

$$(G\varphi_{j-1})(i) = \sum_{i'} \varphi_{j-1}(i')G(2i - i'). \tag{1}$$

The site index i on the left-hand side runs on the coarser lattice, whereas i' on the finer one. More detailed properties of G, together with a brief introduction to wavelet theory, are described in Appendix A. In order to guarantee that the fluctuations of φ_j remain orderf one, a standard RG procedure normalizes G with a factor γ_j which guarantees that $\langle|\varphi_j(i)|^2\rangle_{p_j} - \langle\varphi_j(i)\rangle^2_{p_j} = 1$:

$$\varphi_j = \gamma_j^{-1} G\varphi_{j-1}.$$

A different normalization condition will be chosen for the WCRG later. The Kadanoff block-averaging scheme is a particular example of the coarse-graining procedure in Eq. (1). In the $d = 1$ case, it sums pairs of consecutive samples:

$$\varphi_j(i) = \frac{\varphi_{j-1}(2i) + \varphi_{j-1}(2i-1)}{\sqrt{2}\gamma_j}. \tag{2}$$

In the renormalization-group procedure, one computes scale by scale the energy function E_j of the marginal probability distribution restricted to the coarse-grained field φ_j, as given by

$$p_j(\varphi_j) = \frac{1}{Z_j} e^{-E_j(\varphi_j)} \quad \text{with} \quad E_j(0) \coloneqq 0, \tag{3}$$

where $Z_j = \int e^{-E_j(\varphi_j)} d\varphi_j$. RG proceeds from fine (microscopic) to coarse (macroscopic) scales by calculating E_j from E_{j-1}. The marginal probability distribution $p_j(\varphi_j)$ is computed by integrating $p_{j-1}(\varphi_{j-1})$ along the degrees of freedom of φ_{j-1} that are not in φ_j [15]. These fast degrees of freedom correspond to high-wave-vector fluctuations, which can be computed by using Fourier or wavelet bases. Wavelet bases provide localized representations of large classes of energy functions, which leads to an efficient calculation of marginal integrations, as we discuss below.

2. Fast wavelet transform

We now introduce the orthogonal decomposition, which allows us to represent a microscopic field in terms of its coarse-grained version and fast degrees-of-freedom contributions. This is done with the fast wavelet transform algorithm introduced in Ref. [43]. It computes an orthogonal representation with a cascade of filtering and subsampling illustrated in Fig. 2, which separates the field fluctuations at different scales. For completeness we recall below and in the Appendix A some properties of wavelet theory.

The degrees of freedom of φ_{j-1} that are not in φ_j are encoded in orthogonal wavelet field $\bar{\varphi}_j$. The representation $(\varphi_j, \bar{\varphi}_j)$ is an orthogonal change of basis calculated from φ_{j-1}. The coarse-grained field φ_j is calculated in Eq. (1) with a low-pass scaling filter G and a subsampling. In spatial dimension d, the wavelet field $\bar{\varphi}_j$ has $2^d - 1$ subfields (or channels) computed with a convolution and subsampling operator $\bar{G}$. By including the normalization factor γ_j, we get

$$\varphi_j = \gamma_j^{-1} G\varphi_{j-1} \quad \text{and} \quad \bar{\varphi}_j = \gamma_j^{-1}\bar{G}\varphi_{j-1}. \tag{4}$$

In spatial dimension d, we define a separable low-wave-number filter G in Eq. (A7), which computes a coarse-grained field $\varphi_j(i)$ which has $(L2^{-j})^d$ sites. Instead, the wavelet filter $\bar{G}$ computes $2^d - 1$ wavelet subfields $\bar{\varphi}_j(m, i)$ indexed by $1 \le m \le 2^d - 1$, with separable high-pass filters $\bar{G}_m$ specified in Eq. (A8):

$$\bar{\varphi}_j(m, i) = \gamma_j^{-1} \sum_{i'} \varphi_{j-1}(i')\bar{G}_m(2i - i').$$

For each m, $\bar{\varphi}_j(m, i)$ carries the fast fluctuations of φ_{j-1} along a particular spatial orientation. Cascading Eq. (4) for $1 \le j \le J$ computes the decomposition of the microscopic field φ_0 into its orthogonal wavelet representation over J scales denoted as

$$\{\varphi_J, \bar{\varphi}_j\}_{1\le j\le J}. \tag{5}$$

In $d = 1$, there is a single wavelet field in $\bar{\varphi}_j$. The Kadanoff scheme is computed with a block-averaging filter G in Eq. (2). The corresponding wavelet filter $\bar{G}$ in Eq. (A4) computes the wavelet field with normalized increments:

$$\bar{\varphi}_j(i) = \frac{\varphi_{j-1}(2i-1) - \varphi_{j-1}(2i)}{\sqrt{2}\gamma_j}. \tag{6}$$

If $d = 2$, then there are three channels, as illustrated in Fig. 2.

The wavelet orthonormal filters G and $\bar{G}$ define a unitary transformation, which satisfies

$$\bar{G}G = G\bar{G} = 0 \quad \text{and} \quad G^T G + \bar{G}^T \bar{G} = \mathrm{Id}, \tag{7}$$

where Id is the identity. Appendix A 1 gives in Eqs. (A1) and (A3) a condition on the Fourier transform on G and on $\bar{G}$ to build such filters. The filtering equations in Eq. (4) can be inverted with the adjoint operators G^T and $\bar{G}^T$ as given by

$$\varphi_{j-1} = \gamma_j G^T \varphi_j + \gamma_j \bar{G}^T \bar{\varphi}_j. \tag{8}$$

The adjoint G^T enlarges the field size φ_j by inserting a zero between coefficients and filters the output as

$$(G^T \varphi_j)(i) = \sum_{i'} \varphi_j(i') G(2i' - i).$$

The adjoint $\bar{G}^T$ performs the same operations over the $2^d - 1$ channels and adds them,

$$(\bar{G}^T \bar{\varphi}_j)(i) = \sum_{m=1}^{2^d-1} \sum_{i'} \bar{\varphi}_j(m, i') \bar{G}_m(2i' - i).$$

The fast inverse wavelet transform [43] reconstructs φ_0 from its wavelet representation in Eq. (5) by progressively recovering φ_{j-1} from φ_j and $\bar{\varphi}_j$ with Eq. (8), for j going from J to 1.

The fast wavelet transform computes each $\bar{\varphi}_j$ from φ_0 by applying $j - 1$ times the convolutional operator G and then applying $\bar{G}$. This cascade of convolutional operators defines a single convolutional operator with an equivalent filter which is called a wavelet. When j increases, the wavelet theory [44,45] reviewed in Appendix A 2 proves that wavelet filters converges to wavelet functions of a continuous spatial variable $x \in \mathbb{R}^d$. Wavelet functions belong to the space $\mathbf{L^2}(\mathbb{R}^d)$ of square-integrable functions, $\int |f(x)|^2 dx < \infty$. Dilating and translating these $2^d - 1$ wavelet functions defines a wavelet orthonormal basis of $\mathbf{L^2}(\mathbb{R}^d)$. Wavelet fields $\bar{\varphi}_j$ can be rewritten in Eq. (A14) as decomposition coefficients in this wavelet orthonormal basis. The wavelet basis plays an important role in understanding the properties of wavelet fields, depending upon the representation of operators involved in energy functions. Actually, Wilson, in his seminal paper [25], computes such a wavelet decomposition by using a specific wavelet that is now called the Shannon wavelet. The Kadanoff RG scheme instead can be interpreted as a decomposition in a Haar wavelet basis. Figure 3 shows three representative wavelets that we discuss in this paper, Haar [Figs. 3(a) and 3(b)], Shannon [Figs. 3(c) and 3(d)], and Daubechies [Figs. 3(e) and 3(f)]. We see that the results are improved by using wavelets having a better localization both in the spatial and Fourier domains, as a Daubechies wavelet [46] in Figs. 3(e) and 3(f).

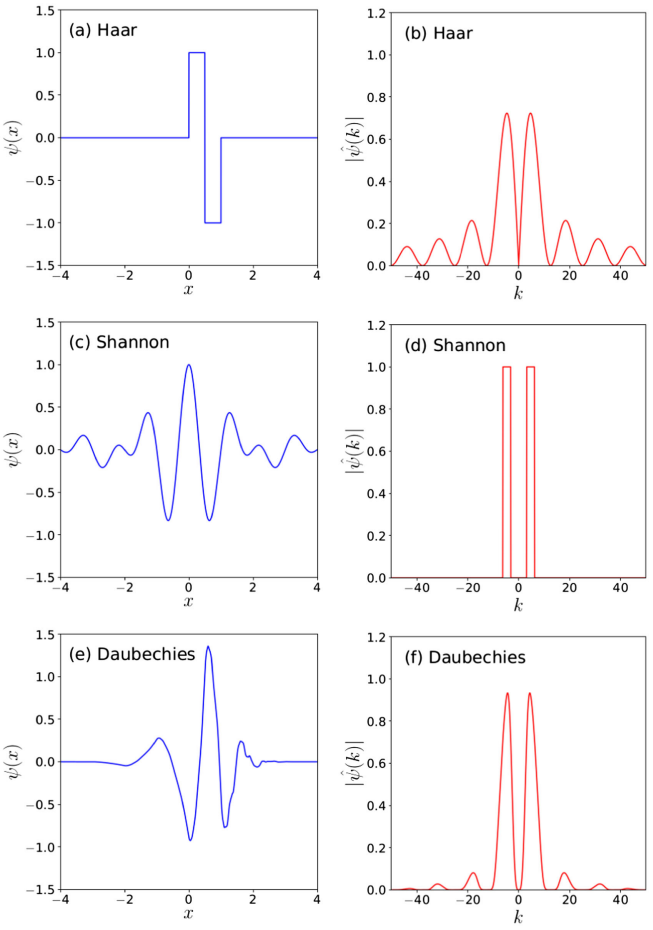

FIG. 3. The left graphs show one-dimensional wavelets $\psi(x)$ in real space, and the right graphs give their Fourier-transform amplitude $|\hat{\psi}(k)|$. (a,)(b) Haar wavelet. (c),(d): Shannon wavelet. (e),(f): Daubechies wavelet with $q = 4$ vanishing moments [46].

For later convenience, we define a projection operator $P = G^T G$ and its complement $\bar{P} = \bar{G}^T \bar{G}$, where $P + \bar{P} = \mathrm{Id}$ because of Eq. (7). P and $\bar{P}$ play a role in projecting a field φ_{j-1} onto the low- and high-wave-vector components, respectively,

$$P\varphi_{j-1} = \gamma_j G^T \varphi_j \quad \text{and} \quad \bar{P}\varphi_{j-1} = \gamma_j \bar{G}^T \bar{\varphi}_j.$$

3. Wavelet renormalization

The standard RG scheme begins from a microscopic energy function E_0, and iteratively computes E_j from E_{j-1}. To compute $p_j(\varphi_j) = Z_j^{-1} e^{-E_j(\varphi_j)}$ from $p_{j-1}(\varphi_{j-1}) = Z_{j-1}^{-1} e^{-E_{j-1}(\varphi_{j-1})}$, one can represent φ_{j-1} as orthogonal wavelet coordinates $(\varphi_j, \bar{\varphi}_j)$ by Eq. (8), and perform a marginal integration along the wavelet field $\bar{\varphi}_j$ as

$$E_j(\varphi_j) = -\log \int e^{-E_{j-1}(\varphi_{j-1})} d\bar{\varphi}_j + c_j, \tag{9}$$

where c_j is a constant chosen such that $E_j(0) = 0$, thus, $c_j = \log \int e^{-E_{j-1}(\bar{P}\varphi_{j-1})} d\bar{\varphi}_j$. Equation (9) plays the same role as a momentum shell integration in the standard RG, in which a similar marginal integration is calculated in a Fourier basis.

Large classes of energy functions involve differential operators such as Laplacians and gradients, which are diagonal in a Fourier basis. However, they often also involve local (pointwise) nonlinear operators as, e.g., in the φ^4 model, whose representations are delocalized on a Fourier basis. For appropriate wavelets, which are well localized in both real-space and Fourier domains, energy functions $E_{j-1}(\varphi_{j-1})$ depend upon local interactions of the wavelet fields $\bar{\varphi}_\ell$ for $\ell \geq j$. This is very useful both for numerical and analytical approaches, as it provides a compact short-range representation of the theory both in real and Fourier space. In his original calculation, Wilson used such "diagonal" approximations in a Shannon wavelet basis to analyze the properties of the φ^4 model at the phase transition [25]. Indeed, if E_{j-1} can be approximated by a sum over sites i of terms that are functions of the local value of the wavelet field $\bar{\varphi}_j(m, i)$, then the multidimensional marginal integral in Eq. (9) becomes a product of one-dimensional integrals, and the RG flow can be easily analyzed. Yet, the Shannon wavelet is not well adapted to the φ^4 energy function because it has a slow spatial decay in real space, as shown in Fig. 3(c).

We stress that there are several studies of RG in wavelet bases [26,27], but they remain on a formal level or are focused on $d = d_U - \epsilon$ perturbation theory. Our interest is instead developing a nonperturbative method both for forward RG and estimation by the WCRG.

4. The φ^4 model

This section introduces a discrete lattice version of the φ^4-field theory as a specific model, which provides a concrete example of energy functions. The φ^4-field theory played a central role in the theory of critical phenomena [10], because it captures the essential properties of standard second-order phase transitions. Its microscopic energy on a lattice reads

$$E_0(\varphi_0) = -\frac{\beta}{2}\varphi_0^T \Delta \varphi_0 + \sum_i (\varphi_0^4(i) - (1+2\beta)\varphi_0^2(i)), \tag{10}$$

where Δ is a discrete Laplacian,

$$-\varphi_0^T \Delta \varphi_0 = \sum_i \sum_{i' \in \mathcal{N}(i)} \frac{1}{2}(\varphi_0(i) - \varphi_0(i'))^2, \tag{11}$$

where $\mathcal{N}(i)$ is a set of the nearest-neighbor sites of i. It contains $2d$ neighbor sites in the d-dimensional cubic lattice. The first term in Eq. (10) disfavors spatial fluctuations. The second term is an even local potential with a double-well shape that favors two opposite values of the field (corresponding to the two wells). The parameter β specifies the relative magnitude between the two terms. For $\beta \ll 1$, fluctuations are entropically favored, and hence, the system is in a disordered phase. For $\beta \gg 1$, the two terms compete and lead to a ferromagnetic phase. The φ^4 model in Eq. (10) undergoes a phase transition in the thermodynamics limit for $\beta = \beta_c \simeq 0.68$ [47]. For $\beta > \beta_c$ the system is in the ordered broken-symmetry phase, whereas for $\beta < \beta_c$ the system is in a disordered phase.

When expressed on a Fourier basis, the Laplacian operator in the energy function in Eq. (10) becomes diagonal, but the $\varphi_0^4(i)$ local potential term instead produces global interactions between all wave vectors in the Fourier domain. On the other hand, appropriate wavelet bases (see below) provide nearly diagonal representations for the Laplacian operators [45] and the $\varphi_0^4(i)$ term because wavelets are well localized in both spatial and Fourier domains. The nearly diagonal representations by such wavelet bases are proven in Appendix A 3.

Kadanoff and Wilson's RG approaches can be interpreted as different choices of wavelet basis to represent high-wave-number fluctuations. The Shannon wavelet shown in Figs. 3(c) and 3(d) is well localized in the Fourier domain, but it has a slow spatial decay in real space. It thus does not provide good local approximations of pointwise polynomial nonlinearities in the φ^4 model. Instead, Kadanoff block averaging [40] is equivalent to decomposition in a wavelet basis generated by a Haar wavelet presented in Figs. 3(a) and 3(b). It has a narrow compact support in real space, but it is discontinuous and therefore extended in the Fourier space. As a consequence, the Haar wavelet is unable to nearly diagonalize the Laplacian term. The most important consequence is that the Kadanoff block-averaging RG is not able to describe the Gaussian fixed point (associated with pure Laplacian energy), and hence, it is incorrect in high spatial dimensions and unable to capture the existence of the upper critical dimension. Appendix A 2 introduces Daubechies wavelets [46] which have a compact support wider than the Haar wavelet but are well localized in Fourier space. They provide nearly diagonalized differential operators such as a Laplacian as well as pointwise nonlinearities of local potentials. Figures 3(e) and 3(f) show the Daubechies wavelet used in our numerical calculations.

Besides RG, turbulence is another physical case in which a multiscale representation using regular intervals in logarithmic space (in Fourier) plays a very important role. Shell models of turbulence, which allows to study turbulent cascades, are based on this kind of representation [48,49].

B. Energy *Ansatz*

The RG coarse-graining in Eq. (9) cannot be performed exactly in general. A strategy developed both for real space [40] and nonperturbative RG [15] is to approximate each E_j as a linear combination of a few terms, each one determined by coupling parameters and basis functions (or operators) of the field. We proceed in a similar way. In the following, we review the energy *Ansatz* that we use as approximation models. We write such an *Ansatz* in terms of coupling parameters and basis functions, so that one can express the wavelet RG flow equation as an evolution of the coupling parameters.

1. Local potentials

The energy *Ansatz*, i.e., the approximate form of the energy function for $E_j(\varphi_j)$, is defined at each scale 2^j as the sum of bilinear terms characterized by a two-point symmetric positive coupling matrix K_j and a nonlinear local potential composed of a vector of basis functions V and the associated coupling parameters C_j:

$$E_j(\varphi_j) = \frac{1}{2}\varphi_j^T K_j \varphi_j + C_j^T V(\varphi_j), \tag{12}$$

where $V(\varphi)$ is given by

$$V(\varphi) = (V_1(\varphi), V_2(\varphi), \ldots, V_s(\varphi)), \tag{13}$$

which do not depend on j.

Local potentials are defined as the sum of independent contributions from local field values. For each n $(1 \le n \le s)$, they are given by

$$V_n(\varphi_j) = \sum_i v_n(\varphi_j(i)), \tag{14}$$

where v_n is, by definition, the local potential for each site i, which is described below. Thus, the nonlinear potential in Eq. (12) is written by

$$\begin{aligned} C_j^T V(\varphi_j) &= \sum_n C_{j,n} V_n(\varphi_j) = \sum_i \sum_n C_{j,n} v_n(\varphi_j(i)) \\ &= \sum_i C_j^T v(\varphi_j(i)). \end{aligned} \tag{15}$$

The microscopic energy E_0 of the φ^4 model in Eq. (10) has such a local potential, and thus it can be written as the form in Eq. (12). The energy *Ansatz* for the φ^4 model assumes that the expression in Eq. (12) holds at all scales. Numerical applications in this paper are based on local potentials, but this is not a necessary condition for the theory and algorithms that are introduced.

We may further impose that two-point interactions have a finite range; i.e., they are defined over neighborhoods $\mathcal{N}(i)$ of constant size s', which does not depend upon the scale 2^j:

$$\varphi_j^T K_j \varphi_j = \sum_i \sum_{i' \in \mathcal{N}(i)} K_j(i,i')\varphi_j(i)\varphi_j(i').$$

The energy *Ansatz* in Eq. (12) can be rewritten as an inner product

$$E_j(\varphi_j) = \theta_j^T U_j(\varphi_j) \tag{16}$$

between a vector of coupling parameters

$$\theta_j = \left(\frac{1}{2}K_j, C_j\right) \tag{17}$$

of dimension $s + s'$, and the vector of the basis functions

$$U_j(\varphi_j) = \left(\varphi_j\varphi_j^T, V(\varphi_j)\right), \tag{18}$$

where $\varphi_j\varphi_j^T = \{\varphi_j(i)\varphi_j(i')\}_{i,i' \in \mathcal{N}(i)}$ is a band matrix of width s'. For stationary fields, K_j is a convolution kernel whose support size is s', and the matrix $\varphi_j\varphi_j^T$ can be replaced by a translation-invariant vector $\{\sum_i \varphi_j(i)\varphi_j(i-i')\}_{|i-i'| \le s'/2}$. This is the case for all applications in this paper, but we keep the notation $\varphi_j\varphi_j^T$ for simplicity. Having an energy *Ansatz* whose functional form is the same at all scales allows one to capture the self-similarity emerging at phase transitions. One expects that the critical point is described by a scale-invariant theory; i.e., the coupling vectors θ_j do not depend upon the scale 2^j for $1 \ll j \ll J$. Henceforth, we call "local" an *Ansatz* such as Eq. (18), as it is the sum over the sites i of functions of the field evaluated in sites at a finite bounded (local) distance from i.

We now describe the local potential v_n. In the theory of phase transitions, v_n are usually chosen to be polynomials, e.g., quadratic and quartic terms, because it leads to easier analytic computations. However, the fast growth of high-degree polynomials may lead to numerical instabilities. As a consequence, in the following, we employ a piecewise linear approximation of $C_j^T v(\varphi_j(i))$. This is done by considering a basis of the "hat" functions $v_n(x) = h(x - na)$, which are uniform translations of a linear box spline $h(x)$, which is a linear combination of three linear rectifiers:

$$h(x) = \max\{x + a, 0\} - 2\max\{x, 0\} + \max\{x - a, 0\},$$

where $2a$ is the width of the hat function. Figure 4 shows an example of the piecewise linear approximation described above with $a = 1$ for a quadratic function by hat functions.

We guarantee that our approximation of the probability distribution at scale 2^j is normalizable by imposing that the symmetric matrix K_j is strictly positive. The other coupling vector C_j does not affect this convergence because $C_j^T V(\varphi_j)$ has at best a linear growth at infinity by

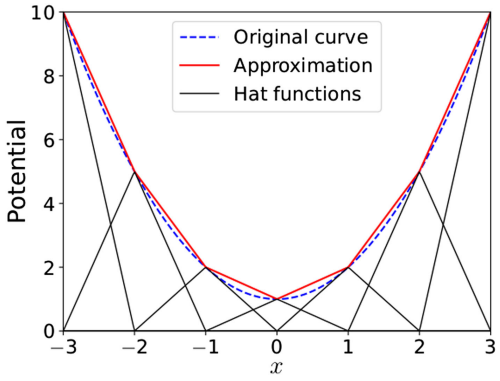

FIG. 4. A quadratic function (original curve, dashed) is represented by a piecewise linear approximation (red) given by a linear combination of hat functions (black).

construction due to the linear growth of the hat function. Note that we do not aim to reproduce the true growth of the potential at infinity, as this is completely determined by rare events and out of the scope of the present work. The choice of having a linear growth of $C_j^T V(\varphi_j)$ is dictated by the requirement of having a stable algorithm in which normalization of the probability distribution is guaranteed by construction. For stationary processes, K_j is a translation-invariant Toeplitz matrix that is diagonalized in the Fourier basis. Normalizability amounts to imposing that its spectrum is strictly positive.

2. RG flow equation

The coupling parameters of energy functions at successive scales are related by an equation derived from the marginal integration in Eq. (9). Inserting the *Ansatz* in Eq. (16) for j and $j-1$ in the RG equation (9) gives

$$-\log \int e^{-\theta_{j-1}^T U_{j-1}(\varphi_{j-1})} d\bar{\varphi}_j + c_j \approx \theta_j^T U_j(\varphi_j). \quad (19)$$

In practice, this is not an equality since the marginal integration produces extra terms with respect to the energy *Ansatz* in Eq. (16). In order to establish an approximate RG treatment, one has to define a procedure to obtain θ_j from the left-hand side of Eq. (19). Several ways have been introduced in the literature: truncations such as in the Migdal-Kadanoff real-space RG [40] and projections as done in nonperturbative RG [15]. Another possibility, that we use later in Sec. V B, is to linearly regress θ_j over φ_j. This also allows us to estimate the approximation error as well.

Note that approximating the RG equation does not lead to instabilities (or divergencies) at the critical point, contrary to the case of perturbation theory directly done on the microscopic scale. The reason is that the integration in Eq. (19) is performed over $\bar{\varphi}_j$ which impacts only the interaction terms between high and low wave vectors and within high wave vectors. See Ref. [15] for a general discussion of the stability of the RG procedure.

In this paper, the usual (forward) RG is introduced only as a preamble to the wavelet-conditional (inverse) RG. Since we do not need to develop an approximate wavelet RG scheme for our purposes, this issue will be analyzed in a future publication.

III. WAVELET-CONDITIONAL RG

A wavelet-conditional RG estimates an efficient representation of the probability distribution of the microscopic field from several examples measured in experiments or obtained by simulations. The aim of the WCRG is to obtain from the data (i) a model of the probability distribution that allows us to generate new samples, (ii) a generative method that is fast even in the presence of long-range correlations, e.g., at criticality, and (iii) an estimate of the underlying microscopic energy function E_0.

A. Conditional probability representation

The conditional-wavelet RG defines a representation of the microscopic energy E_0 into multiscale interaction energies, which represent interactions between high and low wave vectors at each scale. We introduce this new representation below.

