Relevance in the Renormalization Group and in Information Theory

Amit Gordon,[1] Aditya Banerjee,[1] Maciej Koch-Janusz,[2,3] and Zohar Ringel[1]
[1]*Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel*
[2]*Department of Physics, University of Zurich, 8057 Zurich, Switzerland*
[3]*James Franck Institute, The University of Chicago, Chicago, Illinois 60637, USA*

(Received 14 December 2020; revised 10 March 2021; accepted 28 April 2021; published 17 June 2021)

The analysis of complex physical systems hinges on the ability to extract the relevant degrees of freedom from among the many others. Though much hope is placed in machine learning, it also brings challenges, chief of which is interpretability. It is often unclear what relation, if any, the architecture- and training-dependent learned "relevant" features bear to standard objects of physical theory. Here we report on theoretical results which may help to systematically address this issue: we establish equivalence between the field-theoretic relevance of the renormalization group, and an information-theoretic notion of relevance we define using the information bottleneck (IB) formalism of compression theory. We show analytically that for statistical physical systems described by a field theory the relevant degrees of freedom found using IB compression indeed correspond to operators with the lowest scaling dimensions. We confirm our field theoretic predictions numerically. We study dependence of the IB solutions on the physical symmetries of the data. Our findings provide a dictionary connecting two distinct theoretical toolboxes, and an example of constructively incorporating physical interpretability in applications of deep learning in physics.

DOI: 10.1103/PhysRevLett.126.240601

The study of theoretical models is an essential part of physics. For sufficiently complex systems, however, establishing what the correct degrees of freedom are, and building a model in their terms, is a challenge in itself. The process is driven by experimental or numerical observations, but in practice physical intuition and prior knowledge are crucial to constructing a sufficiently simple model capturing the "essence" of the phenomenon, rather than an abundance of raw data [1]. Still, data itself should contain sufficient information for this task, and a tantalizing prospect is to perform it in an unbiased, automatic fashion using modern computational methods, particularly deep learning (DL) [2–4]. A fundamental obstacle to this is the mismatch between the concepts of physics, largely formulated in the language of field theory, and the theory and engineering practice of DL, all but ensuring questions of interpretability [5]. To bridge this divide a framework is required capable of expressing, and allowing for practical computation, of quantities on both sides. Information theory, deeply connected to physics and computer science [6–8], is a natural candidate.

In its classical formulation information theory was intentionally agnostic to the contents of the information, focusing on its efficient transmission [9]. Though often only part of the information is pertinent to the problem, defining a formal notion of "relevance" in sufficient generality has proven difficult [10]. This was addressed in the seminal information bottleneck (IB) paper Ref. [11]: relevant information in a random variable was defined by correlations, or sharing information, with an auxiliary relevance variable, providing an implicit filter indicating what to keep and what to discard. An example of such a relevance variable for the task of compressing a recorded speech is its written transcript. Compressing data to preserve the implicitly defined relevant part most efficiently was cast as a Lagrangian optimization problem, for which DL methods have recently been introduced [12].

In physics, however, there already exists a fundamental and *a priori* independent notion of relevance, based on the properties of the operators under scale transformations embodied in the celebrated renormalization group (RG) flow [13–15]. RG relevance is the most precise definition we possess of what it means for an observable to determine macroscopic physical properties of the system; it directly connects to the powerful formalism of conformal field theories (CFTs) [16–19], which revolutionized the understanding of critical phenomena [20–22].

Here we show that these two notions, belonging to entirely different theoretical frameworks, are in fact equivalent in physical systems, i.e., the information about long-range properties relevant in the information-theoretic sense is formally determined by the most relevant operators in the sense of the RG. Information loss in the context of the RG has been attracting interest since the observation of irreversibility of its flow [23–31]; we introduce a formal connection to compression theory which is constructive, quantitative, and computable. This allows us to verify our predictions numerically.