Since $(\varphi_j, \bar{\varphi}_j)$ is an orthogonal representation of φ_{j-1}, computing $p_{j-1}(\varphi_{j-1})$ from $p_j(\varphi_j)$ amounts to computing

$$p_{j-1}(\varphi_{j-1}) = \alpha_j \bar{p}_j(\bar{\varphi}_j|\varphi_j) p_j(\varphi_j), \quad (20)$$

where $\alpha_j = \gamma_j^{(L/2^{j-1})^d}$ is the Jacobian factor resulting from the normalization in Eq. (8). This procedure can be iterated to express $p_j(\varphi_j)$ as a telescopic product:

$$p_j(\varphi_j) = p_J(\varphi_J) \prod_{l=j+1}^{J} \alpha_l \bar{p}_l(\bar{\varphi}_l|\varphi_l). \quad (21)$$

At the microscopic scale ($j=0$), we can express $p_0(\varphi_0)$ as

$$p_0(\varphi_0) = \alpha p_J(\varphi_J) \prod_{j=1}^{J} \bar{p}_j(\bar{\varphi}_j|\varphi_j). \quad (22)$$

Thus, the probability distribution p_0 can be represented in terms of all the conditional probabilities $\bar{p}_j$ and the coarsest probability p_J. Each conditional probability distribution can be exactly rewritten as

$$\bar{p}_j(\bar{\varphi}_j|\varphi_j) = \frac{1}{\bar{Z}_j} \exp\Big(-\bar{E}_j(\varphi_{j-1}) + \bar{F}_j(\varphi_j)\Big), \quad (23)$$

where $\bar{E}_j$ and $\bar{F}_j$ are the energy associated with scale interactions between $\bar{\varphi}_j$ and φ_j, and the free energy,

respectively. More formally, the scale interaction energy $\bar{E}_j(\varphi_{j-1})$ is constructed such that $\bar{E}_j(P\varphi_{j-1}) = \bar{E}_j(\varphi_{j-1})|_{\bar{\varphi}_j=0} = 0$ for all φ_{j-1}. Then the free energy $\bar{F}_j$ in Eq. (23) is defined as

$$\bar{F}_j(\varphi_j) = -\log \int e^{-\bar{E}_j(\varphi_{j-1})} d\bar{\varphi}_j + c'_j, \tag{24}$$

where c'_j is a constant chosen such that $\bar{F}_j(0) = 0$. Hence, $c'_j = \log \int e^{-\bar{E}_j(\bar{P}\varphi_{j-1})} d\bar{\varphi}_j$. The normalization constant in Eq. (23) is defined as

$$\bar{Z}_j = \int e^{-\bar{E}_j(\bar{P}\varphi_{j-1})} d\bar{\varphi}_j.$$

It results from Eq. (20) that

$$E_{j-1} = E_j + \bar{E}_j - \bar{F}_j. \tag{25}$$

Using the representation in Eq. (22) of p_0, and cascading by Eq. (25), one finds

$$E_0 = E_J + \sum_{j=1}^{J} (\bar{E}_j - \bar{F}_j). \tag{26}$$

These results show that instead of working with p_0 directly, one can work with its alternative representation in Eq. (22). Sampling a field φ_0 can thus be achieved by successive sampling of $\bar{\varphi}_j$ using the multiscale interaction energy functions,

$$\{E_J, \bar{E}_j\}_{1\le j\le J}. \tag{27}$$

Those energy functions specify completely E_0, since each free energy $\bar{F}_j$ is calculated from $\bar{E}_j$ with Eq. (24). In conclusion, the multiscale interaction energy representation in Eq. (27) offers an alternative way to sample and to define the microscopic probability distribution p_0. It is a way to represent probability distributions that describe fluctuations progressively from coarse to fine scales.

We show above how to obtain all the energy functions E_j from the multiscale interaction energies $\{E_J, \bar{E}_j\}_{1\le j\le J}$. One can also obtain the inverse relationship that constructs the scale interaction energies and free energies from E_j. By subtracting Eq. (25) from its counterpart evaluated for $\bar{\varphi}_j = 0$, one obtains

$$\bar{E}_j(\varphi_{j-1}) = E_{j-1}(\varphi_{j-1}) - E_{j-1}(P\varphi_{j-1}). \tag{28}$$

Evaluating Eq. (25) for $\bar{\varphi}_j = 0$ gives

$$\bar{F}_j(\varphi_j) = E_j(\varphi_j) - E_{j-1}(P\varphi_{j-1}). \tag{29}$$

In a standard RG, the fields φ_j are normalized with a multiplicative factor γ_j in Eq. (4), so that at second-order phase transitions, the probability distribution is scale invariant under RG transformations. It results that $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ is also scale invariant. In order to find a fixed point, one has to impose a normalization condition for the fields. In our case, we choose the normalization condition such that

$$\sum_{m=1}^{2^d-1} \langle |\bar{\varphi}_j(m,i)|^2 \rangle_{p_{j-1}} = 1. \tag{30}$$

This normalization condition has an important role even when the system is not at a critical point. It leads to coupling parameters associated with scale interaction energies of order one at all scales.

B. Conditional couplings parameters

We now introduce a model (or *Ansatz*) for conditional probabilities $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ with associated coupling parameters. This is at the core of our parametric representation of the conditional probabilities. In the next section, we show that the estimation of such conditional coupling parameters remains stable at phase transitions, as opposed to coupling parameters θ_j of $p_j(\varphi_j)$. Moreover, we see that it defines richer multiscale models of the microscopic probability distribution $p_0(\varphi_0)$. This is crucial for systems with long-range interactions such as cosmological fields.

1. Scale interactions

For a Gaussian field, the conditional probability $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ contains a quadratic interaction between $\bar{\varphi}_j$ and φ_j. However, this is not the case for non-Gaussian processes, which often have nonlinear dependences across scales. To simplify notations, we write $\varphi_J = \bar{\varphi}_{J+1}$. Since φ_j is specified by its orthogonal wavelet coefficients $\{\bar{\varphi}_\ell\}_{j<\ell\le J+1}$, one finds

$$\bar{p}_j(\bar{\varphi}_j|\varphi_j) = \bar{p}_j(\bar{\varphi}_j|\{\bar{\varphi}_\ell\}_{j<\ell\le J+1}).$$

Several studies have shown that such conditional dependences can indeed be represented with low-dimensional models taking advantage of the sparsity of wavelet fields [6]. For instance, wavelet scattering and phase harmonic models are conditioned by the correlation of the amplitude and the phase of $\bar{\varphi}_j$ relative to $\bar{\varphi}_\ell$ for $j < \ell \le J+1$. They have been shown to approximate well the distribution of complex physical fields such as fluid, gas turbulence, and cosmological matter density [35,50–52].

2. WCRG conditional coupling model

We define a conditional probability model by approximating the scale interaction energy $\bar{E}_j$ of $\bar{p}_j$ with a

conditional coupling vector $\bar{\theta}_j$ and a vector of basis functions $\bar{U}_j$ such that

$$\bar{E}_j(\varphi_{j-1}) = \bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) \quad \text{and} \quad \bar{U}_j(P\varphi_{j-1}) = 0. \tag{31}$$

The last requirement means that $\bar{U}_j(\varphi_{j-1})|_{\bar{\varphi}_j=0} = 0$. It guaranties that $\bar{E}_j(P\varphi_{j-1}) = 0$ and hence that the model taken for $\bar{E}_j$ contains the interactions between the wavelet field $\bar{\varphi}_j$ and the coarse-grained field φ_j, and not the interactions of φ_j with itself.

These models of scale interaction energies $\bar{E}_j$ define models of the microscopic energy E_0 through Eq. (26). Free energies are defined by

$$\bar{F}_j(\varphi_j) = -\log \int e^{-\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1})} d\bar{\varphi}_j + c_j'', \tag{32}$$

where c_j'' is chosen such that $\bar{F}_j(0) = 0$. According to Eq. (26), we get

$$E_0 = \theta_J^T U_J + \sum_{j=1}^{J} (\bar{\theta}_j^T \bar{U}_j - \bar{F}_j), \tag{33}$$

which proves that the microscopic energy function is entirely specified by the WCRG conditional coupling model given by

$$\{\theta_J, \bar{\theta}_j\}_{1\le j\le J}. \tag{34}$$

These models are more general than the energy models discussed in Sec. II B, because they may include different local potentials at all scales, as we describe in more detail below.

3. From local to multiscale potential models

Equation (28) implies that a model for the energy functions E_j translates into a model for the scale interaction energies with

$$\bar{U}_j(\varphi_{j-1}) = U_{j-1}(\varphi_{j-1}) - U_{j-1}(P\varphi_{j-1}) \tag{35}$$

and

$$\bar{\theta}_j^T \bar{U}_j = \theta_{j-1}^T \bar{U}_j. \tag{36}$$

The conditional coupling parameter $\bar{\theta}_j$ is uniquely defined as the orthogonal projection of the coupling vector θ_{j-1} of E_{j-1} over the space generated by the interaction potential vector $\bar{U}_j$.

Let us consider models of E_j with a single local potential at the scale 2^j defined by $U_j(\varphi_j) = (\varphi_j \varphi_j^T, V(\varphi_j))$. Appendix B 1 shows that the corresponding scale interaction energy model obtained from Eq. (35) can be written as

$$\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) = \frac{1}{2} \sum_{\ell=j}^{J+1} \bar{\varphi}_j^T \bar{K}_{j,\ell} \bar{\varphi}_\ell + \bar{C}_j^T \Big(V(\varphi_{j-1}) - V(P\varphi_{j-1}) \Big), \tag{37}$$

with a scale interaction potential

$$\bar{U}_j = \Big(\{\bar{\varphi}_\ell \bar{\varphi}_j^T\}_{j\le\ell\le J+1}, V(\varphi_{j-1}) - V(P\varphi_{j-1}) \Big) \tag{38}$$

and the associated coupling vector

$$\bar{\theta}_j = \Big(\frac{1}{2} \{\bar{K}_{j,\ell}\}_{j\le\ell\le J+1}, \bar{C}_j \Big). \tag{39}$$

Each $\bar{K}_{j,\ell}$ is a convolution operator and hence depends upon the wave-number overlap between wavelets at length scales 2^j and 2^ℓ. It is large for $\ell = j$, small for $\ell = j+1$, and negligible beyond if the wavelet is sufficiently well localized in the Fourier domain. We typically estimate only $\bar{K}_{j,\ell}$ for $\ell = j$ and $j+1$, and we set to zero all matrices for $\ell > j+1$ (see below).

Local scalar energy models E_0 have a single local potential at the finest scale 2^0, as in the φ^4-field theory. Section V A shows that their expansion in Eq. (33) corresponds to particular cases where local potential parameters $\bar{C}_j$ cancel with the free-energy terms $\bar{F}_j$ at all scales 2^j but the finest and coarsest ones, thus leading indeed to a single local potential defined on the microscopic scale. However, general scale interaction models in Eq. (37) may have arbitrary conditional coupling parameters $\bar{C}_j$ which do not disappear. The resulting microscopic energy E_0 then has a different local potential at each scale, which can capture long-range interactions.

C. Coarse to fine WCRG sampling

We now discuss the first important outcome of the WCRG: sampling independent fields or configurations in an efficient way, which fully circumvents the problem of critical slowing-down for second-order phase transitions and numerical instabilities associated with metastable states. In order to sample by the WCRG, one has to first determine the model of scale interaction energy functions by estimating the associated conditional coupling parameters in Eq. (31) from the training dataset. This can be done efficiently (again, in a way that is not affected by critical slowing-down in the presence of long-range correlations) and is explained in the next section. Once the value of the coupling parameters is learned, the WCRG representation of the probability distribution is known as explained above.

Sampling is then performed as follows. At a coarsest scale 2^J, which is of the order of the system size L, we draw a sample φ_J of the probability distribution whose energy is $E_J = \theta_J^T U_J$. This is done with an MCMC Metropolis

sampling algorithm and is fast because φ_J is reduced to a single point if $2^J = L$ or a few grid points if 2^J is of the order of L. The corresponding distribution is simple: It is a Gaussian in the disordered phase of second-order phase transition (if L is large enough). We then iteratively draw $\bar{\varphi}_j$ given φ_j by sampling with an MCMC Metropolis algorithm using the conditional probability distribution in Eq. (23). Note that we do not need to compute the free energy $\bar{F}_j$ because it is a normalization factor and thus not needed in the MCMC algorithm. As we explain and show in the next sections, this sampling is also fast. Indeed, the MCMC Metropolis algorithm used here has a decorrelation time that does not increase near phase transitions. It does not suffer from any critical slowing-down. In consequence, at scale 2^j, the number of operations needed to draw a sample of $\bar{\varphi}_j$ is just of the order of the number of degrees of freedom, i.e., $(2^d - 1)(L2^{-j})^d$. We update only the highest wave numbers through local interactions. Since $2^{Jd} = L^d$, by summing up all scales from J to 1, we find that the number of MC steps to sample a new configuration is proportional to $\sum_{j=1}^{J}(2^d - 1)(L2^{-j})^d = L^d(1 - 2^{-Jd}) = \mathcal{O}(L^d)$. The unit of time of an MC step is L^d. The resulting characteristic MCMC time to sample using the WCRG is therefore independent of the system size L, namely $\tau_{\mathrm{MC}} \sim \mathcal{O}(1)$, where τ_{MC} is an MCMC decorrelation (or mixing) timescale. This result is an important improvement relative to the state of the art. Indeed, standard MCMC simulations for models with long-range correlations, such as Ising or φ^4 models at criticality, require an MCMC time $\tau_{\mathrm{MC}} \sim L^z$, where z is a dynamic critical exponent ($z \simeq 2$ in two dimensions). More efficient cluster algorithms such as the Swendsen-Wang algorithm and the Wolff algorithm have an exponent $z \simeq 0.2$–0.3 [53,54]. The WCRG sampling is thus faster. It is done in "one shot" because the number of operations needed to draw a sample is of the order of the number of degrees of freedom—the best possible in terms of computational complexity since to draw L^d variables, one needs to do at least $\mathcal{O}(L^d)$ operations. It can be applied to generic systems, including nonequilibrium fields, as long as the conditional probability represents the underlying distribution well, whereas cluster algorithms are used only for simple systems such as the Ising model and its variants.

We note that various Monte Carlo algorithms and techniques have been developed in order to achieve faster sampling, such as the Hamilton Monte Carlo [55], the multigrid Monte Carlo [56], umbrella sampling [57], and parallel tempering [58] to name but a few. The Hamilton Monte Carlo [55] uses proposal updates along the direction of the gradient of the Hamiltonian of the system instead of random updates used in the standard MCMC. Although it leads to collective movements, which might be more efficient than the standard MCMC in some cases, the algorithm still remains within the paradigm of physical dynamics, and hence does not lead to drastic improvements in terms of efficiency. The multigrid Monte Carlo method [56], which is a stochastic version of the multigrid method in the field of numerical analysis, uses proposal updates that are performed in a scale-by-scale manner. They are not renormalization-group algorithms because they do not use the conditional probability distribution at each scale, but directly project the microscopic energy on coarse-grained fields [56]. Although the multigrid Monte Carlo can completely eliminate the critical slowing-down for the Gaussian model, it still produces a finite dynamic critical exponent for non-Gaussian systems. For example, the multigrid MC method for the φ^4-field model shows the same exponent as the standard MC ($z \simeq 2$). As we stress in the Introduction, all these methods work in a very different setting from ours: they start from a complete knowledge of the energy function but with no data. In our case instead, the dataset allows us to estimate the representation in terms of conditional probability. This is instrumental to obtain fast sampling.

IV. PRECONDITIONED ESTIMATION OF CONDITIONAL COUPLINGS AND THE ABSENCE OF CRITICAL SLOWING-DOWN

This section discusses the determination of the coupling parameters from the training dataset. When inferring coupling vectors from data using the maximum-likelihood estimation, there are two crucial difficulties that can make the method unstable and unfeasible. First, a dramatic slowing-down of the gradient decent because we must use an extremely small step size in order to avoid instabilities. Second, a large increase of the decorrelation (or mixing) time τ_{MC} of the MCMC, which must be computed at each gradient decent step. These two issues are actually facets of the same phenomenon originating from critical slowing-down.

Let us focus on the case of second-order phase transitions (the φ^4-field theory in particular) to explain intuitively what the problem is, which in this context is related to the phenomenon called critical slowing-down. Let us consider the setting in which the estimation is performed directly on the microscopic lattice, i.e., on the finest scale. We call it a direct coupling estimation method. For instance, imagine estimating an energy function formed by two terms corresponding to a local potential and a discrete Laplacian, respectively, as in Eq. (12).

Close to a critical point, a very tiny change in the coupling parameters induces a dramatic change in the spatial correlations of the fields. In consequence, in order to estimate an energy function that correctly reproduces the training data, the method has to be extremely precise, and hence a very small integration step in the gradient descent is required. At the critical point, the susceptibility to changes in the coupling parameters diverges, and hence the required step size vanishes, thus making the gradient descent

unfeasible. Moreover, for each gradient-descent step leading to the maximum likelihood, one has to perform an MCMC to evaluate the gradient term. If the estimated probability distribution is a faithful representation of the original one, as it should be if the method works, then the MCMC has a very long decorrelation time $\tau_{\rm MC}$ close to the second-order phase transition and diverges as a power law of L at the critical point, $\tau_{\rm MC} \sim \mathcal{O}(L^z)$ with $z > 0$.

In summary, the long-range correlations that emerge near the second-order phase transition make a direct coupling estimation of the probability distribution problematic. In the following, we first show in detail why standard algorithms to infer the coupling vectors suffer from a critical slowing-down near the critical point. We then present the theory which shows that the WCRG circumvents this problem and provides preconditioning of the maximum-likelihood gradient estimation and the MCMC, leading to much faster convergence, overcoming the critical slowing-down completely.

A. Direct coupling estimation by gradient descent

We estimate a model $p_{\theta_j} = Z_j^{-1} e^{-\theta_j^T U_j}$ of p_j by maximizing the likelihood $\langle \log p_{\theta_j} \rangle_{p_j}$ [59]. This is equivalent to minimizing the Kullback-Leibler (KL) divergence,

$$D_{\rm KL}(p_j \| p_{\theta_j}) = \int p_j(\varphi_j) \log p_j(\varphi_j) d\varphi_j - \langle \log p_{\theta_j} \rangle_{p_j},$$

which measures the information loss when approximating p_j by p_{θ_j}. The gradient of $\langle \log p_{\theta_j} \rangle_{p_j}$ or $-D_{\rm KL}(p_j \| p_{\theta_j})$ relative to θ_j is

$$\nabla_{\theta_j} \langle \log p_{\theta_j} \rangle_{p_j} = -\nabla_{\theta_j} D_{\rm KL}(p_j \| p_{\theta_j}) = \langle U_j \rangle_{p_{\theta_j}} - \langle U_j \rangle_{p_j},$$

where U_j is obtained from Eq. (18):

$$U_j(\varphi_j) = \Big(\varphi_j \varphi_j^T, \mu_j^{-1} V(\varphi_j) \Big), \tag{40}$$

where μ_j is a normalization factor that is determined below. The optimal $\theta_j^\star$, which satisfies $\langle U_j \rangle_{p_{\theta_j^\star}} = \langle U_j \rangle_{p_j}$, is searched by a gradient descent with a step size (or learning rate) ϵ,

$$\theta_j^{(t+1)} - \theta_j^{(t)} = \epsilon \Big(\langle U_j \rangle_{p_{\theta_j^{(t)}}} - \langle U_j \rangle_{p_j} \Big), \tag{41}$$

where t represents discrete time step. Since $D_{\rm KL}(p_j \| p_{\theta_j})$ is a convex function of θ_j, as it can be easily checked, the gradient descent is guaranteed to converge if ϵ is smaller than the inverse of the largest eigenvalue of the Hessian of $D_{\rm KL}(p_j \| p_{\theta_j})$ relative to θ_j. The first right-hand-side term $\langle U_j \rangle_{p_{\theta_j^{(t)}}}$ of Eq. (41) is computed from the energy $\theta_j^T U_j$ of $p_{\theta_j^{(t)}}$ with an MCMC Metropolis algorithm. The second term $\langle U_j \rangle_{p_j}$ is estimated with an empirical average of $U_j(\varphi_j)$ over the training dataset. Note that this coupling estimation by maximum likelihood or minimization of the KL divergence is equivalent to an application of the maximum entropy principle [7,60].

The above gradient-descent algorithm is simple but it converges extremely slowly for multiscale processes or near critical points. The convergence rate of the gradient descent depends upon the condition number κ of the Hessian H_{θ_j} of $D_{\rm KL}(p_j \| p_{\theta_j})$ relative to θ_j. If the eigenvalues of H_{θ_j} are between $\lambda_{\rm min}$ and $\lambda_{\rm max}$ then the gradient descent converges for $\epsilon \le \lambda_{\rm max}^{-1}$ with a rate $\kappa^{-1} = \lambda_{\rm min}/\lambda_{\rm max}$. The convergence is thus very slow if κ is large. A direct calculation shows that H_{θ_j} is equal to the covariance of U_j,

$$H_{\theta_j} = {\rm Cov}(U_j)_{p_{\theta_j}} = \langle U_j U_j^T \rangle_{p_{\theta_j}} - \langle U_j \rangle_{p_{\theta_j}} \langle U_j \rangle_{p_{\theta_j}}^T. \tag{42}$$

When p_{θ_j} becomes close to p_j, this Hessian becomes close to ${\rm Cov}(U_j)_{p_j}$, which must be well conditioned to have a fast convergence rate. At phase transitions, $\varphi_0(i)$ is a stationary field whose correlation length ξ grows to the system size L. As we explain below, this is the key reason for the bad conditioning of the Hessian. We determine the normalization factor μ_j in Eq. (40) such that ${\rm Cov}(\mu_j^{-1} V)_{p_j}$ is order one. To do so, we fix μ_j by

$$\mu_j^2 = {\rm TrCov}(V)_{p_j} = \sum_{n=1}^{s} \Big\{ \langle V_n^2 \rangle_{p_j} - \langle V_n \rangle_{p_j}^2 \Big\}. \tag{43}$$

The covariance of $U_j(\varphi_j) = [\varphi_j \varphi_j^T, \mu_j^{-1} V(\varphi_j)]$ is an operator formed by different connected correlation functions (covariances) of $\varphi_j \varphi_j^T$, $V(\varphi_j)$ and between them. Since the probability distribution of φ_0 is invariant by translation (it is stationary), all covariances are also be invariant by translation, and hence diagonal in Fourier. In order to understand intuitively why long-range correlations lead to very large maximum eigenvalues for these covariances, let us focus first on the covariance of the continuous field $\varphi(x)$ itself (this covariance does not appear in the Hessian, but it plays an essential role). The connection between continuous and discrete lattice fields $\varphi(x)$ and $\varphi_j(i)$ is summarized in Appendix A. Approaching the phase transition, this covariance displays a power-law behavior in real space as [61]

$$\langle \varphi(x) \varphi(x') \rangle - \langle \varphi(x) \rangle \langle \varphi(x') \rangle \sim \frac{1}{|x - x'|^{d-2+\eta}}$$

for $1 \ll |x - x'| \ll \xi$, where ξ is the correlation length. Accordingly, the eigenvalue (Fourier transform) of the covariance associated with the field φ, $\lambda_\varphi(k)$ has a power-law decay in Fourier space as $|k|^{-(2-\eta)}$ with

$2-\eta>0$ for $2\pi/\xi \ll |k|$. Approaching critical points, the correlation length ξ grows to the order of the system size L. This leads to a very large ratio between the minimum and maximum eigenvalue for the covariance of the field.

The Hessian does not only contain the covariance of the field but also more complicated covariances between $\varphi\varphi^T$ and $V(\varphi)$. However, the same reasoning holds. The long-range correlation of the field $\varphi(x)$ also induces a power-law decay for $1 \ll |x-x'| \ll \xi$ for these covariances, although with possibly different critical exponents [61]. As in the previous case, this behavior in real space leads to a power-law decay in Fourier space for $2\pi/\xi \ll |k|$. The blowup at small k produces a very large ratio $\kappa = \lambda_{\max}/\lambda_{\min}$ and a small $\lambda_{\max}^{-1}$, and thus a bad conditioning. In the Gaussian case, the computation can be done explicitly since the covariance of U_j is composed only of fourth-order moments that can be computed using Wick's theorem. This gives an explicit confirmation of the general argument explained above.

The emergence of long-range correlations also leads to a very large decorrelation time for the MCMC. This effect called critical slowing-down [13,14] is a consequence of spatial long-range correlations. An MCMC performing local updates has to break long-range correlated patterns in order to show decorrelation and hence reach ergodicity. This is further explained in Sec. IV C.

In summary, long-range correlations in fields make the direct coupling estimation by gradient descent perform very poorly. In many cases, the requirement of a very small learning rate and very large decorrelation time is so stringent to make estimation practically unfeasible.

B. Conditional coupling estimation

The conditional coupling estimation for the WCRG is performed scale by scale, unlike what is done when estimating the full-probability distribution directly in the previous section (direct coupling estimation). One starts from the coarsest scale 2^J of the order of the system size L, i.e., $2^J \approx L$. In this case, estimating the coupling parameters θ_J of an energy model $E_J = \theta_J^T U_J$ is easy and fast since φ_J is reduced to a single or few values. It amounts to performing a low-dimensional estimation from a large training dataset. We then estimate all conditional probabilities $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ needed to recover the probability distribution with Eq. (22).

For each given scale, the procedure is done as follows. First of all, the renormalization factor γ_j in Eq. (4) is adjusted according to Eq. (30). In order to do this, the variances of wavelet fields are estimated from R examples $\bar{\varphi}_{j,r}$ calculated from R examples $\varphi_{0,r}$ of φ_0 in the dataset. The parameter γ_j is adjusted to normalize these variances. As we discuss in the next section, this renormalization has an important role in making the method well conditioned.