We prove that within the IB approach the most relevant operators can be extracted from the data, along with information about physical symmetries, based on intrinsic information-theoretic quantities characterizing the distribution, i.e., without invoking field-theoretic objects. This result is thus not only of theoretical, but also of practical importance. It provides a route towards automating theoretical tasks, e.g., deriving Ginzburg-Landau effective descriptions, and detecting symmetries hidden or emergent in a controlled and by construction interpretable way, by using the toolbox of statistics and deep learning on complex data.

To wit, while we focus on theoretical foundations, in a parallel work these results and recent DL advances [32,33] are leveraged to construct an efficient algorithm, the real-space mutual information neural estimator (RSMI-NE) [34,35], extracting the physically most relevant operators from much larger inputs, and characterizing spatial correlations, phase transitions, and order parameters. We show that RSMI is a limit of the IB problem, providing a theoretical underpinning for this promising numerical method.

Below we briefly review IB theory and its relation to the RSMI approach to the real-space RG in the context of statistical mechanical systems described by a CFT. We then present the main result: an analytical solution to the IB equations at strong compression which provides an explicit dictionary between IB relevancy, RG relevancy, and eigenvectors of the transfer matrix in any dimension. We compare these predictions with numerics, obtaining agreement to high precision. In addition we show how symmetries are manifested in the compressed and coarse-grained degrees of freedom. The Supplemental Material gives technical details and background information.

Relevant features of any data, physical or not, are only meaningfully defined relative to the task at hand, and their identification is complicated by multiple "irrelevant" (for the question asked) structures or regularities which may simultaneously exist in the data. The information bottleneck provides a rigorous framework for unsupervised learning of such most relevant features. With joint probability distribution of "data" V and an auxiliary "relevance" variable E as inputs, the IB finds the optimal (lossy) compression H of V preserving information about E (see Fig. 1). The correlations with E thus define what is relevant in V, rather than arbitrary measures. The IB can be posed as the following variational problem:

$$\min_{P(H|V)} \mathcal{L}_{\mathrm{IB}}[P(H|V)] \equiv \min_{P(H|V)} I(V;H) - \beta^I I(H;E), \quad (1)$$

where the optimization is over conditional probability distributions $P(H|V)$ describing the encoding of V into H. The mutual information terms I in $\mathcal{L}_{\mathrm{IB}}$ quantify total

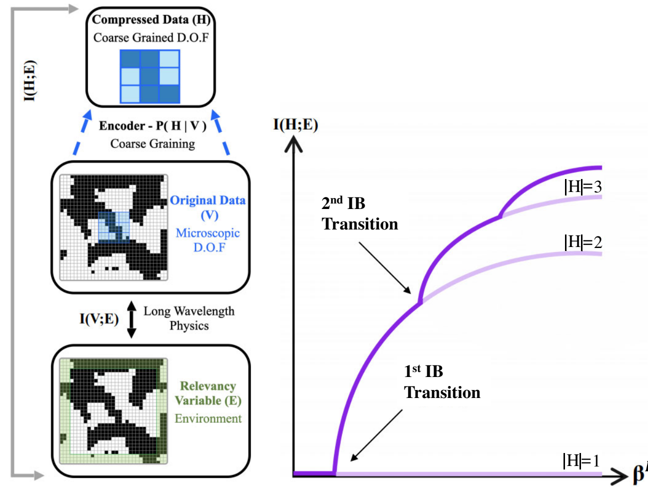

FIG. 1. Left: The general outline of the IB scheme, and in the physical setup of the RSMI RG [36,37]: an optimal encoder extracting information about relevance variable E contained in V is constructed. Right: IB curves depicting relevant information $I(H;E)$ retained by solutions to the IB equations (encoders), as a function of the tradeoff β^I [see Eq. (1)]. At critical values of β^I IB transitions occur: new solutions, with compressed variable H of increased cardinality (i.e., tracking additional features) appear, while the old ones become unstable minima of $\mathcal{L}_{\mathrm{IB}}$.

retained information (i.e., compression rate), and the relevant information thus preserved, respectively, with parameter $\beta^I \geq 0$ controlling the tradeoff between them.

The optimal encoder is found either by iteratively solving a set of coupled "IB equations" obtained from the $\delta\mathcal{L}_{\mathrm{IB}}/\delta P(H|V) = 0$ variation (see [38] for numerical algorithms), or more practically, applying ML variational inference techniques [12]. For the formal analysis here the IB equations are used. Strikingly, the optimal encoders undergo a sequence of sharp "IB" transitions as β^I is varied (see Fig. 1), which are bifurcations of the minima of $\mathcal{L}_{\mathrm{IB}}$. Particularly, the encoder is trivial (retaining zero information) until a finite value of $\beta^I_{c,1}$ at which the first IB transition occurs, when the gain due to retaining some (most) relevant aspect of data outweighs the penalty for keeping any information at all. At each subsequent transition the encoder begins to track another distinct feature of data. This discontinuous behavior, both for discrete [39,40] and continuous variables [41], is crucial, allowing us to identify such well-defined features.

While the IB may be applied to any data, it is of fundamental interest to confront the notion of relevance it gives rise to, and the features it extracts, with the physical relevance, as defined by the RG. The former being entirely determined by the relevance variable, we need to define E ensuring the IB retains precisely the RG-relevant information, and prove this is indeed the case. An appropriate definition for the real-space RG was postulated in the context of RSMI [36,37]: for a random variable V

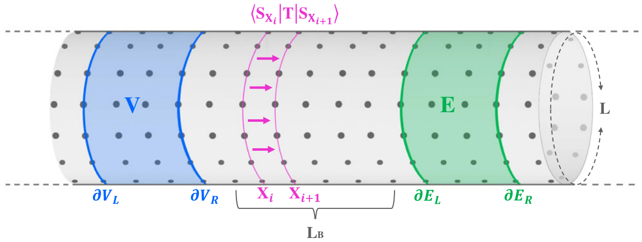

FIG. 2. The transfer matrix (TM) setup used. For a system on a cylinder the IB equations can be solved in terms of TM eigenvectors, which are related to the CFT data in the limit of large circumference L.

representing the marginal distribution of degrees of freedom in an area to be coarse-grained the variable E (the "environment"), is the remainder of the system beyond a shell of nonzero thickness around V (the "buffer," see Fig. 2). The thickness of the excluded buffer, formally taken to infinity, sets the length scale separating short-range correlations to be discarded, from information about long-range properties of the system. Despite conceptual appeal (the system itself defines relevance), and partial numerical [36] and theoretical evidence [37], the validity and the relation of this approach to field theory were unclear. There are also subtle differences between the IB and RSMI approaches. We now can resolve these issues.

To this end consider a statistical mechanical system on a cylinder; the subsystem to be coarse-grained V, the buffer of size L_B, and the relevance variable E are its subsections as per Fig. 2. We assume the system is governed by short-range interactions, and use the classic transfer-matrix (TM) method [42–44]: the partition function can be written as $\mathcal{Z} = \langle \mathrm{BC}|\mathcal{T}^{L_\infty}|\mathrm{BC}\rangle$, where is L_∞ the system length, and the entries of $\mathcal{T}$ are matrix elements of the exponentiated Hamiltonian between configurations of degrees of freedom on elementary slices of the cylinder (on a lattice; in continuum they are taken between slices of the states in the discretized path integral). We use bra-ket notation for such configurations, in particular $|\mathrm{BC}\rangle$ are boundary conditions at the cylinder ends. The unique advantage of the TM approach is that, on the one hand, all distributions entering the IB equations can be cast as matrix elements and partial traces of powers of $\mathcal{T}$, and on the other hand the eigenvalues λ_i and eigenvectors $|i\rangle$ of $\mathcal{T}$ have a direct relation to the operator content of the CFT describing the system [45–47]. Specifically, $\lambda_i/\lambda_0 = e^{-(2\pi/L)\Delta_i}$ in the limit of large cylinder circumference L, where Δ_i are the total scaling dimensions of the CFT primaries (which determine the RG scaling dimensions, and so the critical exponents) in ascending order. The TM thus serves as a theoretical dictionary helping to establish a quantitative map between the field-theoretic and information-theoretic objects.