We wish to estimate the conditional probability model

$$\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j) = \frac{1}{\bar{Z}_j} e^{-\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) + \bar{F}_j(\varphi_j)} \tag{44}$$

of $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$. Given $p_j(\varphi_j)$, this conditional probability model defines the following parametrized model of $p_{j-1}(\varphi_{j-1})$:

$$p_{\bar{\theta}_j}(\varphi_{j-1}) = \bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j) p_j(\varphi_j). \tag{45}$$

This model is optimized by maximizing the log-likelihood $\langle \log p_{\bar{\theta}_j} \rangle_{p_{j-1}}$, which is equivalent to minimizing the Kullback-Leibler divergence $D_{\mathrm{KL}}(p_{j-1} \| p_{\bar{\theta}_j})$. It is calculated with the same gradient-descent algorithm as in the previous section, but we now estimate only high-wave-vector conditional coupling parameters $\bar{\theta}_j$. As we show later, this procedure eliminates the ill-conditioning due to low wave vectors. The gradient of the log-likelihood or the KL divergence relative to $\bar{\theta}_j$ is

$$\begin{aligned} \nabla_{\bar{\theta}_j} \langle \log p_{\bar{\theta}_j} \rangle_{p_{j-1}} &= -\nabla_{\bar{\theta}_j} D_{\mathrm{KL}}(p_{j-1} \| p_{\bar{\theta}_j}) \\ &= \langle \bar{U}_j \rangle_{p_{\bar{\theta}_j}} - \langle \bar{U}_j \rangle_{p_{j-1}}. \end{aligned}$$

Therefore, the KL divergence gradient descent iteratively computes this gradient with a step size ϵ:

$$\bar{\theta}_j^{(t+1)} - \bar{\theta}_j^{(t)} = \epsilon \left(\langle \bar{U}_j \rangle_{p_{\bar{\theta}_j^{(t)}}} - \langle \bar{U}_j \rangle_{p_{j-1}} \right). \tag{46}$$

The second expected value $\langle \bar{U}_j \rangle_{p_{j-1}}$ is estimated with an empirical average of $\bar{U}_j(\varphi_{j-1})$ over examples of φ_{j-1} calculated from the training dataset. For $\bar{\theta}_j = \bar{\theta}_j^{(t)}$, since $p_{\bar{\theta}_j}(\varphi_{j-1}) = \bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j) p_j(\varphi_j)$, we compute samples $\varphi_{j,r}$ of p_j by coarse-graining the samples $\varphi_{0,r}$ of p_0 in the dataset. For each $\varphi_{j,r}$, a sample φ_{j-1} of $p_{\bar{\theta}_j}$ is calculated from a sample $\bar{\varphi}_j$ of $\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_{j,r})$, with

$$\varphi_{j-1} = \gamma_j G^T \varphi_{j,r} + \gamma_j \bar{G}^T \bar{\varphi}_j. \tag{47}$$

Samples $\bar{\varphi}_j$ are calculated with an MCMC Metropolis algorithm, which performs random local updates of $\bar{\varphi}_j$ according to the interaction energy $\bar{\theta}_j^T \bar{U}_j$ of $\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_{j,r})$. We estimate $\langle \bar{U}_j \rangle_{p_{\bar{\theta}_j}}$ by averaging $\bar{U}_j(\varphi_{j-1})$ over the φ_{j-1} in Eq. (47), obtained along all the MCMC chains which update $\bar{\varphi}_j$ for all $\varphi_{j,r}$. As it is shown in Sec. VI E on the φ^4 model, the MCMC on wavelet fields has a fast mixing time (decorrelation) which is not affected by critical slowing-down near phase transitions.

To accelerate computations, one can successively estimate the couplings from coarse to fine scales and initialize

$\bar{\theta}_j^{(t)}$ for $t=0$ with the final gradient-descent value $\bar{\theta}_{j+1}^{(\infty)}$ estimated at the previous scale 2^{j+1}.

The gradient descent in Eq. (46) converges if the gradient step ϵ is smaller than the inverse of the largest eigenvalues of the Hessians $H_{\bar{\theta}_j}$ of $D_{\mathrm{KL}}(p_{j-1}\|p_{\bar{\theta}_j})$ relative to $\bar{\theta}_j$. Since $p_{\bar{\theta}_j}(\varphi_{j-1})$ is factorized by Eq. (45) with the exponential conditional probability distribution in Eq. (44), one can verify that

$$H_{\bar{\theta}_j} = \left\langle \mathrm{Cov}(\bar{U}_j)_{\bar{p}_{\bar{\theta}_j}} \right\rangle_{p_j}, \tag{48}$$

where $\mathrm{Cov}(\bar{U}_j)_{\bar{p}_{\bar{\theta}_j}}$ is the covariance of $\bar{U}_j(\varphi_{j-1})$ relative to $\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j)$, given φ_j from $p_j(\varphi_j)$. The rate of convergence of the gradient descent in Eq. (41) is equal to the inverse of the condition number of $H_{\bar{\theta}_j}$. As we show below, this Hessian is well conditioned even at phase transitions, so the gradient descent does not suffer from critical slowing-down.

C. Preconditioning by renormalization

The estimation of coupling parameters and sampling a microscopic energy function suffer from a critical slowing-down near phase transitions. This section shows that there is no such critical slowing-down with the WCRG. The central result is a theorem for the Gaussian model which proves that although the covariance of the microscopic field φ_0 is badly conditioned at phase transitions, the covariance of wavelet fields $\bar{\varphi}_j$ remain uniformly well conditioned at all scales. We then argue, based on this result, that the WCRG is well conditioned and does not suffer from critical slowing-down. In the next section on numerical applications, we verify that this conclusion indeed holds numerically for the two-dimensional Gaussian and φ^4-field theories and cosmological data. Constructing a fully rigorous proof is an open mathematical challenge that we leave for future works.

1. Preconditioned gradient descent

As previously explained, the rate of convergence of the log-likelihood gradient descent is defined by the condition number of the Hessian in Eq. (48). It computes the covariance of $\bar{U}_j$ relative to $\bar{\varphi}_j$ conditioned by φ_j according to $\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j)$. This covariance is then averaged over φ_j according to p_j. The scale interaction potential $\bar{U}_j$ that we use is given by

$$\bar{U}_j = \left(\{\bar{\varphi}_\ell \bar{\varphi}_j^T\}_{j\le\ell\le j+1}, \nu_j^{-1}(V(\varphi_{j-1}) - V(P\varphi_{j-1})) \right), \tag{49}$$

where ν_j is an additional normalization factor. The quadratic interactions are reduced to neighboring scales because they are otherwise negligible, which amounts to replacing $J+1$ by $j+1$ in Eq. (38).

The Hessian in Eq. (48) includes the cross-covariance of the matrices $\bar{\varphi}_\ell \bar{\varphi}_j^T$ for $\ell = j, j+1$, the cross-covariance of $\bar{\varphi}_\ell \bar{\varphi}_j^T$ and $V(\varphi_{j-1})$, and the covariance of $V(\varphi_{j-1})$. One can verify that the subtraction of $V(P\varphi_{j-1})$ does not modify the Hessian. We fix the normalization constant ν_j similarly to Eq. (43) by setting

$$\nu_j^2 = \mathrm{TrCov}\Big(V(\varphi_{j-1}) - V(P\varphi_{j-1}) \Big)_{p_{j-1}}. \tag{50}$$

As already discussed in the direct coupling estimation case, since the probability distribution is translation invariant, the covariances in Eq. (48) are translation invariant and hence diagonal in Fourier space. To understand why they do not have power-law decay, we first focus on the covariance of $\bar{\varphi}_j$. We show that they are uniformly well conditioned at all scales 2^j, which plays a key role.

To understand the covariance properties of the wavelet fields $\bar{\varphi}_j(m,i)$, we relate wavelet fields defined over the discrete grid $i \in \mathbb{Z}^d$ to fields defined over continuous space $x \in \mathbb{R}^d$, decomposed over orthogonal wavelet bases. Appendix A 2 explains that $\bar{\varphi}_j$ are the decomposition coefficients of a field $\varphi(x)$ for $x \in \mathbb{R}^d$ in a wavelet orthonormal basis of the space of finite-energy functions of $x \in \mathbb{R}^d$. The microscopic field $\varphi_0(i)$ on a discrete lattice is a projection of $\varphi(x)$ at a length scale 1 (or $j=0$), which amounts to setting to zero all wavelet coefficients at scales $2^j < 1$, and hence eliminate the highest wave numbers to sample the field at unit intervals. The microscopic scale $2^j = 1$ plays the role of a cutoff scale. In order to analyze the wavelet field covariance at all scales 2^j, we consider the asymptotic properties of the continuous field $\varphi(x)$ over $\mathbb{R}^d$ (see Appendix A 2). The following theorem shows that the covariance of the wavelet field $\bar{\varphi}_j$ has a behavior very different from the one of φ. It states on a rigorous basis what it means for the wavelet field $\bar{\varphi}_j$ to be a fast degree of freedom.

Theorem IV.1 Let $\varphi(x)$ be stationary field over $x \in \mathbb{R}^d$, whose covariance has eigenvalues $\lambda_\varphi(k) = c|k|^{-\zeta}$ for $k \in \mathbb{R}^d$. If $\hat{G}(k) = \sqrt{2} + \mathcal{O}(|k|^q)$ for $q \ge \zeta/2$, then there exists $A > 0$ and B such that for all $j \in \mathbb{Z}$ and $k \in [-\pi, \pi]^d$,

$$A \le \lambda_{\bar{\varphi}_j}(k) \le B. \tag{51}$$

This theorem proves that the eigenvalues (i.e., the Fourier transform) $\lambda_{\bar{\varphi}_j}$ of the covariance of $\bar{\varphi}_j$ vary by a bounded factor at all scales 2^j, which means that in real space, the wavelet field has local fluctuations of order one with short-range correlation. There are two essential ingredients for the theorem to hold. The first one is to use wavelets which are sufficiently well localized in Fourier space; more precisely, the wavelets have a sufficiently large

number q of vanishing moments relative to the exponent ζ associated with the phase transition, with $\zeta = 2 - \eta$ using the standard notation of phase transition literature. Appendix A 2 explains that for this to hold, it is sufficient to impose that the wavelet filter G has a compact support and a Fourier transform which satisfies $\hat{G}(k) = \sqrt{2} + \mathcal{O}(|k|^q)$ with $q \geq \zeta/2$. The second important ingredient is that wavelet fields are normalized at each scale, i.e., $\bar{\varphi}_j$ satisfies Eq. (30). Conditional probabilities amount to slicing the estimation over multi-wave-number bands over which the scale interaction potentials have well-conditioned covariances. The wavelet field normalization has the flavor of a second-order method (Newton method), where the normalization of the wavelet field plays a similar role to an inverse Hessian to adjust each gradient step.

The theorem proof is in Appendix A 3. It is a particular case of a more general class of results concerning the representation of Calderón-Zygmund operators in wavelet bases [45]. Such operators include differential and pseudo-differential operators. This theorem considers a phase transition where the covariance of φ is a singular operator whose eigenvalues have a homogeneous and isotropic blowup at low wave numbers, as it happens at the critical point in second-order phase transitions [61]. For systems close to the critical point in which there is a finite but very large correlation length, a modification of the theorem proof leads to the same result in Eq. (51) with constants that do not depend upon ξ. This theorem result can be extended to a wider class of covariances by imposing only that $\lambda_\varphi(k)$ varies at most by a constant over any wave-number annulus $c_1 2^{-(j+1)} \leq |k| \leq c_2 2^{-j}$, where c_1 and c_2 are some constants.

To ensure a fast convergence of the gradient descent in Eq. (46), we must ensure the Hessian $H_{\bar{\theta}_j}$ in Eq. (48) is well conditioned. We give a qualitative argument which justifies this property, but not a formal proof. The Hessian is equal to the covariance of the scale interaction potential in Eq. (49). For a direct coupling estimation, the long-range spatial correlations of φ_0 lead to bad conditioning. The Hessian associated with the WCRG depends upon covariances relative to fluctuations of the wavelet fields $\bar{\varphi}_j$, conditioned and averaged over φ_j. Since the $\bar{\varphi}_j$ have a short-range correlation (as in Theorem IV.1) and the scale interaction energies $\bar{U}_j$ are local, the resulting covariances are also short ranged. Moreover, all terms in $\bar{U}_j$ have fluctuations of order one because we normalize the variance of $\bar{\varphi}_j$ and of the potential term in $\bar{U}_j$. Local interactions of terms whose fluctuations are of order one have covariance matrices whose eigenvalues remain order one. They do not blow up at small k and remain of order one for all k. The resulting Hessian is therefore well conditioned. These are the reasons why the WCRG is a well-conditioned method.

Turning these arguments into a rigorous proof is an open mathematical challenge that we leave for future works. The first difficulties are related to the nonlinear potential functions, although this seems within reach when expressed as polynomials because they transform local wavelets into localized functions. Another difficulty is to prove that averaging covariances conditioned by φ_j does not affect their condition number. It requires us to further decompose φ_j into normalized larger-scale wavelet fields $\bar{\varphi}_\ell$ for $\ell > j$ and evaluate cross-correlations with the terms in $\bar{\varphi}_j$. The proof of the purely Gaussian models relies on the fact that the homogeneous singular operators whose Fourier eigenvalues $\lambda_\varphi(k)$ have a power-law decay like $(|k|^\zeta + (2\pi/\xi)^\zeta)^{-1}$ are fully preconditioned by diagonal terms in a wavelet basis [45], not just scale per scale. In other words, the symmetric covariance matrix of all normalized wavelet fields $\{\bar{\varphi}_j\}_{1 \leq j \leq J+1}$, even when taking into accounts interactions across scales, has eigenvalues that remain of order one. In this Gaussian case, the Hessian $H_{\bar{\theta}_j}$ involves fourth-order moments which can be explicitly related to this full covariance matrix. The log-likelihood gradient descent has an exponential convergence with no critical slowing-down.

2. Preconditioned MCMC and Langevin dynamics

In the previous section, we show that if the Hessian $H_{\bar{\theta}_j}$ is well conditioned, then the gradient descent converges on times of order one. The condition on the Hessian requires analyzing a specific set of correlation functions. The requirements on correlation functions to guarantee a short decorrelation time of the MCMC performed on $\bar{\varphi}_j$ conditioned by φ_j are much more restrictive. As proven in Ref. [62], a necessary and sufficient condition for the decorrelation time of MCMC to be finite (not diverging with the system size) is that the so-called point-to-set length for the field $\bar{\varphi}_j$ conditioned by a typical φ_j is finite [62]. Physically, this requirement imposes that all possible spatial connected correlation functions [for the measure $\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j)$] have a short spatial range (and not just a specific set as for the Hessian). Only in this case, one can be sure that $\bar{\varphi}_j$ conditioned by a typical φ_j is noncritical with respect to any kind of ordering, and hence that the MCMC decorrelates quickly.

Although a general proof for the WCRG is a challenge, RG approaches to dynamical critical phenomena [63] show that this result holds for the φ^4 model, and more generally at critical points. These techniques are well studied in theoretical physics and considered to be fully under control, but they are not rigorous. In fact, it is at the core of the dynamical RG procedure that the short-scale degrees of freedom (the wavelet field in our case) are short ranged and relax quickly at each scale, hence the name fast degrees of freedom as shown by perturbative methods (in $d_U - d$, where d_U is the upper critical dimensions) [63] and by nonperturbative approximations [64]. As in the previous case, the Gaussian field theory provides a framework where

these results can be worked out in detail, even rigorously. In this case, the field $\bar{\varphi}_j$ is short ranged if all eigenvalues $\lambda_{\bar{\varphi}_j}$ of its covariance is of order one. Theorem IV.1 proves that this is true even for critical Gaussian fields such as the Ornstein-Uhlenbeck process and fractional Brownian motion. Therefore, even if the MCMC for the critical field has a decorrelation time τ_{MC} that diverges as L^z with $z = \zeta$, the MCMC for the field $\bar{\varphi}_j$ conditioned by a typical φ_j converges on times of order one.

One can directly obtain this result by considering Langevin dynamics for the field $\bar{\varphi}_j$. It is known physically and rigorously [65,66] that MCMC tends to Langevin dynamics in the continuum time limit. Let us then focus on the latter since the analysis is more straightforward. For simplicity, we focus on $d = 1$, where only one wavelet channel exists. The Langevin dynamics associated with the field $\bar{\varphi}_j$ is a multidimensional Ornstein-Uhlenbeck process:

$$\frac{\partial \bar{\varphi}_j(i,t)}{\partial t} = -\sum_{i'} \bar{K}_{j,j}(i-i')\bar{\varphi}_j(i',t) + \xi(i,t), \quad (52)$$

where $\xi(i,t)$ are independent and identically distributed Gaussian white noise of variance equals 2. The matrix $\bar{K}_{j,j}$ [see Eq. (B3)] is diagonal in Fourier space. As a consequence, the above set of Langevin equations decouple in Fourier space. For each Fourier component k, we obtain an independent Ornstein-Uhlenbeck process whose decorrelation time is the inverse of the eigenvalues of $\bar{K}_{j,j}$, which are equal to the eigenvalues $\lambda_{\bar{\varphi}_j}$ of the covariance of $\bar{\varphi}_j$, namely, $\hat{\bar{K}}^{-1}_{j,j}(k) = \lambda_{\bar{\varphi}_j}$ (where the hat denotes the Fourier transform). Theorem IV.1 proves that all of these eigenvalues are of order one and hence that the normalized Langevin process in Eq. (52) converges to the equilibrium stationary measure on times of order one.

In conclusion, the crucial condition that makes RG work is that fast degrees of freedom' are not critical, i.e., without long-range correlations in space and time. It is the same key requirement for the WCRG, thus strengthening the connection between the standard RG framework and WCRG.

V. RECOVERING THE MICROSCOPIC ENERGY

In the previous sections, we show how to estimate the multiscale representation of the probability distribution $p_0(\varphi_0)$ in Eq. (22), allowing us to perform sampling and obtain new data. In this section, we show how to recover the microscopic energy function $E_0(\varphi_0)$ with a well-conditioned method. This is a second important outcome of the WCRG, which can be useful for obtaining essential physical information, and which is out of reach of many other generative models. It is done by computing a functional representation of the scale interaction free energies at each scale.

A. Scale interaction free-energy models

In order to compute the microscopic energy in Eq. (33), we must construct a functional representation of scale interaction free energy $\bar{F}_j(\varphi_j)$ defined in Eq. (32). Modeling $\bar{F}_j(\varphi_j)$ amounts to defining another basis function $\tilde{U}_j(\varphi_j)$ and the associated coupling vector $\tilde{\theta}_j$ such that $\tilde{\theta}_j^T \tilde{U}_j$ provides an accurate approximation of $\bar{F}_j(\varphi_j)$ up to an additive constant:

$$\bar{F}_j(\varphi_j) = \tilde{\theta}_j^T \tilde{U}_j(\varphi_j) + \tilde{c}_j. \quad (53)$$

We show that it defines multiscale potential models of the microscopic energy function $E_0(\varphi_0)$, which are more general than local microscopic potential models.

Replacing each $\bar{F}_j$ by $\tilde{\theta}_j^T \tilde{U}_j + \tilde{c}_j$ in Eq. (33) gives

$$E_0 = \theta_J^T U_J + \sum_{j=1}^{J} \left(\bar{\theta}_j^T \bar{U}_j - \tilde{\theta}_j^T \tilde{U}_j \right) + c_0, \quad (54)$$

where c_0 is a constant. We consider the scale interaction energy model in Eq. (37),

$$\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) = \frac{1}{2} \sum_{\ell=j}^{J+1} \bar{\varphi}_j^T \bar{K}_{j,\ell} \bar{\varphi}_\ell + \bar{C}_j^T \Big(V(\varphi_{j-1}) - V(P\varphi_{j-1}) \Big). \quad (55)$$

The normalization constant μ_j introduced in Eq. (49) is absorbed into $\bar{C}_j$ for simplicity. The term $\bar{C}_j^T V(P\varphi_{j-1})$ is not modified by the free-energy integration in Eq. (32), because it does not depend on $\bar{\varphi}_j$. We thus define a model of $\bar{F}_j(\varphi_j)$ which includes this term

$$\tilde{\theta}_j{}^T \tilde{U}_j(\varphi_j) = \frac{1}{2} \varphi_j^T \tilde{K}_j \varphi_j + \tilde{C}_j^T V(\varphi_j) - \bar{C}_j^T V(P\varphi_{j-1}) \quad (56)$$

with $\tilde{\theta}_j = \left(\frac{1}{2}\tilde{K}_j, \tilde{C}_j \right)$. It results from the expansion in Eq. (54) that E_0 can be written by

$$E_0(\varphi_0) = \frac{1}{2} \varphi_0^T K_0 \varphi_0 + \sum_{j=0}^{J} C_j'^T V(\varphi_j), \quad (57)$$

where $C_0' = \bar{C}_1$, $C_J' = C_J - \tilde{C}_J$, and for $1 \le j \le J-1$,

$$C_j' = \bar{C}_{j+1} - \tilde{C}_j \quad (58)$$

with

$$\varphi_0^T K_0 \varphi_0 = \varphi_J^T K_J \varphi_J + \sum_{j=1}^{J} \left(\sum_{\ell=j}^{J+1} \bar{\varphi}_j^T \bar{K}_{j,\ell} \bar{\varphi}_\ell - \varphi_j^T \tilde{K}_j \varphi_j \right). \quad (59)$$

The multiscale coupling parameters C'_j for $1 \le j \le J$ give rise to long-range potential interactions. Local microscopic energy models correspond to a particular case where K_0 is a convolution operator with a small support, and $C'_j = 0$ for all $1 \le j \le J$, hence, $\tilde{C}_j = \bar{C}_{j+1}$, and $\tilde{C}_J = C_J$. Such models are sufficient to describe the φ^4-field theory. Instead, for cosmological weak-lensing fields, we see that $C'_j \neq 0$ for all $0 \le j \le J$. In this case, the microscopic energy includes local potentials at all scales, producing long-range interactions at the microscopic scale.

B. Free-energy regression

A coupling model $\tilde{\theta}_j^T \tilde{U}_j$ of the scale interaction free energy $\bar{F}_j$ is calculated by minimizing the mean-square error,

$$\ell(\tilde{\theta}_j) = \left\langle \left(\bar{F}_j - \tilde{\theta}_j^T \tilde{U}_j \right)^2 \right\rangle_{p_j}. \tag{60}$$

This section together with Appendix B 2 explains how to estimate the coupling vector $\tilde{\theta}_j$ from R samples $\varphi_{0,r}$ of p_0, with a thermodynamic integration [67]. The free-energy model is validated by verifying that the regression error is small: $\ell(\tilde{\theta}_j) \ll \langle \bar{F}_j^2 \rangle_{p_j}$.

The mean-square loss $\ell(\tilde{\theta}_j)$ is minimized for

$$\tilde{\theta}_j^\star = \langle \tilde{U}_j \tilde{U}_j^T \rangle_{p_j}^{-1} \langle \bar{F}_j \tilde{U}_j \rangle_{p_j}, \tag{61}$$

and the resulting error is

$$\ell(\tilde{\theta}_j^\star) = \langle \bar{F}_j^2 \rangle_{p_j} - \tilde{\theta}_j^{\star T} \langle \bar{F}_j \tilde{U}_j \rangle_{p_j}. \tag{62}$$

To estimate the expected values, we compute R samples $\varphi_{j,r}$ of p_j by coarse-graining R samples $\varphi_{0,r}$ in the training dataset. We then estimate $\langle \tilde{U}_j \tilde{U}_j^T \rangle_{p_j}$ with an empirical average over all $\varphi_{j,r}$.

The main difficulty is to estimate $\langle \bar{F}_j^2 \rangle_{p_j}$ and $\langle \bar{F}_j \tilde{U}_j \rangle_{p_j}$, where $\bar{F}_j = -\log \int e^{-\bar{\theta}_j^T \bar{U}_j} d\bar{\varphi}_j$ up to an additive constant. We evaluate these moments with a thermodynamic integration [67]. This is done by introducing a family of models $\bar{\theta}_j^T \bar{U}_{j,\lambda}$, which performs an interpolation between a Gaussian model and the full-scale interaction energy model. For this purpose, the nonlinear potential term in Eq. (55) is multiplied by $\lambda \in [0, 1]$:

$$\bar{\theta}_j^T \bar{U}_{j,\lambda}(\varphi_{j-1}) = Q_{\varphi_j}(\bar{\varphi}_j) + \lambda \bar{C}_j^T V(\varphi_{j-1}), \tag{63}$$

with a first term which is quadratic in $\bar{\varphi}_j$ for fixed φ_j,

$$Q_{\varphi_j}(\bar{\varphi}_j) = \frac{1}{2} \sum_{\ell=j}^{J+1} \bar{\varphi}_j^T \bar{K}_{j,\ell} \bar{\varphi}_\ell - \bar{C}_j^T V(P\varphi_{j-1}). \tag{64}$$

The thermodynamic integration interpolates linearly the Gaussian measure ($\lambda = 0$) and the original one ($\lambda = 1$) by estimating expected values with empirical averages. Appendix B 2 shows that it involves MCMC calculations on $\bar{\varphi}_j$ for fixed φ_j, which is fast and not affected by the critical slowing-down. As explained in the previous section, it mostly results from Theorem IV.1, which proves that wavelet fields $\bar{\varphi}_j$ have covariances that are well conditioned.

In summary, the free-energy estimation is made jointly by thermodynamic integration and linear regression. This method is numerically stable even close to critical points because it requires sampling only the wavelet field, keeping the coarse-grained field frozen. The numerical experiments in Sec. VI over cosmological data show that small errors introduced by the free-energy estimation induce small errors on the microscopic energy E_0. However, these small microscopic energy errors may lead to large errors when sampling the probability distribution model p_0 using E_0. This is due to the instabilities of microscopic coupling parametrizations of E_0. For instance, it is inherent to the phenomenon of phase transition that a small change in the parameters of the energy (the coupling constants) can change the physical behavior drastically. The WCRG circumvents this problem by generating new samples using the representation in Eq. (22), and it avoids the instabilities by using the parametrization based on conditional coupling parameters $\bar{\theta}_j$. We discuss this issue further when applying the WCRG to cosmological data.