Consider then the IB equations for the optimal encoder $P(h|v)$ at a fixed tradeoff β^I ([48], [11]):

$$P(h|v) \propto P(h)e^{\beta^I \sum_e P(e|v)\log[P(e|h)]}$$
$$P(e|h) = \sum_v p(e|v)p(v|h), \tag{2}$$

where e, h, v are configurations of E, H, and V. Equations (2) are highly nonlinear and coupled. Remarkably though, after rewriting them in the basis of eigenvectors of $\mathcal{T}$, the only interdependence of the conditional probabilities in the limit of large buffer L_B can be shown to be via

$$r_v = \frac{\langle 0|\phi_{\Delta_1}|\partial V_R\rangle}{\langle 0|\partial V_R\rangle} \qquad r_e = \frac{\langle \partial E_L|\phi_{\Delta_1}|0\rangle}{\langle \partial E_L|0\rangle}, \tag{3}$$

i.e., the matrix elements of the CFT primary fields ϕ_{Δ_i} of lowest scaling dimensions Δ_i, where $\phi_{\Delta_1}|0\rangle$ is the subleading TM eigenvector. In particular, we prove

$$P(h|v) = P(h|r_v) \propto P(h)e^{\beta^I \epsilon^2 r_v \langle r_v\rangle_h}, \tag{4}$$

with $\epsilon = (\lambda_1/\lambda_0)^{L_B}$ given by the lowest TM eigenvalues, and $\langle r_v\rangle_h$ the expectation value of r_v given h.

Equation (4) is one of our key results: the optimal IB encoder depends on V only via r_v, i.e., the matrix element of the most relevant operator in the sense of the RG (in $\langle r_v\rangle_h$ v dependence is averaged over). The solution changes from one system (CFT) to another through the values of r_v and $\langle r_v\rangle_h$. This is the mathematical statement of the equivalence in the title, between the RG notion of relevance in terms of operators and scaling dimensions, and one expressed entirely information-theoretically by IB supplemented with the geometric RSMI definition of the environment variable. Consequently, the "features" r_v that the IB, and thus the RSMI, extract are not arbitrary, but correspond to physically most relevant operators, yielding, e.g., magnetization for the Ising model (see below), the vertex operator for the free boson, or electric charge operators in the interacting dimer model [34].

Though Eq. (4) is implicit, as $\langle r_v\rangle_h$ depends on $P(h|v)$, it can be analytically solved around the first IB transition, i.e., for $\beta^I = \beta^I_{c,1} + t$. Below $\beta^I_{c,1}$ no information is retained: the encoder is independent of V and trivial: $P(h|r_v) = 1/|H|$, with $|H|$ the cardinality of the coarse-grained variable. Equiprobability of h reflects a structural symmetry of the encoder under permutations of h labels. Any nontrivial encoder must break it, introducing dependence of some h on V to preserve information: $\beta^I_{c,1}$ marks the first such breaking (in fact all IB transitions reflect successive breaking of permutation symmetry). Above $\beta^I_{c,1}$, following [39,40], the encoder can be perturbatively expanded around the trivial solution (see the