Table I summarizes the different steps to compute the microscopic energy function, with a direct approach or with the WCRG. The WCRG estimates the coupling parameters $\bar{\theta}_j$ of $\bar{p}_j$, and the parameters $\tilde{\theta}_j$ of the free energy $\bar{F}_j$ at all scales 2^j.

VI. NUMERICAL APPLICATIONS OF WCRG: SAMPLING, ESTIMATION OF ENERGY FUNCTION, AND ABSENCE OF CRITICAL SLOWING-DOWN

This section presents three different numerical applications of the WCRG on two-dimensional ($d = 2$) fields. For each example, we first show the results of WCRG sampling and then the recovery of microscopic energy functions.

We start with a Gaussian field theory associated with the Ornstein-Uhlenbeck stochastic process. It provides a simple illustration and a validation test of our numerical method over a system with long-range correlated fields. As a second application, we focus on the φ^4 model, which is central in the theory of phase transitions and provides a challenging estimation problem with nonlinear potential. The third application concerns the characterization of weak-gravitational-lensing maps in cosmology associated with the distribution of dark matter. In this last example, the field is inherently out of equilibrium, and there is no

TABLE I. Summary of multiscale WCRG estimation. It decomposes the microscopic probability $p_0(\varphi_0) = \alpha p_J(\varphi_J)\prod_{j=1}^{J}\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ and its energy function $E_0(\varphi_0) = \theta_J^T U_J(\varphi_J) + \sum_{j=1}^{J}(\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) - \tilde{\theta}_j^T \tilde{U}_j(\varphi_j)) + c_0$. The coupling parameters $\bar{\theta}_j$ of $\bar{p}_j$ are estimated by minimizing a KL divergence. The parameters $\tilde{\theta}_j$ of $\bar{F}_j$ are computed by linear regression.

Estimation method	Parametrized models	Inference
Direct estimation of $p_0(\varphi_0)$	$p_{\theta_0}(\varphi_0) = Z_0^{-1}\exp[-\theta_0^T U_0(\varphi_0)]$	$\min_{\theta_0} D_{\mathrm{KL}}(p_0 \| p_{\theta_0})$
Conditional estimation of $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$	$\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j) = \bar{Z}_j^{-1}\exp[-\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) + \bar{F}_j(\varphi_j)]$	$\min_{\bar{\theta}_j} D_{\mathrm{KL}}(p_{j-1} \| \bar{p}_{\bar{\theta}_j} p_j)$
Estimation of free energies $\bar{F}_j(\varphi_j)$	$\tilde{\theta}_j^T \tilde{U}_j(\varphi_j)$	$\min_{\tilde{\theta}_j} \langle (\bar{F}_j - \tilde{\theta}_j^T \tilde{U}_j)^2 \rangle_{p_j}$

explicit energy model to compare with. However, we show that the resulting model reproduces visually and statistically nearly identical fields. This highly nontrivial example shows the potentiality of our method. The algorithmic aspects of our implementation of the WCRG are in Appendix C.

A. Gaussian field theory

We consider a model in which the field φ_0 is Gaussian at the finest scale, and thus has an energy function

$$E_0(\varphi_0) = \frac{1}{2}\varphi_0^T K_0 \varphi_0$$

without a nonlinear potential energy term [i.e., $V_0(\varphi_0) = 0$]. In particular, K_0 is given by

$$K_0 = \alpha \mathrm{Id} - \beta\Delta \quad \text{with} \quad \beta > 0, \tag{65}$$

where Δ is a discretized approximation of the Laplacian, as in Eq. (11). The corresponding Gibbs measure is the equilibrium measure of the Ornstein-Uhlenbeck stochastic process on the field φ_0. The convolutional matrix K_0 is the inverse of the covariance of φ_0, which is diagonal in Fourier space with eigenvalues

$$\lambda_{\varphi_0}(k) = \hat{K}_0^{-1}(k) = \frac{1}{\beta(|k|^2 + \alpha/\beta)}, \tag{66}$$

where k is the Fourier wave vector. In consequence, the ratio α/β defines a correlation length $\xi = 2\pi(\alpha/\beta)^{-\frac{1}{2}}$, which is the control parameter specifying the model. We normalize the overall factor β such that $\langle|\varphi_0(i)|^2\rangle_{p_0} = 1$. The field φ_0 displays critical fluctuations for $|k| > 2\pi/\xi$, and thus for length scale smaller than ξ.

We now present results obtained with WCRG models and sampling of this Gaussian process. Figures 5(a) and 5(c) show typical samples of the Ornstein-Uhlenbeck process in Eq. (65) of size $L = 32$, for $\xi = 4$ and $\xi = 32$ (training data). These are short-ranged and long-ranged correlated fields, respectively. The values $\xi = L = 32$ correspond to the critical point of the system. Figures 5(b) and 5(d) show two samples synthesized by the WCRG coarse-to-fine sampling, for the same values of ξ. These visual textures cannot be discriminated from the ones in Figs. 5(a) and 5(c). Appendix C 1 gives more details on these simulations.

For a Gaussian process, we evaluate the accuracy of the WCRG model by computing the precision of the estimated microscopic energy kernel K_0. Figure 6 compares the (radially averaged) Fourier eigenvalues $\hat{K}_0(k)$ of K_0 for the original and the estimated model by the WCRG, for $\xi = 4$ and $\xi = 32$. They perfectly superimpose because the conditional probability models are exact at all scales, since the energy functions remain Gaussian. Appendix C 1 explains the implementation. It demonstrates that WCRG calculations have no critical slowing-down in the presence of long-range correlations $\xi = L$.

B. The φ^4-field theory

The φ^4-field theory is the simplest model, which contains all the key ingredients of standard second-order phase transitions, such as large-scale collective behaviors, critical properties, long-range correlations, and self-similarity at the critical point. It has also played a central role in testing new techniques and ideas [68]. Here, we follow the same

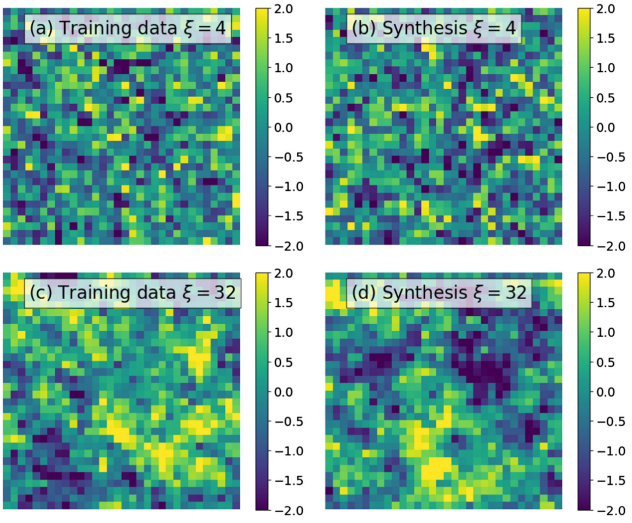

FIG. 5. (a),(c) Training samples of the Ornstein-Uhenbeck process for $\xi = 4$ (a) and $\xi = 32$ (c). (b),(d) Synthesized fields generated by WCRG sampling for $\xi = 4$ (b) and $\xi = 32$ (d).

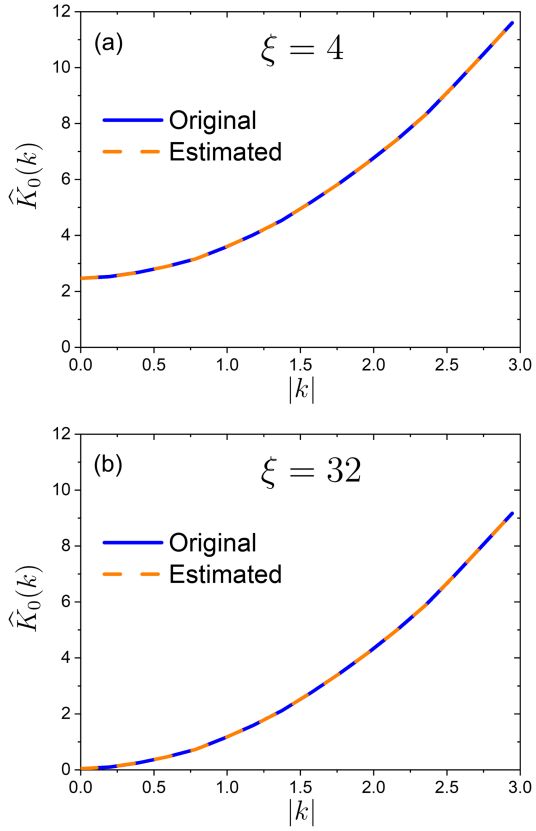

FIG. 6. Fourier eigenvalues $\hat{K}_0(k)$ of the Ornstein-Uhlenbeck kernel K_0 radially averaged over constant Fourier wave-vector modulus $|k|$, for $\xi = 4$ in (a) and $\xi = 32$ in (b). Original eigenvalues are shown with a solid curve and the estimated ones with a dashed curve.

strategy and apply the WCRG method to the φ^4-field theory in two dimensions.

The microscopic φ^4 model on a discrete lattice has a local potential introduced in Eq. (10),

$$E_0(\varphi_0) = \frac{1}{2}\varphi_0^T K_0 \varphi_0, + C_0^T V(\varphi_0). \tag{67}$$

The quadratic kernel is $K_0 = -\beta\Delta$ where β plays a role of inverse temperature, and Δ is the Laplacian. The local potential $C_0^T V(\varphi_0) = \sum_i C_0^T v(\varphi_0(i))$ has a double well

$$C_0^T v(\varphi_0(i)) = \varphi_0^4(i) - (1 + 2\beta)\varphi_0^2(i). \tag{68}$$

Previous numerical work has shown the existence of a second-order phase transition in the thermodynamics limit at $\beta_c \simeq 0.67$ [47]: For $\beta < \beta_c$ the system is disordered, whereas for $\beta > \beta_c$ the system is in the ordered, or broken-symmetry, phase.

We consider a two-dimensional system with $L = 32$ unless otherwise stated. We focus on four values of β: $\beta = 0.5, 0.6, 0.67$, and 0.76, which covers the three different regimes of the model, the disordered ($\beta = 0.5, 0.6$), critical ($\beta = 0.67 \simeq \beta_c$), and ordered ($\beta = 0.76$) phases. For each value of β, we generate $R = 10\,000$ samples of the field φ_0, which we use as a training dataset to perform the WCRG. Typical configurations are shown in the left panels of Fig. 7. See Appendix C 2 for more details and Appendix C 5 for a larger system size $L = 64$.

The right panels of Fig. 7 display samples synthesized with a coarse-to-fine WCRG, for $\beta = 0.5$, 0.67, and 0.76. They visually cannot be discriminated from samples of the original process in the left panels, even at $\beta \simeq \beta_c$. Figure 8 superimposes the marginal histogram of field values $\varphi_0(i)$ obtained from the training dataset and the one generated by coarse-to-fine WCRG sampling for $\beta \simeq \beta_c$. We also observe an excellent agreement and numerically confirm that the WCRG sampling correctly captures the critical behavior two point, as it reproduces well the spatial correlation function characterizing the power-law decay associated with critical phenomena (see Appendix C 4).

Figure 9(a) shows the nonlinear local potential $C_j^T v$ calculated at each scale 2^j for $\beta < \beta_c$. Recall that $C_j^T V(\varphi_j) = \sum_i C_j^T v(\varphi_j(i))$ in Eqs. (12) and (15). At large scales, the potential has a single well centered at zero because the disordered phase is nearly Gaussian at coarser scales. During the WCRG flow from coarse to fine scales,

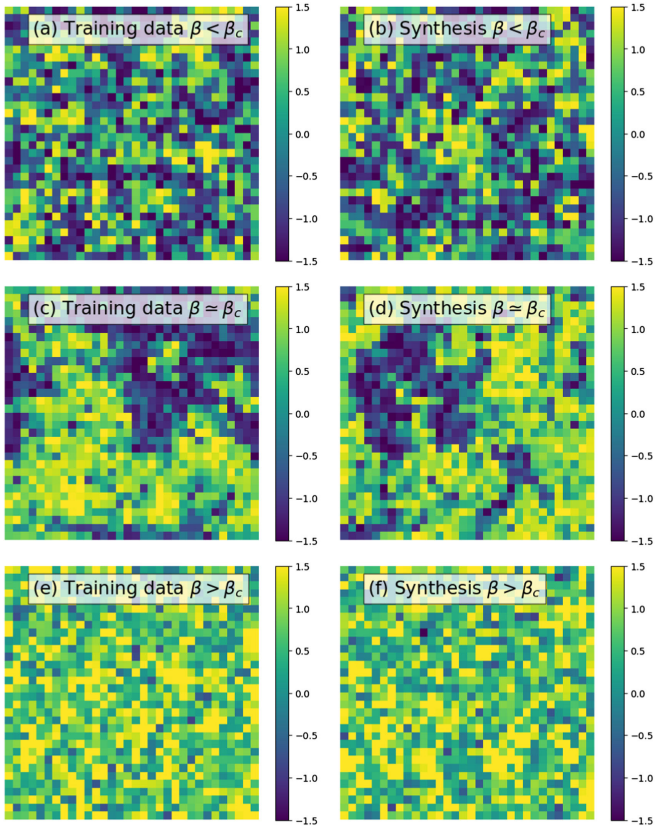

FIG. 7. Original training samples of φ^4 and synthesized fields generated by the WCRG, for $\beta = 0.5 < \beta_c$, $\beta = 0.67 \simeq \beta_c$, and $\beta = 0.76 > \beta_c$ from top to bottom: (a, b, c) training data, (d, e, f) synthesis.

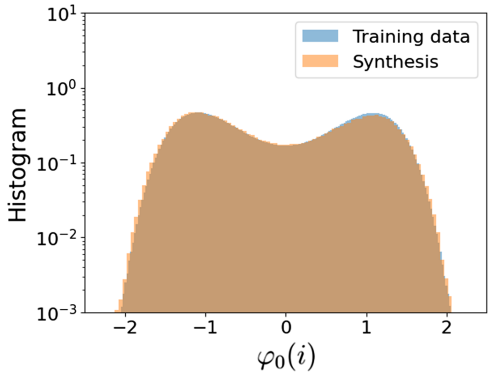

FIG. 8. Superimposed histograms of field values computed from φ^4 training data (blue) and WCRG synthesized fields (orange), for $\beta \simeq \beta_c$.

$C_j^T v$ progressively acquires a double-well shape. This figure illustrates the inverse RG flow induced by the WCRG. It looks like a backward run of the forward nonperturbative RG flows [68]. The analogous WCRG flow for $\beta \simeq \beta_c$ is presented in Fig. 9(b). The potential $C_j^T v$ has a double-well shape which remains stable for many intermediate scales. This is more evidence that the WCRG correctly captures the physical behavior associated with criticality. In fact, on the basis of RG theory, one expects to find a scale-invariant theory on intermediate scales. The stable local potential found by the WCRG is a striking manifestation of this phenomenon. Finally, Fig. 9(c) presents the result in the ordered phase. We start from a broken-symmetry phase at a large scale, where the data are all in the same positively magnetized phase. Remarkably, the WCRG recovers the symmetric potential at the finest scale. In theory, the nonlinear potential should have a slight asymmetry because an infinitesimal positive field has to be applied to obtain the positively magnetized phase at a large scale. Yet, this effect is very small and not detected numerically. The ability of the WCRG to recover the microscopic (broken) symmetry is important in applications of this method to characterize properties of microscopic energies.

The accuracy of the WCRG model of φ^4 is evaluated by comparing the values of K_0 and C_0, which define the microscopic energy function in Eq. (67). We focus on the phase transition $\beta \simeq \beta_c$, which is the most challenging case. Figure 10 shows that the estimated local potential at the finest scale is nearly equal to the original one. Figure 11(a) shows the convolution kernel of K_0 estimated by the WCRG. Figure 11(b) superimposes the eigenvalues of the original Laplacian K_0 with its WCRG estimation. The excellent agreement demonstrates that the WCRG provides a precise estimation of the φ^4 microscopic energy.

Since the φ^4 model is *a priori* known to be local, to regress the free energies $\bar{F}_j$ we impose that $\tilde{C}_j = \bar{C}_{j+1}$ for all $1 \le j \le J-1$ and $\tilde{C}_J = C_J$ in Eq. (56). This is a usual assumption in RG treatments of the φ^4 model. In our case, we can validate *a posteriori* this assumption by computing

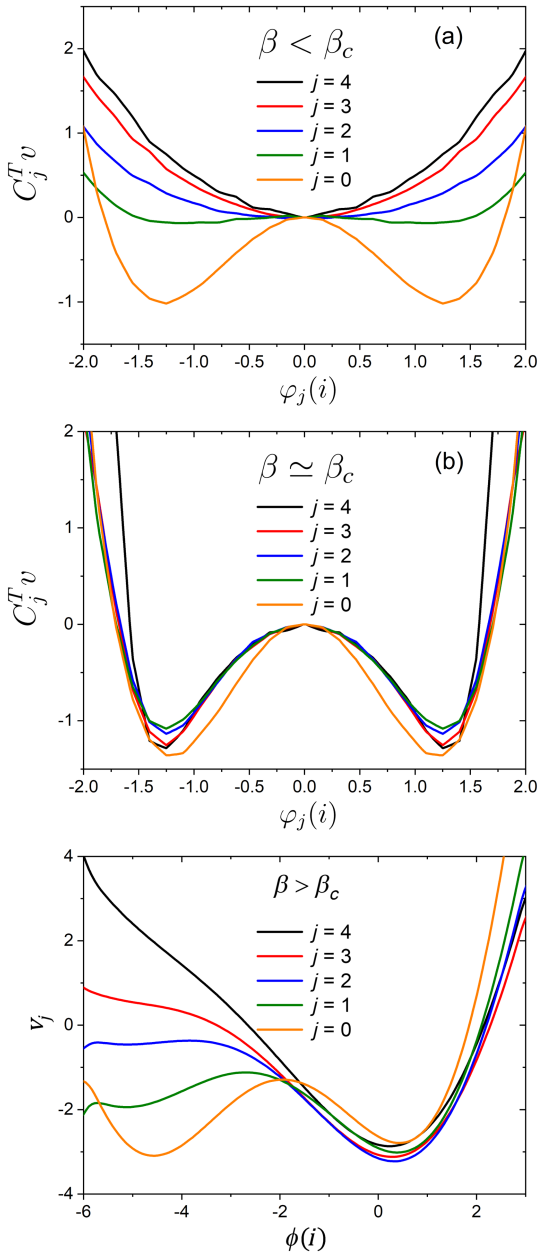

FIG. 9. WCRG flow of nonlinear potentials $C_j^T v(\varphi_j(i))$ across scales 2^j, for $\beta = 0.5$ in (a), $\beta = 0.67 \simeq \beta_c$ in (b), and $\beta = 0.76$ in (c).

the resulting linear regression error of the free energy. It is measured by the ratio between the mean-square regression error $\ell(\tilde{\theta}_j^\star)$ in Eq. (60) and the average squared free-energy $\langle \bar{F}_j^2(\varphi_j) \rangle_{p_j}$. The relative error with this locality assumption is of order 10^{-5} for all the β that we considered. It confirms *a posteriori* that a local potential provides an accurate model of φ^4 training examples.

C. Cosmological data

This section applies the WCRG estimation and sampling to model weak-lensing images in cosmology. Gravitational

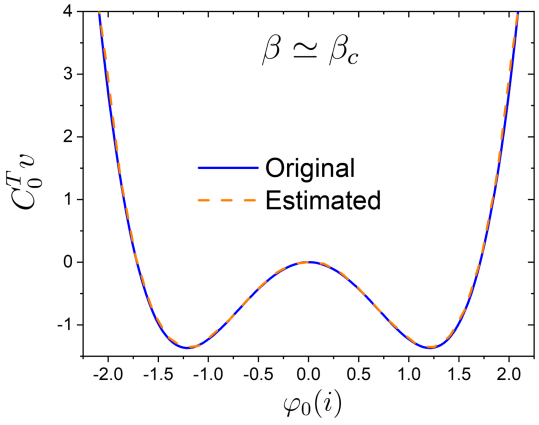

FIG. 10. Comparison for $\beta \simeq \beta_c$ of the original nonlinear local potential $C_0^T v(\varphi_0(i))$ at the microscopic scale (solid curve) and its WCRG estimation (dashed curve).

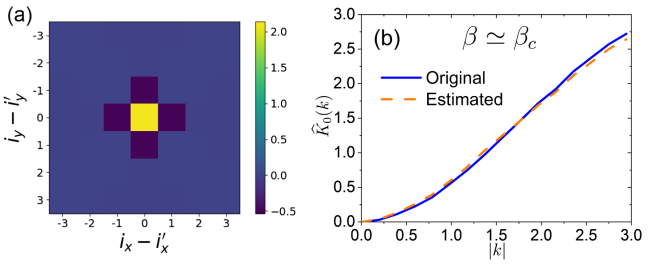

FIG. 11. (a) Estimated convolutional kernel $K_0(i - i')$ in the $(d = 2)$-dimensional space for φ^4. (b) Comparison for $\beta \simeq \beta_c$ of the Fourier eigenvalues $\hat{K}_0(k)$ of K_0 of the original model (solid curve) and the one estimated by the WCRG (dashed curve).

lensing deforms images of background objects such as galaxies near a foreground mass [69,70]. Galaxy clusters are the largest gravitationally bound structures in the Universe, with approximately 80% of cluster content in the form of dark matter. It can cause strong and weak statistically coherent distortions of background sources on the order of 10% (cluster weak lensing) [70]. A major scientific challenge is to study the weak-gravitational-lensing signature of large-scale structures in the Universe and to understand fundamental physics such as the nature of dark energy and the total mass of neutrinos. In particular, several groups have been trying to capture rich information that is beyond the traditional two-point statistics using non-Gaussian statistics through high-order moments [71–75]. More recently, iterated wavelet transforms called scattering transform have been used to generate weak-lensing images conditioned by scattering moments [36]. However, none of these techniques are able to define the explicit microscopic energy function which generates such fields.

In the following, we apply the WCRG to study the statistics of weak-gravitational-lensing maps. We use a set of simulated convergence maps computed by the Columbia lensing group [76,77] as a training dataset. It can be considered as test convergence maps for the next-generation space telescope *Euclid* [78]. Details about the data and numerical implementation can be found in Appendix C 3. The top row of Fig. 12 shows several examples of the convergence maps used as the training dataset. Compared to the φ^4-field theory images, one sees rare high-amplitude local fluctuations (shown in yellow) reflecting a higher concentration of dark matter. This characterizes a more complex statistics typical of highly non-Gaussian processes (with a long tail; see below).

On the bottom row of Fig. 12, we show the synthesized images generated by WCRG sampling using the estimated scale interaction energy functions introduced in Sec. III C. The original (top row) and synthesized (bottom row) define textured images which cannot be discriminated visually. It shows that the WCRG model is able to capture the high-amplitude non-Gaussian fluctuations. A more quantitative analysis is performed by comparing the histograms of $\varphi_0(i)$ for the training dataset and the fields generated by sampling the WCRG models. Figure 13 shows that both distributions are nearly equal. In particular, the WCRG model reproduces the long tail involving rare events, which has been hampering various approaches in statistical analysis.

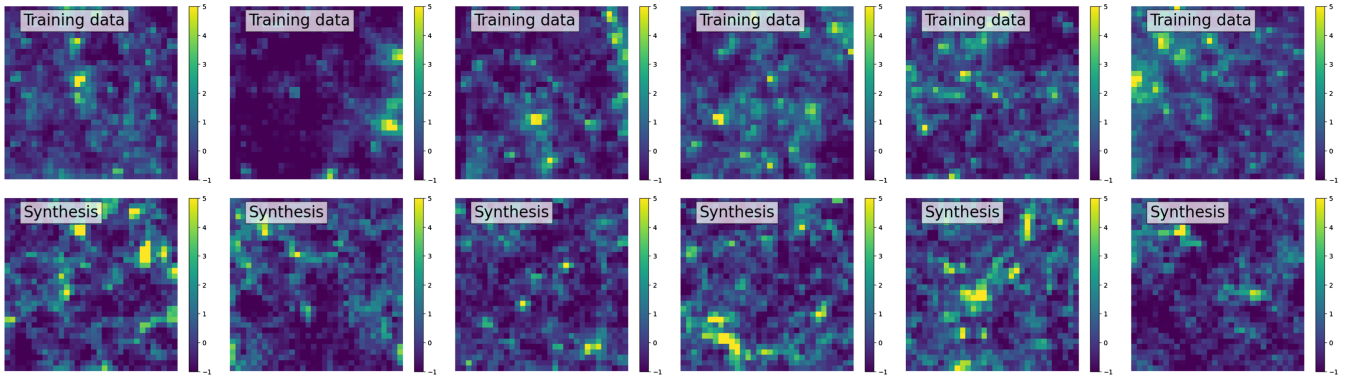

FIG. 12. Weak-gravitational-lensing maps in cosmology. Top: training dataset from the Columbia lensing group [76,77]. Bottom: synthesized fields generated by the WCRG sampling algorithm.

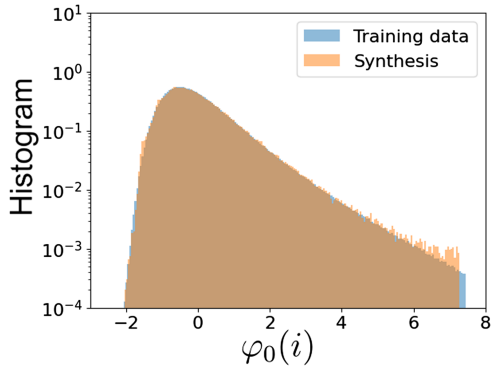

FIG. 13. Superposition of the normalized histogram of $\varphi_0(i)$ for the weak-lensing training data (blue) and samples of multiscale WCRG models (orange).

It demonstrates that the WCRG is a powerful tool to tackle highly nontrivial scientific problems, including cosmology.

We now turn to the recovery of the energy function at the microscopic scale. For weak lensing, the energy function has a different meaning from the φ^4 model in thermal equilibrium. Since cosmological data are obtained from an inherently out-of-equilibrium process, the energy function is not the Hamiltonian in the sense of a generator of the dynamics. However, it provides a compact and explicitly interpretable parametrization of the high-dimensional probability distribution. It can be regarded as an effective Hamiltonian leading to a Boltzmann-like representation of the probability distribution. Looking for an effective Hamiltonian in a nonequilibrium problem is generally a challenge. It has been possible to obtain it only in a few cases [79,80], mostly in simple $d = 1$ systems. The estimation of the microscopic energy function of a cosmological system is particularly challenging because of the existence of long-range interactions resulting due to gravitation.

The WCRG multiscale microscopic energy model is decomposed in Eq. (57) as a sum of local potential interactions at all scales,

$$E_0(\varphi_0) = \frac{1}{2}\varphi_0^T K_0 \varphi_0 + \sum_{j=0}^{J} C_j'^T V(\varphi_j). \tag{69}$$

We see that $C_0' = \bar{C}_1$ and $C_j' = \bar{C}_{j+1} - \tilde{C}_j$ for $j \geq 1$, where $\bar{C}_j^T V$ is potential of the interaction energy $\bar{E}_j$, and $\tilde{C}_j^T V$ is the potential of the free energy $\bar{F}_j$. For stationary fields, K_0 is a convolution operator. Figure 14 displays the estimated convolution kernel $K_0(i - i')$ and its Fourier eigenvalues $\hat{K}_0(k)$. It is the counterpart of Fig. 11 for φ^4. It appears to be close to a discrete Laplacian, which was not expected. We also find that the operators K_j computed at all scales $2^j \geq 1$ remain close to a discrete Laplacian.

Figure 15(a) shows the evolution of the estimated interaction potential $\bar{C}_j^T V(\varphi_{j-1}) = \sum_i \bar{C}_j^T v(\varphi_{j-1}(i))$ across

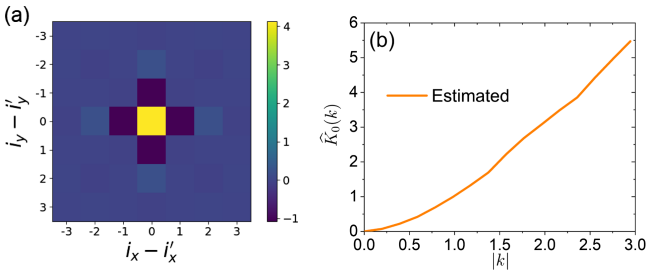

FIG. 14. (a) Estimated convolutional kernel $K_0(i - i')$ in the $(d = 2)$-dimension space from the WCRG model of weak-gravitational-lensing maps. (b) Fourier eigenvalues $\hat{K}_0(k)$.

scales. As opposed to φ^4, Fig. 15(b) shows that the multiscale potential terms $C_j' = \bar{C}_{j+1} - \tilde{C}_j$ are nonzero at all scales 2^j for $1 \leq j \leq J$. They capture nonlocal interactions in the microscopic-scale energy in Eq. (69). The multiscale potentials $C_j'^T v(\varphi_j(i))$ in Fig. 15(b) favor large values of the field at smaller scales, which increases the relevance of the secondary metastable minimum of $\bar{C}_j^T v(\varphi_{j-1}(i))$ in Fig. 15(a). However, this effect diminishes at large scales, thus avoiding the presence of large regions with high value of

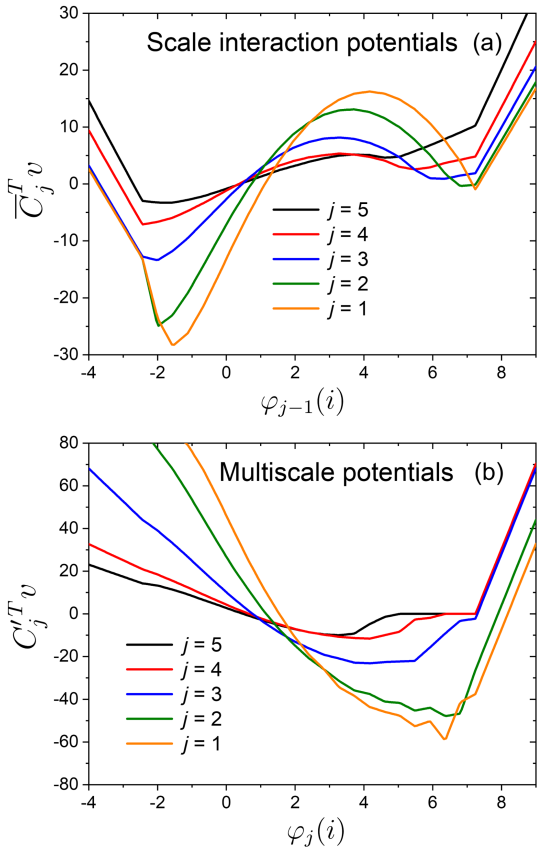

FIG. 15. (a) Scale interaction potential $\bar{C}_j^T v((\varphi_{j-1}(i))$ across scales 2^j estimated from weak-lensing maps. (b) Multiscale potentials $C_j'^T v(\varphi_j(i))$ of the microscopic energy function E_0 for $j \geq 1$.

the field. These nonlocal, long-range interactions that favor high-amplitude values locally but not globally, induce the long tail of the histogram in Fig. 13. This tail corresponds to the bright yellow spots in the weak-lensing images of Fig. 12. This interpretation echoes the aggregation of matters due to long-range gravitational forces.

D. Stable and unstable representations of probability distributions

Behind the fast convergence of WCRG parameter estimation lies the fact that conditional probabilities of wavelet fields provide stable parametrizations of large classes of probability distributions, even close to phase transitions. The WCRG decomposes a distribution p_0 into $p_0 = \alpha p_J \prod_{j=1}^{J} \bar{p}_j$, and it approximates each conditional probability $\bar{p}_j(\bar{\varphi}_j|\varphi_j)$ with a model $\bar{p}_{\bar{\theta}_j}(\bar{\varphi}_j|\varphi_j)$. Small errors on the conditional couplings $\bar{\theta}_j$ do not strongly affect the properties of p_0. Indeed, the conditional probability distribution $\bar{p}_j$ is not singular, as the fast degrees of freedom $\bar{\varphi}_j$ are well behaved with short-range correlations. This is in stark contrast with standard energy-based models which directly approximate p_0 by p_{θ_0}. A small change of the coupling parameters in θ_0 can lead to dramatic changes on p_{θ_0} near phase transitions.

This is vividly illustrated over weak-lensing maps. A model p_{θ_0} of p_0 is defined by the microscopic energy model in Eq. (69) defined by $\theta_0 = (\frac{1}{2}K_0, C'_0, C'_1, \ldots, C'_J)$, with $C'_0 = \bar{C}_1$ and $C'_j = \bar{C}_{j+1} - \bar{C}_j$ for $j \geq 1$. Figure 16(b) is a sample of p_{θ_0} calculated with an MCMC algorithm updating φ_0 directly. It has very different statistical properties from original weak-lensing images as in Fig. 12, which clearly appears in the tail of the superimposed histograms in Fig. 16(d). The excess tail corresponds to high-amplitude clusters of sites organized in real space in an ant colony shape; see Fig. 16(b). Their typical width is 2 to 3 pixels, which indicates that the multiscale potential parameter C'_j has an error at $j = 1$. Such statistical errors do not appear with a coarse-to-fine wavelet sampling of the WCRG model, as shown in Fig. 13. This model is parametrized by the conditional coupling parameters $\bar{\theta}_j$, which include the potential parameters $\bar{C}_j$ of $\bar{E}_j$. The errors on $C'_j = \bar{C}_{j+1} - \tilde{C}_j$ in E_0 are thus produced by errors when estimating the potential parameters $\tilde{C}_j$ of the free energies $\bar{F}_j$. Yet, we show that these errors are small, and the large statistical errors in Fig. 16(d) are due to instabilities of the coupling parametrization θ_0.

To demonstrate this property, we modify θ_0 by changing only the potential parameter C'_j for $j = 1$. We initialize C'_1 with the value obtained from the free-energy calculation. Its value is updated with a gradient descent on the KL divergence $D_{\mathrm{KL}}(p_0 \| p_{\theta_0})$, with the algorithm of Sec. IV A. Figure 16(a) shows that this optimization produces a very small modification of $C'^T_1 v$. Yet, the images sampled from this slightly modified microscopic energy E_0 have a totally different result in Fig. 16(c), which now match the weak-lensing maps. Indeed, their histogram in Fig. 16(e) is superimposed over the histogram of weak-lensing images.

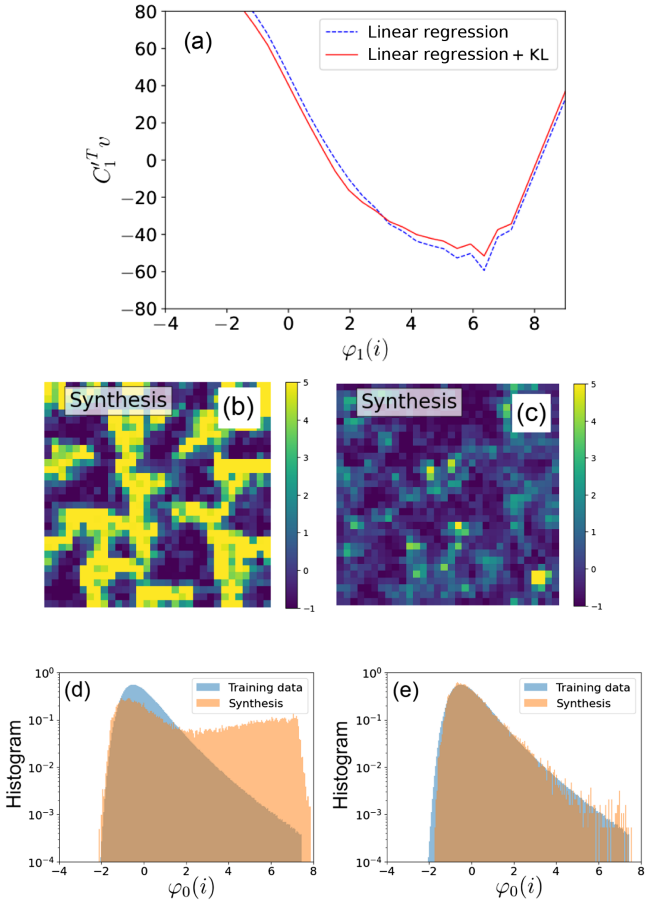

FIG. 16. (a): Multiscale potential $C'^T_1 v(\varphi_1(i))$ at a fine scale $j = 1$ used in the microscopic energy E_0. The dashed curve potential is calculated with a linear regression of the free energy $\bar{F}_1$. The full curve is obtained with a direct KL-divergence minimization. (b),(c) Synthesized images by sampling using the energy E_0. The images (b) and (c) correspond to the potentials $C'^T_1 v$ shown in (a), computed from $\bar{F}_1$ and by minimizing the KL divergence, respectively. (d),(e) Superimposed histograms of $\varphi_0(i)$ of synthesized images (in orange) and original weak-lensing images (in blue). The large errors in the tail of histogram (d) correspond to high-amplitude patches in image (b), whereas the histogram (e) of image (c) is well reproduced.

This numerical experiment shows that the coupling parameter θ_0 of E_0 is estimated with a good accuracy through free-energy regressions. However, this coupling parametrization of E_0 is highly unstable. Such instabilities also appear at phase transitions. WCRG conditional probability representation circumvents this problem by relying only on the parameters $\bar{\theta}_j$ of wavelet-conditional probabilities, which are stable. It thus leads to more reliable generative models of complex many-body problems.

E. Absence of critical slowing-down for the WCRG

This section illustrates the numerical stability of WCRG calculations over the φ^4 model. In particular, we contrast the WCRG with standard direct coupling estimation approaches plagued by critical slowing-down close to the critical point. As we discuss in Sec. IV, the critical slowing-down appears both in the gradient-descent dynamics and to estimate moments by MCMC. Numerical details are given in Appendix C. For Gaussian models, the absence of critical slowing-down of WCRG algorithms is demonstrated in Appendix C 1.

For the φ^4 model, we begin by studying the critical slowing-down of the MCMC for direct coupling estimation when β approaches β_c, and when L is increased at $\beta \simeq \beta_c$. The MCMC simulations shown in Fig. 17 are performed at the last gradient-descent step when the coupling parameters have converged to an optimal value. Let us denote $\varphi_0^{(t)}$ the finest scale field at a Monte Carlo time step t, that is evolving with the MCMC simulation. One MC time step corresponds to L^2 trial MC updates over an image of L^2 pixels. The magnetization is written by $\Phi(t) = |L^{-2}\sum_i \varphi_0^{(t)}(i)|$. Its time autocorrelation function $A_\Phi(t)$ is given by

$$A_\Phi(t) = \frac{\langle \delta\Phi(t)\delta\Phi(0)\rangle_{\text{time}}}{\langle \delta\Phi^2\rangle_{\text{time}}}, \tag{70}$$

where $\langle\cdots\rangle_{\text{time}}$ denotes the time average under the stationary state and $\delta\Phi(t) = \Phi(t) - \langle\Phi\rangle_{\text{time}}$. Figure 17(a) plots $A_\Phi(t)$ for a direct Monte Carlo update of φ_0 for different values of β

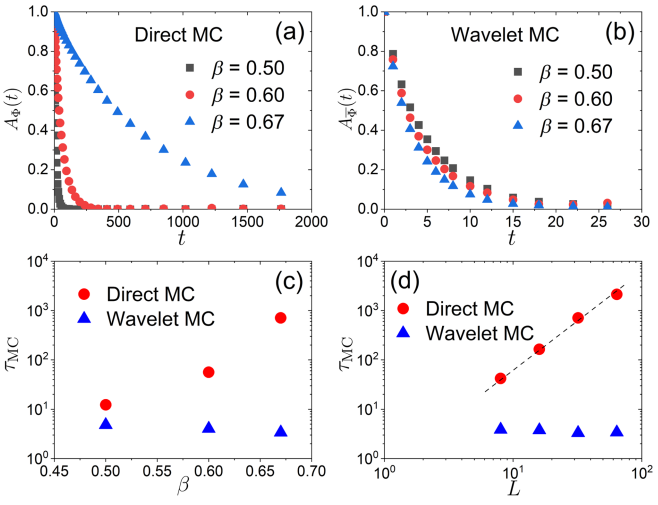

FIG. 17. (a),(b) Autocorrelation functions for direct MC simulations updating φ_0 in (a), and for wavelet MC simulations updating $\bar{\varphi}_1$ in (b). It is computed for different β approaching β_c for the φ^4 model with $L = 32$. (c),(d) Decorrelation timescale τ_{MC} for the direct and wavelet MC simulations. It is computed as a function of β with $L = 32$ in (c), and as a function of L up to $L = 64$ with $\beta \simeq \beta_c$ in (d). The dashed line in (d) corresponds to $\tau_{\text{MC}} \sim L^z$ with $z \simeq 2$.

approaching β_c. The decorrelation timescale τ_{MC} defined by $A_\Phi(\tau_{\text{MC}}) = 1/e$ increases in Fig. 17(c). The critical slowing-down also appears when L increases at $\beta \simeq \beta_c$. As expected, Fig. 17(d) shows that τ_{MC} diverges as $\tau_{\text{MC}} \sim \mathcal{O}(L^z)$ with $z \simeq 2$ at the critical point [10], whereas it remains finite for $\beta < \beta_c$.

In the WCRG, the MCMC is performed at each scale on the wavelet fields $\bar{\varphi}_j$, while φ_j is fixed. We call it a wavelet MC. Figure 17(b) shows the resulting autocorrelation function $A_{\bar{\Phi}}(t)$ of $\bar{\Phi}(t)$ given by

$$\bar{\Phi}(t) = \frac{1}{3(L2^{-j})^2}\sum_{m,i}|\bar{\varphi}_j^{(t)}(m,i)|. \tag{71}$$

It is displayed at the finest scale $j = 1$, but coarser scales behave similarly and require fewer computations. The results show that, remarkably, $A_{\bar{\Phi}}(t)$ decays at the same rate for all β, which is very different from a direct MC updating of φ_0. The decorrelation timescale τ_{MC} is defined by $A_{\bar{\Phi}}(\tau_{\text{MC}}) = 1/e$. Observe that it remains nearly constant as a function of β in Fig. 17(c), and as a function of L in Fig. 17(d). It means that the wavelet MC is not affected at all by the critical slowing-down.

We now study the convergence of the gradient descent, which estimates coupling parameters by minimizing a KL divergence. It also involves a critical slowing-down near the phase transition. The error is measured by the squared norm of moments errors. It measures the amplitudes of the parameter gradients in Eqs. (41) and (46). For the direct gradient-descent algorithm of Sec. IV A, the moment error is

$$\mathcal{L}_{\text{D}}(t) = \|\langle U_j\rangle_{p_{\theta_j^{(t)}}} - \langle U_j\rangle_{p_j}\|^2, \tag{72}$$

where t is the gradient-descent time step. It is computed at the finest scale $j = 0$. For the wavelet-conditional gradient descent of Sec. IV B, the moment error is

$$\mathcal{L}_{\text{C}}(t) = \|\langle \bar{U}_j\rangle_{p_{\bar{\theta}_j^{(t)}}} - \langle \bar{U}_j\rangle_{p_{j-1}}\|^2. \tag{73}$$

It is shown only at the finest scale $j = 1$, where its decay is the slowest. Figure 18 compares the decay of $\mathcal{L}_{\text{D}}(t)$ and $\mathcal{L}_{\text{C}}(t)$, near the phase transition $\beta \simeq \beta_c$. For each method, we compute numerically the maximum step size ϵ such that the gradient decent converges and leads to an accurate microscopic energy function E_0. Numerical results show that $\epsilon = 0.005$ for the direct algorithm, whereas $\epsilon = 0.5$ is 100 times larger for the conditional gradient descent. As a result, $\mathcal{L}_{\text{C}}$ in Fig. 18(a) has a much slower decay than $\mathcal{L}_{\text{C}}$ in Fig. 18(b). It confirms that the WCRG is also able to precondition the gradient-descent Hessian in the φ^4 case and thus eliminate the critical slowing-down.

The remarkable performance of the WCRG comes from the fact that RG can handle a singular theory close to a critical point, contrary to perturbation theory, which treats

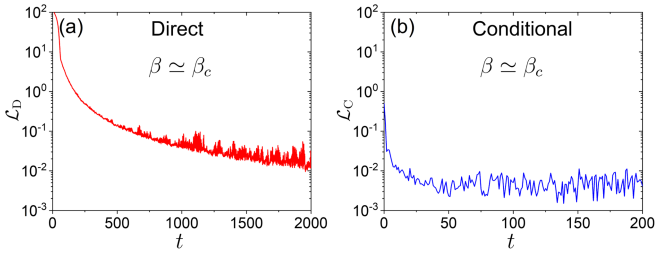

FIG. 18. Moment errors as a function of the number of gradient-descent steps t. (a) Error $\mathcal{L}_D$ in Eq. (72) for the direct algorithm. (b) Error $\mathcal{L}_C$ in Eq. (73) for the wavelet-conditional algorithm.

on the same footing all degrees of freedom. Close to a critical point, where the susceptibility to changes in the coupling parameters diverges, perturbation theory is bound to fail. On the contrary, since RG always works on the fast short-scale degrees of freedom, it is not affected by the critical point, and one can perform approximation on the RG flow safely [15]. In our context, the direct coupling estimation plays the same role as a perturbation theory. It fails when approaching the critical point because it works directly on the degrees of freedom and coupling parameters plagued by a singular behavior. Instead, the WCRG works at each scale on the wavelet fields corresponding to the shortest scales, which are noncritical and hence fast. This is the essential physical ingredient that makes the WCRG numerically stable and unaffected by the critical slowing-down.

VII. CONCLUSIONS AND DISCUSSION

The WCRG provides a new representation of high-dimensional probability distributions, as a product of conditional probabilities. They are associated with scale interaction energy functions, which can be estimated from limited sets of data. It addresses a major open issue in physics and machine learning. This approach is tightly related to the renormalization-group theory developed for second-order phase transitions. Fluctuations of the field, or fast degrees of freedom, are represented at each scale in a wavelet orthogonal basis.

The WCRG differs from a standard forward RG in several aspects. It does not suppose that the microscopic energy is known *a priori*, but it provides an estimation of this energy from data. A standard RG computes the flow of coupling parameters defined over the whole energy function. The WCRG rather computes the flow of conditional coupling parameters, which specify conditional probabilities of wavelet fields given coarse-grained fields. We show that estimations of conditional coupling parameters by maximum likelihood is well conditioned and avoids critical slowing-down. The WCRG generates new fields by sampling wavelet-conditional probability distributions, which also circumvents the critical slowing-down at phase transitions. Explicit expression of the microscopic energy can be recovered from WCRG estimations, with the thermodynamic integration and a linear regression that are introduced. This is applied to Gaussian and φ^4 models in thermal equilibrium, and to weak-gravitational-lensing maps in nonequilibrium. The study of the cosmological data is particulary challenging since the underlying distribution is unknown *a priori*.

We introduce local potential models at each scale, which can capture long-range interactions at the microscopic scale. These multiscale *Ansatzë* open the possibility to build models of complex physical fields including turbulences. Such models have indeed been studied in Refs. [34,35] by specifying dependences across scales through correlations of phases and amplitudes of wavelet coefficients. Applications of the WCRG to systems with different kinds of discrete or continuous symmetries, e.g., $O(N)$ ferromagnetic models, are straightforward as the basic building blocks (multiscale and conditional probabilities) remain the same—only the multiscale *Ansatzë* change. It would also be interesting to study applications of the WCRG to disordered systems, for which it could provide a new efficient sampling method of low-temperature configurations.

Many publications have pointed out the similarities between RG transformations and deep network architectures [19,81–83] and flow-based modeling [18]. The computational architecture of the WCRG bares some close resemblance to deep generative networks. This paper suggests that it may be a conditional RG which is calculated in many deep neural networks. Similar to deep networks, the WCRG decomposes the field at multiple scales with a wavelet transform which cascades convolutions and subsamplings. Local potentials are computed with piecewise linear rectifiers which connect wavelet coefficients at a given scale, stored in a network layer. The normalization is achieved by a normalization of variances, which corresponds to batch normalizations in deep neural networks. Network parameters correspond to coupling interactions in a wavelet-conditional RG, which are learned with a stochastic gradient-descent algorithm.

Artificial neural networks are very expressive but lack interpretability and need a lot of data. These are two major concerns in their applications to physics problems. The WCRG is physics based. It devises an approximate representation of probability distributions for many-body systems. Such a representation is based on the physical properties of the system at hand, in particular, multiscale representations of interaction energies. Its advantage compared to neural-network-based generative models is that it can be analyzed theoretically; i.e., one can explain why sampling is fast and estimation efficient. Moreover, the WCRG provides effective interaction energies governing the physical processes at different scales, thus allowing for physical interpretation and performing RG directly from

the data. Therefore, the WCRG presents some of the features that make neural-network-based generative models very useful while also remaining interpretable. The trade-off (which for us is an advantage, too) is that it requires physical insights to devise suitable *Ansatzë* for the interaction energies. Fewer data are needed the more the *Ansatz* is constrained by physical insights.

It has been verified that the integration of a WCRG decomposition in a deep generative network architecture provides a numerical stabilization and acceleration of image generation with score diffusion [84]. Moreover, it is shown in Ref. [85] that the WCRG can explain the ability of deep neural networks to circumvent the curse of dimensionality. Beyond the φ^4 model and weak-gravitational-lensing maps studied in this paper, local interaction models on wavelet coefficients across scales can indeed generate complex nonstationary complex images such as faces [85]. Remarkably, the nonstationarity of complex image fields, such as faces, is captured at the lowest wave numbers as phase transitions in the φ^4 model. Given nonstationary low wave numbers, local stationary conditional models of wavelet coefficients are used to generate these nonstationary fields.

In deep network diffusion models, the parameter estimation is performed by score matching [86] instead of maximum-likelihood optimization, which avoids using MCMC Metropolis sampling algorithms. Score matching is computationally much faster, but it is not guaranteed to reach an optimal solution if the log probability is not strongly log-concave. Numerical results in Ref. [87] show that the WCRG defines strongly log-concave conditional probability distributions for φ^4 and weak-lensing energy functions, so that one can accelerate computations by replacing maximum likelihood by score matching.

ACKNOWLEDGMENTS

We acknowledge funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, Reference No. ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and from the Simons Foundation Collaboration "Cracking the Glass Problem" (Grant No. 454935 to G. B.). We also thank Zoltan Haiman for sharing the convergence maps obtained by the Columbia Lensing group, and Antonin Brossollet and Simon Coste for careful reading of the manuscript.