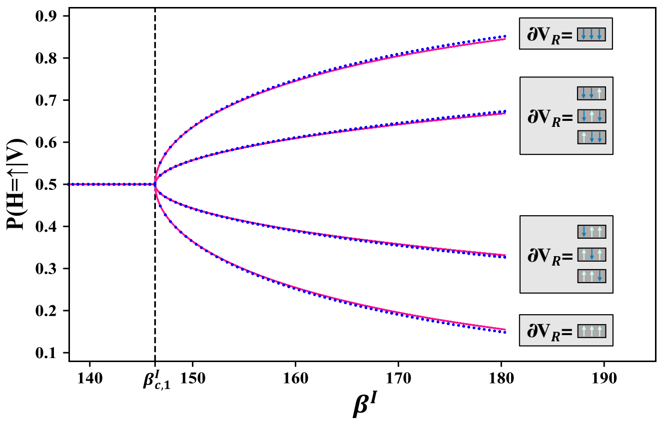

FIG. 3. Comparison of theory with numerics. The analytical prediction (solid red) for the optimal IB compression $P_\beta(h|v)$ (see Eq. (4), and Eq. C6 in the Supplemental Material) in terms of CFT and transfer matrix data is confronted with encoders obtained by numerically solving IB Eqs. (2) on the probability distribution of the system (blue dots), here the critical 2D Ising model. For clarity we use a cylinder of three sites' circumference, V and E as in Fig. 2. The variable H is a spin, whose probability to take value ↑ we plot as a function of the tradeoff β^I [see Eq. (1)]. The encoder is entirely random and independent of V below β^I_c matching the prediction Eq. (5), and above is determined by the physically relevant magnetization on the edge, in excellent agreement with the theory.

Supplemental Material). In particular, comparing to the expansion of Eq. (4) in t yields

$$(\beta^I_{c,1})^{-1} = \epsilon^2 + o(\epsilon^2) \overset{L\to\infty}{\to} e^{-4\pi\Delta_1(L_B/L)} + o(\epsilon^2). \quad (5)$$

Here $o(\epsilon^2)$ reflects the contribution of operators with subleading relevance, containing powers of ϵ greater than two. Since ϵ decays exponentially in L_B/L maintaining $L_B \gg L$ keeps those corrections exponentially small.

Equation (5) is an analytical prediction for the first IB transition, signaling emergence of nontrivial solutions to the IB equations (see Fig. 1 above and Fig. 1 in the Supplemental Material), in terms of field-theoretic data characterizing the physical system underlying the probability distribution. In the Supplemental Material, we derive this nontrivial solution explicitly (see also Fig. 3).

The prediction is generic and verifiable: we can input the probability distribution of the physical system to the IB equations, and find the solutions for changing β^I numerically, as in a compression problem [38]. On the other hand we can use the CFT description and either compute r_v, $\langle r_v \rangle_h$, and ϵ analytically, or by a numerical TM diagonalization, and compare. In Fig. 3(c) numerical IB solutions are plotted as a function of β^I in the case of the critical 2D Ising model. The value β^I at which nontrivial encoders appear matches the predicted $\beta^I_{c,1}$ to five digits' accuracy. The feature the IB extracts is the most relevant local operator, i.e., the magnetization in this case (see the Supplemental Material). Note that thus far analytical solutions to the IB problem were limited to the Gaussian variable case [41].

The validity of this picture is not limited to lattice models. In fact, for the continuum Gaussian field theory the entire IB curve can be computed analytically, including all the IB transitions [59], using Gaussian information bottleneck results [41] and Green's functions.

As mentioned, the RSMI algorithm [36,37] is closely related to the IB. Specifically, it also maximizes the relevant information $I(H;E)$, however contains no trade-off β^I, but instead a fixed cardinality $|H|$. Intuitively, the IB extracts as many features as β^I allows, adding them as β^I grows, while the RSMI from the outset optimizes exactly $|H|$ best features. RSMI is thus a $\beta^I \to \infty$ limit of the IB under the constraint of fixed $|H|$. In practice $|H|$ is also bounded in IB, but this affects solutions only at β^I large enough for $|H|$ features to have already been used.

The quantitative connection between compression- and field-theoretic formalisms thus established opens the exciting possibility of applying distinct theoretical and numerical methods of either area to its counterpart. We discuss such avenues in the conclusions, here, however, we immediately demonstrate one interesting example.