APPENDIX A: WAVELET BASES AND REPRESENTATION OF OPERATORS

The fast wavelet transform decomposes a field φ_0 by using orthogonal filters G and $\bar{G}$. Appendix A 1 briefly reviews the construction of such filters in any spatial dimension d. A wavelet transform stabilizes renormalization-group computations because wavelet bases nearly diagonalize operators involved in the calculation of large classes of energy functions. Appendix A 2 explains that the wavelet fields $\bar{\varphi}_j$ computed by the fast wavelet transform of Sec. II A are decomposition coefficients in a wavelet orthonormal basis of $\mathbf{L^2}(\mathbb{R}^d)$. Theorem IV.1 is proved in Appendix A 3 by showing that a singular homogeneous operator, such as a Laplacian in a Gaussian model, is diagonally dominant in a wavelet orthonormal basis. We use the PYWAVELETS software [88] to compute the fast wavelet transform.

1. Wavelet filter design

We review the construction of wavelet filters G and $\bar{G}$ which satisfy the orthogonality conditions in Eq. (7). We begin in spatial dimension $d = 1$ and then introduce a separable extension of such filters in any dimension d.

a. One-dimensional filters

In dimension $d = 1$, the coarse-graining filter $G^{(1)}$ has a Fourier transform concentrated at low wave numbers. The superscript (1) denotes $d = 1$. The wavelet filter $\bar{G}^{(1)}$ instead computes high-wave-number variations. The orthogonality conditions in Eq. (7) have been proved to be satisfied [44] if the Fourier series $\hat{G}^{(1)}(k)$ of $G^{(1)}(n)$ satisfies for all $k \in [0, 2\pi]$,

$$|\hat{G}^{(1)}(k)|^2 + |\hat{G}^{(1)}(k+\pi)|^2 = 2 \quad \text{and} \quad \hat{G}^{(1)}(0) = \sqrt{2} \tag{A1}$$

and

$$\hat{G}^{(1)}(k) > 0 \quad \text{for } k \in [0, \pi/2]. \tag{A2}$$

The Fourier series $\hat{\bar{G}}^{(1)}(k)$ of the wavelet filter $\bar{G}^{(1)}(n)$ satisfies

$$\hat{\bar{G}}^{(1)}(k) = e^{-ik}\hat{G}^{(1)*}(k+\pi), \tag{A3}$$

where $i = \sqrt{-1}$ and $*$ is the symbol for the complex conjugate. This implies that

$$\bar{G}^{(1)}(n) = (-1)^{1-n}G^{(1)}(1-n). \tag{A4}$$

The Haar filter is a simple solution of Eq. (A1) given by

$$G^{(1)}(n) = 2^{-1/2} \quad \text{if } n = 0, 1 \quad \text{and}$$
$$G^{(1)}(n) = 0 \quad \text{otherwise.} \tag{A5}$$

It computes the one-dimensional Kadanoff block averaging in Eq. (2). The corresponding wavelet filter is $\bar{G}^{(1)}(0) = -2^{-1/2}$, $\bar{G}^{(1)}(1) = 2^{-1/2}$, and $\bar{G}^{(1)}(n) = 0$ if $n \neq 0, 1$. It iteratively computes Haar wavelet fields with Eq. (6).

The Shannon filter is another simple solution of Eq. (A1) whose Fourier transform is supported in $k \in [-\pi/2, \pi/2]$,

$$\hat{G}^{(1)}(k) = 2^{1/2}\mathbf{1}_{|k|\leq\pi/2}, \tag{A6}$$

where $\mathbf{1}$ is the characteristic function. The next section shows that the Haar and Shannon filters lead to Haar and Shannon wavelets, which are badly localized in the Fourier and real-space domains, respectively. Therefore, we rather use Daubechies filters defined in Ref. [46]. They define wavelets which are well localized both in the real-space and Fourier domains, which is important to efficiently represent operators involved in the calculations of energy functions.

b. Separable multidimensional filters

In dimension d, wavelet filters which satisfy the orthogonality conditions in Eq. (7) can be defined as separable products of the one-dimensional filters $G^{(1)}$ and $\bar{G}^{(1)}$. The d-dimensional low-pass filter $G^{(d)}$ is constructed as

$$G^{(d)}(n_1, n_2, \ldots, n_d) = \prod_{\ell=1}^{d} G^{(1)}(n_l). \tag{A7}$$

Let us write $\bar{G}_0^{(1)} = G^{(1)}$ and $\bar{G}_1^{(1)} = \bar{G}^{(1)}$. In dimension d, there are $2^d - 1$ wavelet channels with the associated $2^d - 1$ filters $\bar{G}_m^{(d)}$ for $1 \leq m \leq 2^d - 1$. These are obtained with different separable products of $\bar{G}_0^{(1)}$ and $\bar{G}_1^{(1)}$. By using a binary digit $b_1, b_2, \ldots, b_d$ with $b_\ell \in \{0, 1\}$, we define

$$\bar{G}_m^{(d)}(n_1, n_2, \ldots, n_d) = \prod_{\ell=1}^{d} \bar{G}_{b_\ell}^{(1)}(n_\ell). \tag{A8}$$

For example, in $d = 2$, there are three channels ($m = 1, 2, 3$):

$$\begin{aligned}
\bar{G}_1^{(2)}(n_1, n_2) &= G^{(1)}(n_1)\bar{G}^{(1)}(n_2),\\
\bar{G}_2^{(2)}(n_1, n_2) &= \bar{G}^{(1)}(n_1)G^{(1)}(n_2),\\
\bar{G}_3^{(2)}(n_1, n_2) &= \bar{G}^{(1)}(n_1)\bar{G}^{(1)}(n_2).
\end{aligned}$$

One can verify that these filters satisfy the orthogonality condition in Eq. (7) [37]. In this paper, the filter $G^{(d)}$ and the vector of $2^d - 1$ filters $\bar{G}_m^{(d)}$ are often denoted simply as G and $\bar{G}$, respectively.

2. Wavelets bases of $\mathbf{L^2}(\mathbb{R}^d)$ from filters

A fast wavelet transform computes wavelet fields $\bar{\varphi}_j$ as a cascade of filtering and subsamplings with the wavelet filters G and $\bar{G}$. We explain that these wavelet fields can be rewritten as decomposition coefficients in a wavelet orthonormal basis of the Hilbert space $\mathbf{L^2}(\mathbb{R}^d)$ of square-integrable functions, $\int |f(x)|^2 dx < \infty$. These wavelet bases are obtained by dilations and translations of wavelet functions, which result from the cascade of wavelet filters G and $\bar{G}$.

a. Wavelet bases

The fast wavelet transform computes φ_j and $\bar{\varphi}_j$ by iterating j times on G and $\bar{G}$ from φ_0. Discrete wavelets are the equivalent filters which relate φ_j and $\bar{\varphi}_j$ to φ_0. We first set all normalization factors $\gamma_j = 1$ for simplicity. Since $\varphi_j = G\varphi_{j-1}$, where G is a convolution and subsampling operator, we get

$$\varphi_j = (G)^j \varphi_0. \tag{A9}$$

Since G is convolution and subsampling by 2, one can verify that $(G)^j$ computes the inner product with an equivalent filter denoted by $\tilde{\psi}_j^0$, which is translated at intervals 2^j:

$$\varphi_j(i) = \sum_{i'} \varphi_0(i')\tilde{\psi}_j^0(i' - 2^j i).$$

Similarly, since $\bar{\varphi}_j = \bar{G}\varphi_{j-1}$, we get

$$\bar{\varphi}_j = \bar{G}(G)^{j-1}\varphi_0. \tag{A10}$$

We thus verify that wavelet fields are computed as inner products with $2^d - 1$ different wavelet filters $\tilde{\psi}_j^m$ translated at intervals 2^j,

$$\bar{\varphi}_j(m, i) = \sum_{i'} \varphi_0(i')\tilde{\psi}_j^m(i' - 2^j i).$$

Note that the superscript m on $\tilde{\psi}_j^m$ specifies the wavelet channels, and it is not an exponent. The same convention is used below.

The multiresolution theory in Ref. [44] proves that $2^{-jd/2}\tilde{\psi}_j^0(2^{-j}i)$ converges to a scaling function $\psi^0(x)$ with $x \in \mathbb{R}^d$, whose Fourier transform is

$$\hat{\psi}^0(k) = \prod_{p=1}^{\infty} \frac{\hat{G}(2^{-p}k)}{\sqrt{2}}. \tag{A11}$$

Moreover, each discrete wavelet $2^{-jd/2}\tilde{\psi}_j^m(2^{-j}i)$ for $1 \leq m \leq 2^d - 1$ converges to a wavelet $\psi^m(x)$ with $x \in \mathbb{R}^d$, whose Fourier transform is

$$\hat{\psi}^m(k) = \frac{\hat{\bar{G}}_m(2^{-1}k)}{\sqrt{2}}\hat{\psi}^0(2^{-1}k). \tag{A12}$$

The function $\psi^0(x)$ is an averaging filter called a scaling function which satisfies $\int \psi^0(x)dx = 1$, whereas $\psi^m(x)$ for $1 \leq m \leq 2^d - 1$ are called wavelet functions which satisfy $\int \psi^m(x)dx = 0$. Dilated and translated wavelets are written by

$$\psi^m_{j,i}(x) = 2^{-jd/2}\psi^m(2^{-j}x - i) \tag{A13}$$

for $0 \le m \le 2^d - 1$. The main result proved in Refs. [44,45] is that the family of wavelets

$$\{\psi^m_{j,i}(x)\}_{1\le m\le 2^d-1, j\in\mathbb{Z}, i\in\mathbb{Z}^d} \tag{A14}$$

is an orthonormal basis of $\mathbf{L^2}(\mathbb{R}^d)$.

In the renormalization-group decomposition, the wavelet transform is normalized by γ_j in Eq. (4). One can prove [44] that at any scale 2^j, the family of scaling functions $\{\psi^0_{j,i}(x)\}_{i\in\mathbb{Z}^d}$ is also orthonormal within $\mathbf{L^2}(\mathbb{R}^d)$. For any microscopic field φ_0 on a discrete lattice, one can verify that there exists $\varphi \in \mathbf{L^2}(\mathbb{R}^d)$ such that for any $j \ge 0$ the coarse-graining approximations φ_j provide the renormalized decomposition coefficients of φ in these bases of scaling functions:

$$\varphi_j(i) = a_j^{-1}\int_{\mathbb{R}^d}\varphi(x)\psi^0_{j,i}(x)dx, \tag{A15}$$

where $a_j = \prod_{\ell=1}^{j}\gamma_\ell$ is the renormalization factor. If φ_0 is stationary over $\mathbb{Z}^d$ then $\varphi(x)$ can be defined as a stationary process in $\mathbb{R}^d$. Similarly, for $1 \le m \le 2^d - 1$, the wavelet fields are normalized decomposition coefficients of $\varphi(x)$ in the wavelet orthonormal basis:

$$\bar{\varphi}_j(m,i) = a_j^{-1}\int_{\mathbb{R}^d}\varphi(x)\psi^m_{j,i}(x)dx. \tag{A16}$$

The coarse-grained fields $\varphi_j(i)$ and wavelet fields $\bar{\varphi}_j(m,i)$ computed by a fast wavelet transform thus correspond to decomposition coefficients of a field $\varphi(x)$ over orthogonal functions in $\mathbf{L^2}(\mathbb{R}^d)$. This result is important to understand the action of operators over such fields when computing energy functions.

b. Choice of wavelet to represent energy functions

We now explain how to choose wavelet filters for a WCRG. Energy functions involve differential operators such as Laplacians or gradients, but they also include pointwise nonlinearities as in the φ^4 model. Differential operators are diagonal on a Fourier basis, whereas polynomial pointwise nonlinearities are local in the spatial domain (real space) but produce global interactions between Fourier coefficients. Both types of operators induce local interactions over wavelet coefficients if the wavelets $\psi^m(x)$ are sufficiently well localized in the spatial domain, and if their Fourier transform $\hat{\psi}^m(k)$ is also well localized along wave vectors.

The wavelets $\psi^m(x)$ of a wavelet orthonormal basis are entirely specified by the one-dimensional filter $G^{(1)}$ which satisfies Eqs. (A1) and (A2). Indeed, we derive $\bar{G}^{(1)}$ with Eq. (A3), the separable d-dimensional filters $G^{(d)}$ and $\bar{G}^{(d)}$ with Eqs. (A7) and (A8), and each $\hat{\psi}^m$ with Eqs. (A11) and (A12). Therefore, the wavelet properties are adjusted with an appropriate choice of filter $G^{(1)}$.

If $G^{(1)}$ is a one-dimensional filter having $s+1$ nonzero coefficients, then one can verify that each $\psi^m(x)$ has a compact support of width s [46]. Moreover if its Fourier transform satisfies $\hat{G}^{(1)}(k) = \sqrt{2} + \mathcal{O}(|k|^q)$, then one can verify that $\hat{\psi}^m(k)$ for $1 \le m \le 2^d - 1$ satisfies $|\hat{\psi}^m(k)| = \mathcal{O}(|k|^q)$ at low wave vectors k. The integer q is called the number of vanishing moments of ψ^m because this last property implies that ψ^m is orthogonal to any polynomial $P(x)$ of degree strictly less than q: $\int\psi^m(x)P(x)dx = 0$. For a fixed q, a Daubechies filter [46] is a filter $G^{(1)}$ satisfying Eqs. (A1) and (A2), having q vanishing moments and a support of minimum size $s = 2q - 1$.

Haar wavelets are defined by the Haar filter in Eq. (A5). In dimension $d = 1$ (hence, $m = 1$ and we drop off the superscript), one can verify that it defines a Haar wavelet $\psi(x) = \mathbf{1}_{[0,1/2)} - \mathbf{1}_{[1/2,1)}$. It is discontinuous, with a compact support of size $s = 1$ and $q = 1$ vanishing moment. The Kadanoff scheme corresponds to decomposition on a Haar wavelet basis. Yet it is not sufficiently well localized in the Fourier domain to accurately approximate singular differential operators with a nearly diagonal matrix. Wilson instead performed approximate RG calculations [25] using Shannon wavelets obtained with the Shannon filter in Eq. (A6). In dimension $d = 1$, the Fourier transform of ψ is $\hat{\psi}(k) = \mathbf{1}_{[-2\pi,-\pi)} + \mathbf{1}_{[\pi,2\pi)}$. It is infinitely differentiable and has an infinite number of vanishing moments. It is therefore well localized in the Fourier domain and provides a nearly diagonal approximation of differential operators. However, these wavelets have infinite support with a slow spatial decay in real space and are not absolutely integrable. Pointwise polynomial nonlinearities thus produce long-range interactions over Shannon wavelet coefficients, as in the Fourier case.

Haar and Shannon wavelets can be interpreted as Daubechies wavelets having, respectively, $q = 1$ and $q = \infty$ vanishing moments. To obtain accurate approximations of differential operators and pointwise nonlinearities requires us to choose $1 < q < \infty$. In this paper, numerical calculations are performed with a Daubechies wavelet having $q = 4$ vanishing moments, which is called a db4 wavelet (or D8 wavelet). Figure 3 shows the graph of a Haar, Shannon, and Daubechies wavelet having four vanishing moments in $d = 1$ dimension.

3. Proof of Theorem IV.1

Theorem IV.1 considers a singular covariance operator whose eigenvalues $\lambda_\varphi(k)$ have a power-law decay $|k|^{-\zeta}$ and shows that it is represented by nearly diagonal matrices in a wavelet basis, for wavelets having $q \ge \zeta/2$ vanishing moments. Because the wavelet fields are normalized, it proves that the eigenvalues $\lambda_{\bar{\varphi}_j}$ of the covariance of $\bar{\varphi}_j$ have

a lower bound $A > 0$ and an upper bound $B < \infty$ which do not depend upon j.

We first compute A and then B. We write $\lambda_{\varphi_{j-1}}$ the eigenvalues of the covariance of φ_{j-1} at the site n and n',

$$C_{\varphi_{j-1}}(n,n') = \langle \varphi_{j-1}(n)\varphi_{j-1}(n')\rangle_{p_{j-1}}.$$

Since $\bar{\varphi}_j = \gamma_j^{-1}\bar{G}\varphi_{j-1}$ and $\bar{G}$ satisfies the unitary condition in Eq. (7), we verify that the wavelet field $\bar{\varphi}_j$ is obtained from φ_{j-1} with an orthogonal projection weighted by γ_j^{-1}. As a result, the covariance of $\bar{\varphi}_j$ has eigenvalues between the minimum and maximum eigenvalues of the covariance of φ_{j-1} multiplied by γ_j^{-2} and hence,

$$\inf\{\lambda_{\bar{\varphi}_j}\} \geq \gamma_j^{-2}\inf\{\lambda_{\varphi_{j-1}}\}. \tag{A17}$$

A lower bound of $\lambda_{\bar{\varphi}_j}$ is thus obtained by computing a lower bound of $\lambda_{\varphi_{j-1}}$. We see in Eq. (A15) that

$$\varphi_j(i) = a_j^{-1}\int_{\mathbb{R}^d}\varphi(x)\psi^0_{j,i}(x)dx,$$

where $a_j = \prod_{\ell=1}^{j}\gamma_\ell$, and we get

$$C_{\varphi_j}(n,n') = \frac{a_j^{-2}}{(2\pi)^d}\int_{\mathbb{R}^d}\lambda_\varphi(k)\hat{\psi}^0_{j,n}(k)\hat{\psi}^0_{j,n'}(-k)dk.$$

Since $\psi^0_{j,n}(x) = 2^{-dj/2}\psi^0(2^{-j}x-n)$ and $\lambda_\varphi(k) = c|k|^{-\zeta}$, we get

$$C_{\varphi_j}(n,n') = \frac{ca_j^{-2}2^{j\zeta}}{(2\pi)^d}\int_{\mathbb{R}^d}|k|^{-\zeta}|\hat{\psi}^0(k)|^2 e^{ik(n-n')}dk. \tag{A18}$$

By rewriting this integral by a sum of integrals over $[-\pi,\pi]^d$, this last integral can be rewritten as

$$C_{\varphi_j}(n,n') = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\lambda_{\varphi_j}(k)e^{ik(n-n')}dk, \tag{A19}$$

where its eigenvalues for $k\in[-\pi,\pi]^d$ are

$$\lambda_{\varphi_j}(k) = ca_j^{-2}2^{j\zeta}\sum_{\ell\in\mathbb{Z}^d}|k+2\ell\pi|^{-\zeta}|\hat{\psi}^0(k+2\ell\pi)|^2. \tag{A20}$$

By selecting the first term $\ell = 0$ in the sum Eq. (A20), we derive that for $k\in[-\pi,\pi]^d$,

$$\lambda_{\varphi_j}(k) \geq ca_j^{-2}2^{j\zeta}\Gamma\pi^{-\zeta}, \tag{A21}$$

where $\Gamma = \inf_{k\in[-\pi,\pi]^d}\{|\hat{\psi}^0(k)|^2\}$. One can prove that $\Gamma > 0$, because $\hat{\psi}^0(k)$ is continuous, and Eq. (A2) with Eq. (A11) guaranties that $\hat{\psi}^0(k)$ does not vanish on $[-\pi,\pi]^d$.

We derive a lower bound of $\lambda_{\bar{\varphi}_j}$ from Eqs. (A17) and (A21) by inserting $a_j = \gamma_j a_{j-1}$,

$$\inf\{\lambda_{\bar{\varphi}_j}\} \geq ca_j^{-2}2^{(j-1)\zeta}\Gamma\pi^{-\zeta}. \tag{A22}$$

Let us now compute an upper bound of the eigenvalues $\lambda_{\bar{\varphi}_j}$ of the covariance of $\bar{\varphi}_j$. The wavelet field $\bar{\varphi}_j$ has zero average $\langle\bar{\varphi}_j(m,n)\rangle_{p_{j-1}} = 0$ and its covariance is thus,

$$C_{\bar{\varphi}_j}(m,m',n,n') = \langle\bar{\varphi}_j(m,n)\bar{\varphi}_j(m',n')\rangle_{p_{j-1}}.$$

Similar to Eq. (A19), we verify that $C_{\bar{\varphi}_j}(m,m',n,n') = C_{\bar{\varphi}_j}(m,m',n-n')$ and hence that for fixed m, m', it is diagonalized in a Fourier basis. Let us write $\tilde{\lambda}_{\bar{\varphi}_j}(m,m',k)$ the eigenvalue at a wave vector $k\in[-\pi,\pi]^d$. To compute the eigenvalues of $C_{\bar{\varphi}_j}(m,m',n,n')$, for each k we must also diagonalize the matrix $\tilde{\lambda}_{\bar{\varphi}_j}(m,m',k)$ along m, m'. An upper bound of these eigenvalues is obtained by computing the trace of this matrix since all eigenvalues are positive

$$\sup\{\lambda_{\bar{\varphi}_j}\} \leq \sup_k\left\{\sum_{m=1}^{2^d-1}\tilde{\lambda}_{\bar{\varphi}_j}(m,m,k)\right\}. \tag{A23}$$

Similar to Eq. (A20), we verify that for $k\in[-\pi,\pi]^d$,

$$\tilde{\lambda}_{\bar{\varphi}_j}(m,m,k) = ca_j^{-2}2^{j\zeta}\sum_{\ell\in\mathbb{Z}^d}|k+2\ell\pi|^{-\zeta}|\hat{\psi}^m(k+2\ell\pi)|^2. \tag{A24}$$

We then have an inequality,

$$\tilde{\lambda}_{\bar{\varphi}_j}(m,m,k) \leq ca_j^{-2}2^{j\zeta}\Bigg(|k|^{-\zeta}|\hat{\psi}^m(k)|^2 + \pi^{-\zeta}\sum_{\ell\in\mathbb{Z}^d-\{0\}}|\hat{\psi}^m(k+2\ell\pi)|^2\Bigg). \tag{A25}$$

The first term is uniformly bounded for all $k\in[-\pi,\pi]^d$, because the wavelets ψ^m have $q \geq \zeta/2$ vanishing moments and hence $|\hat{\psi}^m(k)| = \mathcal{O}(|k|^q)$ for k in the neighborhood of 0. To control the second term, observe that $\{\psi^m(x-i)\}_{i\in\mathbb{Z}^d}$ is an orthonormal family and hence,

$$\int\psi^m(x-i)\psi^m(x)dx = \delta(i),$$

where $\delta(i)$ is the Kronecker delta. Computing the Fourier transform along i of this equality gives for all $k\in[-\pi,\pi]^d$,

$$\sum_{\ell\in\mathbb{Z}^d}|\hat{\psi}^m(k+2\ell\pi)|^2 = 1. \tag{A26}$$

We derive from Eq. (A25) that there exists a finite Γ' such that for all $k \in [-\pi, \pi]^d$,

$$\tilde{\lambda}_{\bar{\varphi}_j}(m, m, k) \leq c a_j^{-2} 2^{j\zeta} \Gamma'.$$

Inserting this inequality in Eq. (A23) proves that

$$\sup\{\lambda_{\bar{\varphi}_j}\} \leq (2^d - 1) c a_j^{-2} 2^{j\zeta} \Gamma'. \tag{A27}$$

To finish the proof, we relate a_j to $2^{j\zeta}$. The normalization in Eq. (30) implies that

$$\sum_{m=1}^{2^d-1} \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \tilde{\lambda}_{\bar{\varphi}_j}(m, m, k) dk = \sum_{m=1}^{2^d-1} \langle |\bar{\varphi}_j(m, n)|^2 \rangle = 1. \tag{A28}$$

The equality in Eq. (A24) gives

$$a_j^2 2^{-j\zeta} = c\rho \tag{A29}$$

with

$$\rho = \frac{1}{(2\pi)^d} \sum_{m=1}^{2^d-1} \int_{\mathbb{R}^d} |k|^{-\zeta} |\hat{\psi}^m(k)|^2 dk.$$

The constant ρ is finite and strictly positive because of the vanishing moment condition which imposes that $|\hat{\psi}^m(k)|^2 = \mathcal{O}(|k|^{2q})$ with $q \geq \zeta/2$ and because each wavelet is normalized and hence,

$$\|\psi^m\|^2 = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\hat{\psi}^m(k)|^2 dk = 1.$$

Inserting this in Eq. (A27), we prove with Eq. (A23) that

$$\sup\{\lambda_{\bar{\varphi}_j}\} \leq (2^d - 1) \rho^{-1} \Gamma' = B, \tag{A30}$$

which finishes the proof of the upper bound. Inserting Eq. (A29) in the lower bound Eq. (A22) gives

$$\inf\{\lambda_{\bar{\varphi}_j}\} \geq \rho^{-1} \Gamma (2\pi)^{-\zeta} = A > 0, \tag{A31}$$

which finishes the theorem proof.

APPENDIX B: SCALE INTERACTION ENERGIES

1. Scale interaction potential and coupling across scales

This appendix proves the expression Eq. (37) of the interaction potential $\bar{E}_j = \bar{\theta}_j^T \bar{U}_j$.

Equations (36) and (35) imply that

$$\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) = \theta_{j-1}^T U_{j-1}(\varphi_{j-1}) - \theta_{j-1}^T U_{j-1}(P\varphi_{j-1}),$$

where $P = G^T G$ and

$$\theta_{j-1}^T U_{j-1}(\varphi_{j-1}) = \frac{1}{2} \varphi_{j-1}^T K_{j-1} \varphi_{j-1} + C_{j-1}^T V(\varphi_{j-1}).$$

The result is that

$$\bar{\theta}_j^T \bar{U}_j(\varphi_{j-1}) = A + C_{j-1}^T \Big(V(\varphi_{j-1}) - V(P\varphi_{j-1}) \Big) \tag{B1}$$

with

$$A = \frac{1}{2} \varphi_{j-1}^T K_{j-1} \varphi_{j-1} - \frac{1}{2} \varphi_{j-1}^T G^T G K_{j-1} G^T G \varphi_{j-1}.$$

With Eqs. (4) and (8), we get

$$\begin{aligned} A &= \frac{1}{2} \varphi_{j-1}^T K_{j-1} \varphi_{j-1} - \frac{\gamma_j^2}{2} \varphi_j^T G K_{j-1} G^T \varphi_j \\ &= \frac{\gamma_j^2}{2} \Big(2\bar{\varphi}_j^T \bar{G} K_{j-1} G^T \varphi_j + \bar{\varphi}_j^T \bar{G} K_{j-1} \bar{G}^T \bar{\varphi}_j \Big). \end{aligned} \tag{B2}$$

We note that $\varphi_j^T G K_{j-1} \bar{G}^T \bar{\varphi}_j = \bar{\varphi}_j^T \bar{G} K_{j-1} G^T \varphi_j$ because K_{j-1} is symmetric, $K_{j-1}^T = K_{j-1}$. By iterating on Eq. (8) we can decompose φ_j into wavelet fields at all scales

$$\varphi_j = \sum_{\ell=j+1}^{J} a_{j+1,\ell} (G^T)^{\ell-j-1} \bar{G}^T \bar{\varphi}_\ell + a_{j+1,J} (G^T)^{J-j} \bar{\varphi}_{J+1}$$

with $a_{j+1,\ell} = \prod_{\ell'=j+1}^{\ell} \gamma_{\ell'}$ and $\bar{\varphi}_{J+1} = \varphi_J$. Inserting this equation in Eq. (B2) gives

$$A = \frac{1}{2} \sum_{\ell=j}^{J+1} \bar{\varphi}_j^T \bar{K}_{j,\ell} \bar{\varphi}_\ell \tag{B3}$$

with

$$\begin{aligned} \bar{K}_{j,j} &= \gamma_j^2 \bar{G} K_{j-1} \bar{G}^T, \\ \bar{K}_{j,J+1} &= 2\gamma_j^2 a_{j+1,J} \bar{G} K_{j-1} (G^T)^{J-j+1}, \end{aligned} \tag{B4}$$

and for $j + 1 \leq \ell \leq J$,

$$\bar{K}_{j,\ell} = 2\gamma_j^2 a_{j+1,\ell} \bar{G} K_{j-1} (G^T)^{\ell-j} \bar{G}^T.$$

Inserting Eq. (B3) in Eq. (B1) proves Eq. (37) with $\bar{C}_j = C_{j-1}$.

2. Scale interaction free-energy calculation

This section computes a regression of the scale interaction free energy in Eq. (32) with a model $\tilde{\theta}_j^T \tilde{U}_j(\varphi_j)$ with a thermodynamic integration. We see in Eq. (61) that the mean-square error is minimized for

$$\tilde{\theta}_j^\star = \langle \tilde{U}_j \tilde{U}_j^T \rangle_{p_j}^{-1} \langle \bar{F}_j \tilde{U}_j \rangle_{p_j},$$

which requires us to estimate $\langle \bar{F}_j^2 \rangle_{p_j}$ and $\langle \bar{F}_j \tilde{U}_j \rangle_{p_j}$. We estimate these two moments with an empirical average of $\bar{F}_j^2(\varphi_{j,r})$ and $\bar{F}_j(\varphi_{j,r})\tilde{U}_j(\varphi_{j,r})$ over R samples $\varphi_{j,r}$ of p_j computed from R samples $\varphi_{0,r}$ of p_0.

Each $\bar{F}_j(\varphi_{j,r})$ is calculated with a thermodynamic integration. As explained in Eq. (63), we introduce a parametrized energy model which isolates a quadratic term in $\bar{\varphi}_j$ from the potential term multiplied by $\lambda \in [0, 1]$:

$$\bar{\theta}_j^T \bar{U}_{j,\lambda}(\varphi_{j-1}) = Q_{\varphi_j}(\bar{\varphi}_j) + \lambda \bar{C}_j^T V(\varphi_{j-1}), \tag{B5}$$

so that for $\lambda = 0$ we get a Gaussian energy and for $\lambda = 1$ we recover the full-scale interaction energy in Eq. (B1). It results from Eq. (B2) that Q_{φ_j} can be written by

$$Q_{\varphi_j}(\bar{\varphi}_j) = \frac{\gamma_j^2}{2}(2\bar{\varphi}_j^T K_j^{\mathrm{HL}} \varphi_j + \bar{\varphi}_j^T K_j^{\mathrm{HH}} \bar{\varphi}_j) - \bar{C}_j^T V(P\varphi_{j-1}), \tag{B6}$$

with

$$K_j^{\mathrm{HL}} = \bar{G} K_{j-1} G^T \quad \text{and} \quad K_j^{\mathrm{HH}} = \bar{G} K_{j-1} \bar{G}^T.$$

The scale interaction energy model in Eq. (B5) defines a conditional probability parametrized by λ,

$$\bar{p}_{\bar{\theta}_j,\lambda}(\bar{\varphi}_j|\varphi_j) = \frac{e^{-\bar{\theta}_j^T \bar{U}_{j,\lambda}(\varphi_{j-1})}}{\bar{Z}_{j,\lambda}(\varphi_j)},$$

where $\bar{Z}_{j,\lambda}(\varphi_j)$ is the normalization factor. The thermodynamic integration method computes the free energy by integrating a derivative in λ, which yields the following equation:

$$\bar{F}_j(\varphi_j) = \bar{F}_{j,\lambda=1}(\varphi_j) = \bar{F}_{j,\lambda=0}(\varphi_j) + \int_0^1 \left\langle \bar{C}_j^T V(\varphi_{j-1}) \right\rangle_{\bar{p}_{\bar{\theta}_j,\lambda}} d\lambda, \tag{B7}$$

where the free energy for $\lambda = 0$ is

$$\bar{F}_{j,\lambda=0}(\varphi_j) = -\log \int e^{-Q_{\varphi_j}(\bar{\varphi}_j)} d\bar{\varphi}_j.$$

By inserting Eq. (B6), a Gaussian integral calculation gives

$$\bar{F}_{j,\lambda=0}(\varphi_j) = \frac{1}{2}\varphi_j^T M_j \varphi_j - \bar{C}_j^T V(P\varphi_{j-1}) + \tilde{c}_j \tag{B8}$$

with

$$M_j = \gamma_j^2 (K_j^{\mathrm{HL}})^T (K_j^{\mathrm{HH}})^{-1} K_j^{\mathrm{HL}}.$$

The constant $\tilde{c}_j$ is not calculated because we compute the free energy up to an additive constant. As previously explained, the calculation of $\bar{F}_j(\varphi_j)$ is stable because it is performed over the wavelet field $\bar{\varphi}_j$ whose covariance is not singular. As a result, we can approximate the integral over λ by a Riemann sum with few terms.

Numerically, for each known realization $\varphi_{j,r}$ of p_j, we compute $\bar{F}_j(\varphi_{j,r})$ with Eq. (B7). We first evaluate $\bar{F}_{j,\lambda=0}(\varphi_{j,r})$ with Eq. (B8). We then approximate the integral over λ in Eq. (B7) with a Riemann sum of $\langle \bar{C}_j^T V(\varphi_{j-1}) \rangle_{\bar{p}_{\bar{\theta}_j,\lambda}}$ with about ten values of λ in [0, 1]. For each fixed λ, each expected value is estimated by computing a chain of samples $\bar{\varphi}_{j,\lambda}^{(t)}$ of $\bar{p}_{\bar{\theta}_j,\lambda}(\bar{\varphi}_j|\varphi_{j,r})$, with wavelet MC updates at each time t. We then define

$$\varphi_{j-1,r,\lambda}^{(t)} = \gamma_j G^T \varphi_{j,r} + \gamma_j \bar{G}^T \bar{\varphi}_{j,\lambda}^{(t)},$$

and for each $\varphi_{j,r}$ we perform an empirical average of all $\bar{C}_j^T V(\varphi_{j-1,r,\lambda}^{(t)})$ along the time variable t of the MCMC chain. Applying Eq. (B7) provides an estimation of $\bar{F}_j(\varphi_{j,r})$ for each $\varphi_{j,r}$, from which we derive $\langle \bar{F}_j^2 \rangle_{p_j}$ and $\langle \bar{F}_j \tilde{U}_j \rangle_{p_j}$ with empirical averages along r.

APPENDIX C: METHODS FOR NUMERICAL EXPERIMENTS

This appendix contains numerical details on the three type of fields that are studied: Gaussians, φ^4, and weak-gravitational-lensing maps. For each system, we explain the numerical algorithms and give additional data supporting our conclusions.

1. Gaussian field theory

a. Model

The Gaussian model is a particular case of a local energy model. The potential and coupling parameters in Eqs. (17) and (18) are defined by setting the nonlinear potential to zero: $U_j = (\varphi_j \varphi_j^T, 0)$ and $\theta_j = (\frac{1}{2}K_j, 0)$. In this paper, we consider the Ornstein-Uhlenbeck Gaussian stochastic process introduced in Sec. VI A. We define the model in the Fourier space on a lattice given by

$$\hat{K}_0(k) = \beta(|k|^2 + (2\pi/\xi)^2), \tag{C1}$$

where ξ is the correlation length that uniquely characterizes the process. The overall factor β is determined such that $\langle|\varphi_0(i)|^2\rangle_{p_0} = 1$. We study systems with the size $L = 8$, 16, 32, and 64, varying $\xi = 2$, 4, 8, 16, 32, and 64.

b. Direct coupling estimation

The gradient decent dynamics of the direct coupling estimation in Eq. (41) is given by

$$\theta_j^{(t+1)} - \theta_j^{(t)} = \epsilon\Big(\langle U_j\rangle_{p_{\theta_j^{(t)}}} - \langle U_j\rangle_{p_j}\Big). \tag{C2}$$

For each time step t, $\theta_j^{(t)} = (\frac{1}{2}K_j^{(t)}, 0)$ defines a Gaussian field whose covariance is the inverse of $K_j^{(t)}$,

$$\langle U_j\rangle_{p_{\theta_j^{(t)}}} = \langle\varphi_j\varphi_j^T\rangle_{p_{\theta_j^{(t)}}} = \Big(K_j^{(t)}\Big)^{-1}. \tag{C3}$$

We thus compute $\langle U_j\rangle_{p_{\theta_j^{(t)}}}$ by inverting $K_j^{(t)}$ instead of using an MCMC algorithm. For the initialization at $t = 0$, we set $K_j^{(t=0)} = \sigma_j^{-2}\mathrm{Id}$, where $\sigma_j^2 = 1$ is the normalized variance of φ_j.

Figure 19(a) shows decay of a normalized error $\mathcal{L}(t) = \|K_0^{(t)} - K_0\|_s/\|K_0\|_s$, where $\|\cdot\|_s$ is the spectral norm of a matrix. For $L = 32$, the decay becomes progressively slower when ξ increases. The timescale τ_{GD} of the gradient decent dynamics is defined by $\mathcal{L}(\tau_{\mathrm{GD}}) = 10^{-1.5} \simeq 0.03$. Figure 19(c) plots τ_{GD} versus ξ. The decay rate is controlled by the maximum step size $\epsilon_{\max}$ of ϵ in Eq. (C2), where $\epsilon_{\max}$ is the inverse of the maximum eigenvalue $\lambda_{\max}$ of the Hessian matrix H_{θ_0} in Eq. (42). For a Gaussian model, H_{θ_0} contains only the fourth-order moments of φ_0, and they can be written by the product of the second-order moments (Wick's theorem). It thus indicates that τ_{GD} varies like

$$\tau_{\mathrm{GD}} \sim \epsilon_{\max}^{-1} = \lambda_{\max} \sim \Big(\hat{K}_0^{-1}(k=0)\Big)^2 \sim \xi^4. \tag{C4}$$

The numerical results in Fig. 19(c) follow this asymptotic growth. The gradient-descent dynamics of the direct coupling estimation has a critical slowing-down when the spatial correlation increases. Figure 19(d) shows the critical scaling by plotting τ_{GD} as a function of L, at the critical point $\xi = L$. This scaling is worse than the mixing timescale of a direct MC, namely, $\tau_{\mathrm{MC}} \sim \mathcal{O}(\xi^z)$ when $\xi \approx L$ with $z = 2$.

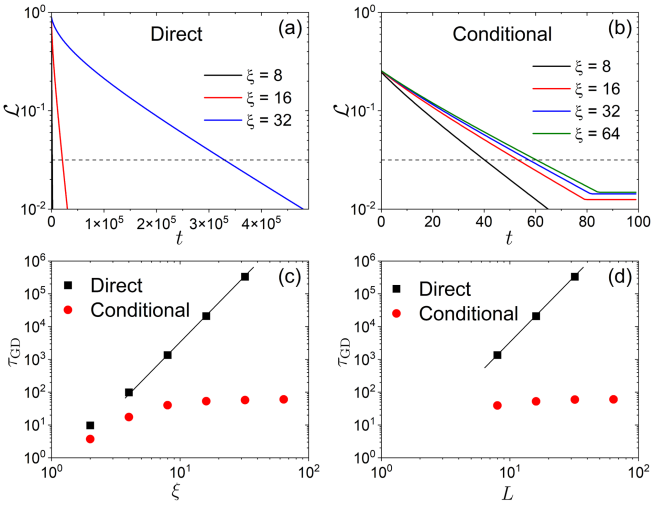

FIG. 19. (a),(b) Decay of the relative error $\mathcal{L}(t)$ for the direct coupling estimation with $L = 32$ in (a), and for the conditional coupling estimation with $L = 64$ in (b), for several ξ up to critical point $\xi = L$. The decay timescale τ_{GD} is defined by $\mathcal{L}(\tau_{\mathrm{GD}}) = 10^{-1.5}$ (horizontal dashed line). (c),(d) Evolution of τ_{GD} from the direct and conditional coupling estimations as a function of ξ for L fixed in (c), and at the critical point $L = \xi$ in (d). The solid straight lines correspond to $\tau_{\mathrm{GD}} \sim \xi^4$ in (c) and $\tau_{\mathrm{GD}} \sim L^4$ in (d).

c. Conditional coupling estimation

The gradient-decent dynamics for the conditional coupling parameters $\bar{\theta}_j$ is computed by Eq. (46),

$$\bar{\theta}_j^{(t+1)} - \bar{\theta}_j^{(t)} = \epsilon\Big(\langle\bar{U}_j\rangle_{p_{\bar{\theta}_j^{(t)}}} - \langle\bar{U}_j\rangle_{p_{j-1}}\Big), \tag{C5}$$

where $\bar{U}_j = (\bar{\varphi}_j\bar{\varphi}_j^T, \bar{\varphi}_{j+1}\bar{\varphi}_j^T, 0)$. Since $p_{\bar{\theta}_j^{(t)}}(\varphi_{j-1}) = \bar{p}_{\bar{\theta}_j^{(t)}}(\bar{\varphi}_j|\varphi_j)p_j(\varphi_j)$ is also Gaussian, the second-order moments $\langle\bar{U}_j\rangle_{p_{\bar{\theta}_j^{(t)}}}$ can be also computed by the matrix inversion without doing an MCMC simulation. In the WCRG framework, the conditional coupling estimation is performed scale per scale, from $j = J$ down to $j = 1$. Figure 19(b) shows the gradient-descent dynamics at $j = 1$ with $L = 64$. It corresponds to the last scale of WCRG and hence to the highest computational cost among all j. We plot the relative error $\mathcal{L}(t) = \|K_0^{(t)} - K_0\|_s/\|K_0\|_s$ as in Fig. 19(a) in order to compare both results. For each time step t, $K_0^{(t)}$ is recovered with Eq. (59). As expected, $\mathcal{L}(t)$ has a much faster decay for the conditional coupling estimation than for the direct one. Note that $\mathcal{L}(t)$ saturates at a small value when t is large. This small error is introduced by the elimination of terms for $\ell > j+1$ in Eq. (38). The decay timescale τ_{GD} is also defined by $\mathcal{L}(\tau_{\mathrm{GD}}) = 10^{-1.5}$. Figure 19(c) plots τ_{GD} versus ξ (with $L = 64$), and Fig. 19(d) plots τ_{GD} versus L (with $L = \xi$). As expected from Fig. 19(b), the decay timescale τ_{GD} of the conditional coupling estimation is much smaller and becomes nearly constant when ξ increases for a fixed L and L increases at the critical point. It shows that the

conditional coupling estimation is not affected by the critical slowing-down.

d. Monte Carlo simulations

Although it is not needed, we sample Gaussian models with MCMC simulation to test the evolution of the mixing time. We have two MC simulation schemes. The first one is a MC using a direct update of φ_j to estimate $\langle U_j \rangle_{p_{\bar{\theta}_j^{(t)}}}$ in the direct coupling estimation in Eq. (C2). The second one is the wavelet MC, which updates $\bar{\varphi}_j$ given φ_j. It is used to estimate $\langle \bar{U}_j \rangle_{p_{\bar{\theta}_j^{(t)}}}$ in the conditional coupling estimation in Eq. (C5) and to compute samples $\bar{\varphi}_j$ in the WCRG sampling algorithm of Sec. III C.

For the direct MC simulation, we make the following proposal for a site i chosen randomly: $\varphi_j(i) \to \varphi_j(i) + \delta(r - 1/2)$, where r is a random number uniformly distributed in (0, 1]. We set $\delta = 6.0$ to minimize the mixing time τ_{MC} (defined below) at the critical point. The proposed update is accepted or rejected with the standard Metropolis-Hastings rule. In order to estimate the decorrelation timescale at $j = 0$, we compute the time autocorrelation function $A_\Phi(t)$ with Eq. (70). A unit MC time step for the direct MC corresponds to L^2 MC trials (one per pixel on average). MC simulations are computed up to 2×10^6 MC steps. The last 10^6 MC steps are used to compute the autocorrelation function $A_\Phi(t)$ by discarding the first 10^6 MC steps. We compute ten independent samples to measure $A_\Phi(t)$. Figure 20(a) shows $A_\Phi(t)$ for several ξ for $L = 32$. The decay of $A_\Phi(t)$ becomes significantly slower when increasing ξ. To evaluate the decay rate, we define a decorrelation (or mixing) timescale τ_{MC} at which $A_\Phi(\tau_{\mathrm{MC}}) = 1/e$. Figure 20(c) shows τ_{MC} versus ξ for $L = 32$, which verifies that $\tau_{\mathrm{MC}} \sim \xi^2$. Figure 20(d) shows a critical slowing-down with $\tau_{\mathrm{MC}} \sim L^2$ when $L = \xi$ increases. In both plots, we find that $\tau_{\mathrm{MC}} \sim \xi^z$ and $\tau_{\mathrm{MC}} \sim L^z$ with $z = 2$. This is justified by an argument given in Sec. IV C 2 for the Ornstein-Uhlenbeck process, which shows that $\tau_{\mathrm{MC}} \sim \hat{K}_0^{-1}(k=0) \sim \xi^2$.

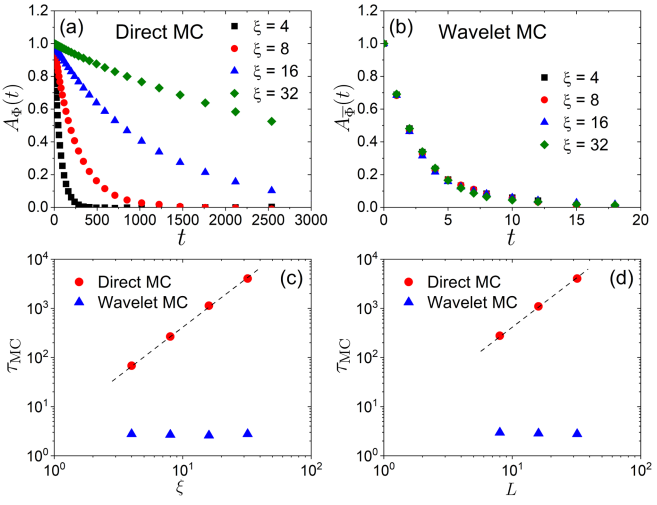

FIG. 20. MCMC simulations at the stationary state for the Gaussian model. (a),(b) Autocorrelation function for direct MC simulations in (a) and for the wavelet MC simulations in (b), for several ξ with $L = 32$. (c),(d) Timescale τ_{MC} at which the autocorrelation is equal to e^{-1}. In (c), ξ varies, whereas $L = 32$. In (d), τ_{MC} is computed at the critical point $\xi = L$ for different L. The dashed straight lines in (c) and (d) correspond to $\tau_{\mathrm{MC}} \sim \xi^2$ and $\tau_{\mathrm{MC}} \sim L^2$, respectively.

e. Wavelet Monte Carlo

A wavelet MC updates only for the wavelet fields $\bar{\varphi}_j$ given φ_j fixed. For each trial update, a channel m and a site i are chosen randomly to propose $\bar{\varphi}_j(m,i) \to \bar{\varphi}_j(m,i) + \delta(r - 1/2)$. We apply the standard Metropolis-Hastings rule by evaluating the scale interaction energy function $\bar{E}_j$. We set $\delta = 12.0$ which minimizes the mixing time τ_{MC} (defined below) at the critical point. We compute the autocorrelation function $A_{\bar{\Phi}}(t)$ with Eq. (71) to estimate the decorrelation timescale. For the wavelet MC case, $3(L/2^j)^2$ MC trials correspond to one MC time step (one per pixel on average). MC simulations are computed over 2×10^3 MC steps. The last 10^3 MC steps are used to compute $A_{\bar{\Phi}}(t)$, and we average over ten independent samples. The decorrelation timescale τ_{MC} is also defined by $A_{\bar{\Phi}}(\tau_{\mathrm{MC}}) = 1/e$. Figure 20(b) plots $A_{\bar{\Phi}}(t)$, which decays much more quickly than with a direct MC in Fig. 20(a). Figure 20(c) shows that τ_{MC} nearly does not depend on ξ. Figure 20(d) also shows that it does not depend on L at the critical point. This is totally different from direct MC simulations. These results confirm that the dynamics of the wavelet fields do not suffer from any critical slowing-down.

2. Two-dimensional φ^4-field theory

a. Model

The second model we consider is a lattice version of the φ^4-field theory introduced in Sec. II A 4, which has been studied widely [47,89–92]. We consider two-dimensional systems of linear size $L = 32$. Let $\Phi = L^{-2}|\sum_i \varphi_0(i)|$. Figure 21 shows that the mean magnetization $\langle \Phi \rangle_{p_0}$ increases with β. The effective critical value of β for $L = 32$ is found to be $\beta_c \simeq 0.67$ by finding the peak of the susceptibility. Because of finite-size effects [93], the magnetization has a smooth crossover around β_c. We focus on four values: $\beta = 0.5$, $\beta = 0.6$, $\beta = 0.67 \simeq \beta_c$, and $\beta = 0.76$, which are shown by vertical dashed line in Fig. 21. It covers disordered, critical, and ordered phases. For each β, we generate $R = 10\,000$ statistically independent samples of the field φ_0, which we use as the training dataset to compute the WCRG.

b. Monte Carlo simulations

Monte Carlo simulations for the φ^4 model are calculated as in the Gaussian model described above, except for the

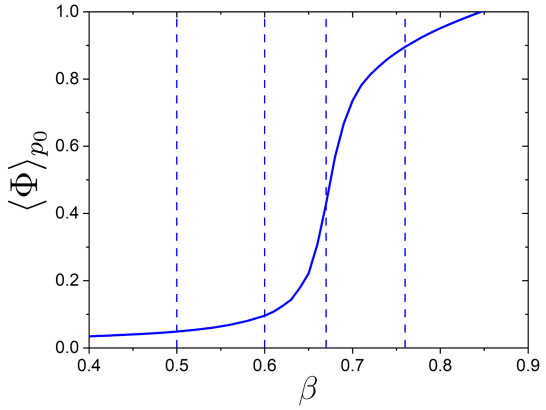

FIG. 21. Mean value of the magnetization of the φ^4 model computed over the training dataset. The four state points studied in this paper are shown by dashed lines: $\beta = 0.50$, 0.60, 0.67, and 0.76.

presence of a nonlinear potential term in E_j and $\bar{E}_j$. To generate the training dataset, a direct MC is computed at the finest scale $j = 0$ with periodic boundary conditions. It updates $\varphi_0(i)$ using MC trials with $\delta = 3.0$ [90,92]. For the direct and conditional coupling estimations in Eqs. (41) and (46), we estimate moments with a direct MC and a wavelet MC, respectively. The value of δ evolves dynamically during simulation so that the acceptance ratio remains nearly 50%. To measure the autocorrelation functions $A_\Phi(t)$ and $A_{\bar{\Phi}}(t)$ in Fig. 17, we fix $\delta = 3.0$ and 1.0, respectively.

c. Coupling parameter estimations

In a direct coupling estimation, the gradient-descent dynamics minimizes the KL divergence with Eq. (41). Prior information on symmetries reduces the number of coupling parameters that needs to be estimated. Because of the translational invariance, K_j is defined by an $L/2^j \times L/2^j$ periodic kernel. We impose that K_j is symmetric to transposition and reflection, which reduces the number of matrix elements to estimate. The nonlinear potential is defined by 36 hat functions when the system is in the ordered broken-symmetry phase with $\beta = 0.76$. When $\beta = 0.5$, 0.6, and 0.67, which correspond to disordered symmetric and critical phases, we use 17 hat functions. For the WCRG approach, the gradient-descent dynamics in Eq. (41) is computed scale per scale, from the coarsest scale, $j = 5$ ($L/2^j = 1$), down to $j = 1$ ($L/2^j = 16$). It is defined by 15 scale interaction matrices $\bar{K}_{j,\ell}$, and each of them is a circulant matrix having transposition and reflection symmetries. We further impose symmetries between different scale interaction matrices originating from rotational symmetries of two-dimensional wavelet transform. For the nonlinear potential, we use the same set of hat functions as the direct coupling estimation at all scales 2^j.

d. Thermodynamic integration and regression

To estimate the energy function E_j at each scale, we compute a thermodynamic integration and a linear regression described in Sec. B 2. The thermodynamic integration [67] starts with the Gaussian model, where the integration can be performed analytically. As we explain in Sec. B 2, we multiply the second term in Eq. (37) by λ and estimate statistical averages of these non-Gaussian terms for each λ. We approximate the integral by a Riemann sum computed over ten values of λ, uniformly distributed in [0, 1]. Figures 10 and 11 compare the estimated energy function $E_0(\varphi_0)$ and the original one for $\beta \simeq \beta_c$.

3. Cosmological data

We use a set of simulated convergence maps computed by the Columbia lensing group [76,77]. These convergence maps are calculated by ray-tracing N-body simulations. The sample dataset is available from Ref. [94]. They simulate convergence maps of the next-generation space telescope *Euclid*. It is a mission of the European Space Agency whose goal is to map the geometry of the Universe and better understand dark matter and dark energy. The cosmology is defined by the matter density parameter $\Omega_m = 0.26$ and the root-mean-square matter fluctuation $\sigma_8 = 0.8$. We subsample the 1024×1024 maps by a factor of 4 with a local averaging, and we extract 32×32 patches from the subsampled dataset. We also introduce a cutoff of the maximum amplitude of peaks, as shown in Fig. 13. The resulting dataset is then normalized to have a mean 0 and variance 1. We use $R = 78126$ images of size 32×32 as a training dataset to compute the WCRG, but $R \simeq 3000$ is enough to get nearly the same results.

The gradient descent of the WCRG is computed with the same setting as the φ^4 model for $\beta = 0.76$, which also has an asymmetric distribution. Namely, we use the same size of scale interaction matrices $\bar{K}_{\ell,j}$ for the three wavelet channels, and we decompose the nonlinear potential over hat functions. The number of hat functions is 23. We also include ReLU functions ($f(x) = x$ for $x > 0$ and equal to zero otherwise) at the edges, with coefficients set to -10 and 10 at the left and right edges, respectively. These ReLU functions confine the field values over an appropriate interval. This is needed for weak-lensing data, which have a long-tail distribution that is nearly Laplacian.

4. Spatial correlation function

We consider the two-point spatial correlation function $c(r)$, which is defined by

$$c(r) = \frac{\langle \delta\varphi_0(r)\delta\varphi_0(0)\rangle}{\langle [\delta\varphi_0(0)]^2\rangle}, \tag{C6}$$

where $\delta\varphi_0$ is the fluctuation of the microscopic field, and the bracket $\langle\cdot\rangle$ denotes averaging over configurations from either the training data or the WCRG sampling (the training

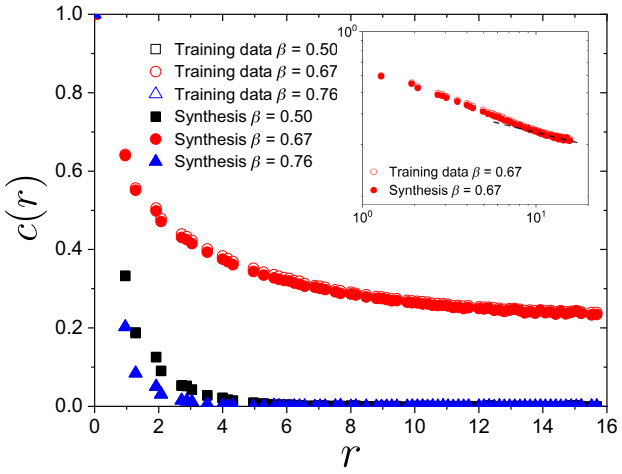

FIG. 22. Two-point spatial correlation function $c(r)$ for the training data (empty symbols) and synthesized ones by the WCRG sampling (filled symbols). The inset shows a log-log plot for $\beta = 0.67 \simeq \beta_c$. The dashed line follows $c(r) \sim r^{-(d-2+\eta)}$ with $\eta = 1/4$ (and $d = 2$).

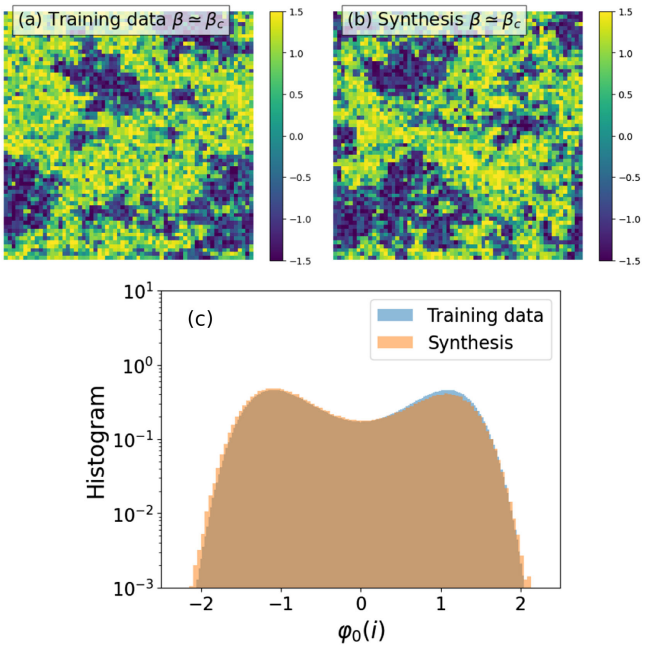

FIG. 23. (a),(b) Original training samples of φ^4 (a) and synthesized fields generated by the WCRG (b) for $L = 64$ at $\beta = 0.67 \simeq \beta_c$. (c) Superimposed histograms of field values computed from φ^4 training data (blue) and WCRG synthesized fields (orange) for $L = 64$ at $\beta \simeq \beta_c$.

data are obtained as done in Sec. VI.B for the other tests on the φ^4-field theory). We present $c(r)$ in Fig. 22 for $\beta = 0.50 < \beta_c$, $\beta = 0.67 \simeq \beta_c$, and $\beta = 0.76 > \beta_c$. Whereas $c(r)$ for $\beta = 0.50$ and $\beta = 0.76$ decay quickly, the one for $\beta = 0.67$ decays very slowly, which is consistent with a power-law decay $c(r) \sim r^{-1/4}$ expected from the Ising universality class in $d = 2$. We numerically confirm that the WCRG sampling correctly reproduces the two-point spatial correlation function from the original training dataset, suggesting that the WCRG method can characterize the important physical features in critical phenomena.

5. Larger system sizes

In the main text, we mainly focus on the systems with the linear size $L = 32$. This system size is large enough to study essential features of critical slowing-down and demonstrate the performance of the WCRG. Here, we study a larger system, the φ^4 model with $L = 64$ at $\beta \simeq \beta_c$, in order to confirm the results of the main text and consolidate our conclusions. Larger systems for the weak-gravitational-lensing maps will be reported in a future publication with a detailed physical discussion. Figure 23 shows comparisons between the synthesized fields by the WCRG sampling and the ones from the training dataset for $L = 64$ in terms of snapshots [Figs. 23(a) and 23(b)] and histogram [Fig. 23(c)]. We confirm that the WCRG sampling works well for the system with $L = 64$. We then estimate the microscopic energy in Fig. 24. Figures 24(a) and 24(b) show the estimated convolution kernel of K_0 and its eigenvalues (compared with the original one), respectively. Figure 24 shows that the estimated local potential at the finest scale is nearly equal

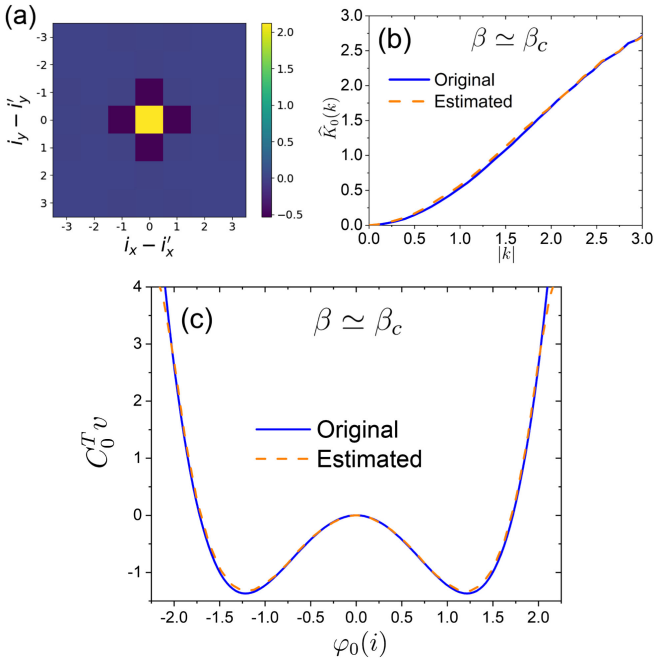

FIG. 24. (a) Estimated convolutional kernel $K_0(i - i')$ in the $(d = 2)$-dimensional space for the φ^4 model. (b) Comparison for $\beta \simeq \beta_c$ of the Fourier eigenvalues $\hat{K}_0(k)$ of K_0 of the original model (solid curve) and the one estimated by the WCRG (dashed curve). (c) Comparison for $\beta \simeq \beta_c$ of the original nonlinear local potential $C_0^T v(\varphi_0(i))$ at the microscopic scale (solid curve) and its WCRG estimation (dashed curve). The results are obtained from $L = 64$.

to the original one. The finding that numerical slowing-down is absent in the generation process for $L = 64$ is reported in the main text. These numerical results demonstrate that the WCRG provides a precise estimation of the φ^4 microscopic energy also at $L = 64$.

[1] N. Perraudin, A. Srivastava, A. Lucchi, T. Kacprzak, T. Hofmann, and A. Réfrégier, *Cosmological N-body simulations: A challenge for scalable generative models*, Comput. Astrophys. Cosmol. **6**, 5 (2019).

[2] J. Carrasquilla and R. G. Melko, *Machine learning phases of matter*, Nat. Phys. **13**, 431 (2017).

[3] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, *Weak pairwise correlations imply strongly correlated network states in a neural population*, Nature (London) **440**, 1007 (2006).

[4] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, *Highly accurate protein structure prediction with* ALPHAFOLD, Nature (London) **596**, 583 (2021).

[5] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721 (1984).

[6] J. Portilla and E. P. Simoncelli, *A parametric texture model based on joint statistics of complex wavelet coefficients*, Int. J. Comput. Vis. **40**, 49 (2000).

[7] S. C. Zhu, Y. N. Wu, and D. Mumford, *Minimax entropy principle and its application to texture modeling*, Neural Comput. **9**, 1627 (1997).

[8] D. Kingma and M. Welling, *Auto-encoding variational Bayes*, in *Proceedings of the ICLR* (2014), https://arxiv.org/abs/1312.6114.

[9] A. Radford, L. Metz, and R. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks*, in *Proceeedings of the ICLR* (2016), https://arxiv.org/abs/1511.06434.

[10] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon Press, Oxford, 2002), Vol. 113.

[11] R. Bellman and R. Kalaba, *A mathematical theory of adaptive control processes*, Proc. Natl. Acad. Sci. U.S.A. **45**, 1288 (1959).

[12] W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, Oxford, 2006), Vol. 13.

[13] P. M. Chaikin, T. C. Lubensky, and T. A. Witten, *Principles of Condensed Matter Physics* (Cambridge University Press Cambridge, England, 1995), Vol. 10.

[14] J. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Oxford University Press, New York, 2021), Vol. 14.

[15] B. Delamotte, *An introduction to the nonperturbative renormalization group*, in *Renormalization Group and Effective Field Theory Approaches to Many-Body Systems*, edited by Achim Schwenk and Janos Polonyi (Springer, New York, 2012), pp. 49–132.

[16] K. Tanaka, S. Kataoka, M. Yasuda, and M. Ohzeki, *Inverse renormalization group transformation in Bayesian image segmentations*, J. Phys. Soc. Jpn, **84**, 045001 (2015).

[17] K. Shiina, H. Mori, Y. Tomita, H. K. Lee, and Y. Okabe, *Inverse renormalization group based on image super-resolution using deep convolutional networks*, Sci. Rep. **11**, 9617 (2021).

[18] S.-H. Li and L. Wang, *Neural network renormalization group*, Phys. Rev. Lett. **121**, 260601 (2018).

[19] P. Mehta and D. J. Schwab, *An exact mapping between the variational renormalization group and deep learning*, arXiv:1410.3831.

[20] J. J. Yu, K. G. Derpanis, and M. A. Brubaker, *Wavelet flow: Fast training of high resolution normalizing flows*, Adv. Neural Inf. Process. Syst. **33**, 6184 (2020).

[21] G. Evenbly and S. R. White, *Entanglement renormalization and wavelets*, Phys. Rev. Lett. **116**, 140403 (2016).

[22] S. Mallat, *Understanding deep convolutional networks*, Phil. Trans. R. Soc. A **374**, 20150203 (2016).

[23] F. Noé, S. Olsson, J. Köhler, and H. Wu, *Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning*, Science **365**, eaaw1147 (2019).

[24] M. Gabrié, G. M. Rotskoff, and E. Vanden-Eijnden, *Efficient Bayesian sampling using normalizing flows to assist Markov chain Monte Carlo methods*, arXiv:2107.08001.

[25] K. G. Wilson, *Renormalization group and critical phenomena. II. Phase-space cell analysis of critical behavior*, Phys. Rev. B **4**, 3184 (1971).

[26] G. Battle, *Wavelets and Renormalization* (World Scientific, Singapore, 1999), Vol. 10.

[27] M. V. Altaisky, *Unifying renormalization group and the continuous wavelet transform*, Phys. Rev. D **93**, 105043 (2016).

[28] A. N. Kolmogorov, *A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number*, J. Fluid Mech. **13**, 82 (1962).

[29] U. Frisch and G. Parisi, *Fully developed turbulence and intermittency*, N. Y. Acad. Sci. Ann. **357**, 359 (1980).

[30] U. Frisch, *From global scaling, à la Kolmogorov, to local multifractal scaling in fully developed turbulence*, Proc. R. Soc. A **434**, 89 (1991).

[31] J. F. Muzy, E. Bacry, and A. Arneodo, *The multifractal formalism revisited with wavelets*, Int. J. Bifurcation Chaos **04**, 245 (1994).

[32] B. B. Mandelbrot, *Multifractals and 1/f Noise: Wild Self-Affinity in Physics (1963–1976)* (Springer, New York, 2013).

[33] E. Bacry, J. Delour, and J. F. Muzy, *Multifractal random walk*, Phys. Rev. E **64**, 026103 (2001).

[34] E. Allys, T. Marchand, J.-F. Cardoso, F. Villaescusa-Navarro, S. Ho, and S. Mallat, *New interpretable statistics for large-scale structure analysis and generation*, Phys. Rev. D **102** (2020).

[35] S. Zhang and S. Mallat, *Maximum entropy models from phase harmonic covariances*, Appl. Comput. Harmon. Anal. **53**, 199 (2021).

[36] S. Cheng and B. Ménard, *Weak lensing scattering transform: Dark energy and neutrino mass sensitivity*, Mon. Not. R. Astron. Soc. **507**, 1012 (2021).

[37] S. Mallat, *A Wavelet Tour of Signal Processing* (Elsevier, 1999).

[38] D. Ron, R. H. Swendsen, and A. Brandt, *Inverse Monte Carlo renormalization group transformations for critical phenomena*, Phys. Rev. Lett. **89**, 275701 (2002).
[39] D. Bachtis, G. Aarts, F. Di Renzo, and B. Lucini, *Inverse renormalization group in quantum field theory*, Phys. Rev. Lett. **128**, 081603 (2022).
[40] L. P. Kadanoff, A. Houghton, and M. C. Yalabik, *Variational approximations for renormalization group transformations*, J. Stat. Phys. **14**, 171 (1976).
[41] K. G. Wilson and M. E. Fisher, *Critical exponents in 3.99 dimensions*, Phys. Rev. Lett. **28**, 240 (1972).
[42] M. E. Fisher, *The renormalization group in the theory of critical behavior*, Rev. Mod. Phys. **46**, 597 (1974).
[43] S. Mallat, *A theory for multiresolution signal decomposition: The wavelet representation*, IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674 (1989).
[44] S. Mallat, *Multiresolution approximations and wavelet orthonormal bases of L^2*, Trans. Am. Math. Soc. **315**, 69 (1989).
[45] Y. Meyer, *Wavelets and Operators* (Cambridge University Press, Cambridge, England, 1992).
[46] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, 1992).
[47] J. Kaupužs, R. Melnik, and J. Rimšāns, *Corrections to finite-size scaling in the φ^4 model on square lattices*, Int. J. Mod. Phys. C **27**, 1650108 (2016).
[48] L. Biferale, *Shell models of energy cascade in turbulence*, Annu. Rev. Fluid Mech. **35**, 441 (2003).
[49] A. A. Mailybaev, *Hidden scale invariance of intermittent turbulence in a shell model*, Phys. Rev. Fluids **6**, L012601 (2021).
[50] J. Bruna and S. Mallat, *Multiscale sparse microcanonical models*, Math. Stat. Learn. **1**, 257 (2019).
[51] E. Allys, T. Marchand, J.-F. Cardoso, F. Villaescusa-Navarro, S. Ho, and S. Mallat, *New interpretable statistics for large-scale structure analysis and generation*, Phys. Rev. D **102**, 103506 (2020).
[52] S. Cheng, R. Morel, E. Allys, B. Menard, and S. Mallat, *Scattering spectra models for physics*, arXiv:2306.17210.
[53] R. H. Swendsen and J.-S. Wang, *Nonuniversal critical dynamics in Monte Carlo simulations*, Phys. Rev. Lett. **58**, 86 (1987).
[54] U. Wolff, *Comparison between cluster Monte Carlo algorithms in the Ising model*, Phys. Lett. B **228**, 379 (1989).
[55] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B **195**, 216 (1987).
[56] J. Goodman and A. D. Sokal, *Multigrid Monte Carlo method. Conceptual foundations*, Phys. Rev. D **40**, 2035 (1989).
[57] G. M. Torrie and J. P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys. **23**, 187 (1977).
[58] E. Marinari and G. Parisi, *Simulated tempering: A new Monte Carlo scheme*, Europhys. Lett. **19**, 451 (1992).
[59] L. Wasserman, *All of Statistics* (Springer, New York, 2021).
[60] A. Batou and C. Soize, *Calculation of Lagrange multipliers in the construction of maximum entropy distributions in high stochastic dimension*, SIAM/ASA J. Uncertainty Quantif. **1**, 431 (2013).
[61] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, England, 1996), Vol. 5.
[62] A. Montanari and G. Semerjian, *Rigorous inequalities between length and time scales in glassy systems*, J. Stat. Phys. **125**, 23 (2006).
[63] P. C. Hohenberg and B. I. Halperin, *Theory of dynamic critical phenomena*, Rev. Mod. Phys. **49**, 435 (1977).
[64] L. Canet and H. Chaté, *A non-perturbative approach to critical dynamics*, J. Phys. A **40**, 1937 (2007).
[65] A. Andreanov, G. Biroli, J.-P. Bouchaud, and A. Lefevre, *Field theories and exact stochastic equations for interacting particle systems*, Phys. Rev. E **74**, 030101(R) (2006).
[66] A. Gelman, W. R. Gilks, and G. O. Roberts, *Weak convergence and optimal scaling of random walk Metropolis algorithms*, Ann. Appl. Probab. **7**, 110 (1997).
[67] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Elsevier, New York, 2001), Vol. 1.
[68] J. Berges, N. Tetradis, and C. Wetterich, *Non-perturbative renormalization flow in quantum field theory and statistical physics*, Phys. Rep. **363**, 223 (2002).
[69] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, Phys. Rep. **340**, 291 (2001).
[70] M. Kilbinger, *Cosmology with cosmic shear observations: A review*, Rep. Prog. Phys. **78**, 086901 (2015).
[71] L. Fu, M. Kilbinger, T. Erben, C. Heymans, H. Hildebrandt, H. Hoekstra, T. D. Kitching, Y. Mellier, L. Miller, E. Semboloni *et al.*, *CFHTLenS: Cosmological constraints from a combination of cosmic shear two-point and three-point correlations*, Mon. Not. R. Astron. Soc. **441**, 2725 (2014).
[72] X. Liu, C. Pan, R. Li, H. Shan, Q. Wang, L. Fu, Z. Fan, J.-P. Kneib, A. Leauthaud, L. Van Waerbeke *et al.*, *Cosmological constraints from weak lensing peak statistics with Canada–France–Hawaii Telescope Stripe 82 Survey*, Mon. Not. R. Astron. Soc. **450**, 2888 (2015).
[73] T. Kacprzak, D. Kirk, O. Friedrich, A. Amara, A. Refregier, L. Marian, J. Dietrich, E. Suchyta, J. Aleksić, D. Bacon *et al.*, *Cosmology constraints from shear peak statistics in dark energy survey science verification data*, Mon. Not. R. Astron. Soc. **463**, 3653 (2016).
[74] N. Martinet, P. Schneider, H. Hildebrandt, H. Shan, M. Asgari, J. P. Dietrich, J. Harnois-Déraps, T. Erben, A. Grado, C. Heymans *et al.*, *KiDS-450: Cosmological constraints from weak-lensing peak statistics – II: Inference from shear peaks using N-body simulations*, Mon. Not. R. Astron. Soc. **474**, 712 (2018).
[75] H. Shan, X. Liu, H. Hildebrandt, C. Pan, N. Martinet, Z. Fan, P. Schneider, M. Asgari, J. Harnois-Déraps, H. Hoekstra *et al.*, *KiDS-450: Cosmological constraints from weak-lensing peak statistics – I. Inference from analytical prediction of high signal-to-noise ratio convergence peaks*, Mon. Not. R. Astron. Soc. **474**, 1116 (2018).
[76] J. M. Zorrilla Matilla, Z. Haiman, D. Hsu, A. Gupta, and A. Petri, *Do dark matter halos explain lensing peaks?*, Phys. Rev. D **94**, 083506 (2016).
[77] A. Gupta, Jose Manual Zorrilla Matilla, D. Hsu, and Z. Haiman, *Non-Gaussian information from weak lensing data via deep learning*, Phys. Rev. D **97**, 103515 (2018).
[78] R. Laureijs, J. Amiaux, S. Arduini, J.-L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, *Euclid definition study report*, arXiv:1110.3193.

[79] B. Derrida, J. Lebowitz, and E. Speer, *Large deviation of the density profile in the steady state of the open symmetric simple exclusion process*, J. Stat. Phys. **107**, 599 (2002).
[80] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, *Macroscopic fluctuation theory*, Rev. Mod. Phys. **87**, 593 (2015).
[81] H. W. Lin, M. Tegmark, and D. Rolnick, *Why does deep and cheap learning work so well?*, J. Stat. Phys. **168**, 1223 (2017).
[82] S. Iso, S. Shiba, and S. Yokoo, *Scale-invariant feature extraction of neural network and renormalization group flow*, Phys. Rev. E **97**, 053304 (2018).
[83] D. A. Roberts, S. Yaida, and B. Hanin, *The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks* (Cambridge University Press, Cambridge, England, 2022).
[84] F. Guth, S. Coste, V. De Bortoli, and S. Mallat, *Wavelet score-based generative modeling*, in *Proceedings of the NeurIPS* (2022), https://proceedings.neurips.cc/paper_files/paper/2022/hash/03474669b759f6d38cdca6fb4eb905f4-Abstract-Conference.html.
[85] K. Kadkhodaie, F. Guth, M. S., and S. Simoncelli, *Learning multi-scale local conditional probability models of images*, in *Proceedings of the ICLR* (2023), https://arxiv.org/abs/2303.02984.
[86] Y. Song, S. Garg, J. Shi, and S. Ermon, *Sliced score matching: A scalable approach to density and score estimation*, in *Uncertainty in Artificial Intelligence* (PMLR, 2020), pp. 574–584, https://proceedings.mlr.press/v115/song20a.html.
[87] F. Guth, E. Lempereur, J. Bruna, and S. Mallat, *Conditionally strongly log-concave generative models*, in *Proceedings of the ICML* (to be published).
[88] G. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, *PyWAVELETS: A PYTHON package for wavelet analysis*, J. Open Source Softwaare **4**, 1237 (2019).
[89] A. Milchev, D. Heermann, and K. Binder, *Finite-size scaling analysis of the Φ^4 field theory on the square lattice*, J. Stat. Phys. **44**, 749 (1986).
[90] M. Hasenbusch, *A Monte Carlo study of leading order scaling corrections of 4 theory on a three-dimensional lattice*, J. Phys. A **32**, 4851 (1999).
[91] A. Tröster, C. Dellago, and W. Schranz, *Free energies of the ϕ^4 model from Wang-Landau simulations*, Phys. Rev. B **72**, 094103 (2005).
[92] W. Zhong, G. T. Barkema, D. Panja, and R. C. Ball, *Critical dynamical exponent of the two-dimensional scalar ϕ^4 model with local moves*, Phys. Rev. E **98**, 062128 (2018).
[93] K. Binder, *Finite size effects at phase transitions*, in *Computational Methods in Field Theory*, edited by H. Gausterer and C. B. Lang (Springer, New York, 1992), pp. 59–125.
[94] http://columbialensing.org/ (Section "Dark Matter").