Symmetries are crucial in analytical understanding of physical systems, and in the RG in particular [60]. They have a direct relation to order parameters, and often effectively determine the long range properties. One thus expects the IB and RSMI to reflect the relevant symmetries of the model. Let s be an element of such symmetry group $\mathcal{S}$ acting on configurations of V and E as a permutation, denoted by multiplication, leaving the system invariant: $P(e,v) = P(se,sv)$. We expect the optimal encoder $P_\beta(h|v)$ to maintain it:

$$P(e,v) = P(se,sv) \Rightarrow P_\beta(h|v) = P_\beta(\phi_s h|sv), \quad (6)$$

so that the coarse-grained system is invariant under a representation ϕ_s of $\mathcal{S}$, potentially trivial. We show this indeed holds true in the IB, as long as $|H|$ is large enough to support a representation of an appropriate dimension. The argument is constructive: below $\beta^I_{c,1}$ the encoder is trivially invariant under all symmetries. For $\beta^I = \beta^I_{c,1} + t$ a solution can be built by an explicit symmetrization procedure, using the knowledge of the perturbative structure of the encoder around the first IB transition [40]. In the Supplemental Material we show this solution to be optimal. The symmetry of the encoder holds for all $\beta^I < \beta^I_{c,2}$ by continuity; numerical experiments support validity of this picture more generally.

Note that the symmetry $\mathcal{S}$ may not be obvious in the microscopic formulation of the system [61] or the experimental data, or may even be emergent [62]. Equation (6) can then be used as a constructive tool, potentially allowing us to systematically learn $\mathcal{S}$ from the symmetries of the entries of the numerically obtained $P_\beta(h|v)$ (see the Supplemental Material and also [63]). Moreover, the structure of the IB in the presence of physical symmetries or symmetries of the data shines light on the question of constructing RG transformations compatible with the symmetries of the system.

The presented results have clear theoretical and practical consequences: physical relevance in the sense of the RG can be defined, and thus probed, entirely in terms of information-theoretic quantities, without any explicit reference to "operators," "scaling dimensions," "field theories," etc. These mathematical objects are generically not available to us on the information theory side—we may have access to samples of probability distributions but without knowing which physical system generated them, what its relevant operators are, how they are expressed in terms of microscopic degrees of freedom, or which correlation functions we should be computing. By formalizing and proving the equivalence of the information-theoretic notion of physical relevance we gave it a concrete meaning, and developed the necessary technology to make it quantitative and computable analytically and numerically. This offers a path to discover the answer to the above questions, when faced with complex data.

Consequently, numerous directions are now open. On a theoretical front, application of IB analysis to extract relevant quantities in the challenging case of disordered and nonequilibrium systems is extremely promising, given its nonreliance on the notion of a Hamiltonian. This may require deeper understanding of the properties of the IB equations, and their constrained version in the RSMI-NE algorithm. Numerically, given the relation to the transfer matrix, the possibility of using the IB or RSMI where TM computations are difficult (e.g., in 3D) is an exciting prospect, as is applying approximate numerical IB or RSMI to experimental data. Finally, we hope that the methodology of using information-theoretical formulations of physical quantities combined with the ability of deep learning to optimize them in a controlled fashion [34,35,64], can provide a blueprint for more theoretically interpretable applications of deep learning in physics.

Z. R. and A. G. acknowledge support from ISF Grant No. 2250/19 as well as helpful discussions with Etam Benger (HUJI). M. K. J. gratefully acknowledges the support of Sebastian Huber, during stay in whose group at ETH Zurich part of the work was performed, the financial support from the Swiss National Science Foundation and the NCCR QSIT, and from the European Research Council under the Grant Agreement No. 771503 (TopMechMat), as well as from European Unions Horizon 2020 programme under Marie Sklodowska-Curie Grant Agreement No. 896004 (COMPLEX ML). A. B. acknowledges financial support from Eldad Bettelheim's ISF Grant No. 1466/15 and Alex Retzker's MicroQC Grant No. 820314.

[1] P. V. Coveney, E. R. Dougherty, and R. R. Highfield, Big data need big theory too, Phil. Trans. R. Soc. A **374**, 20160153 (2016).

[2] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[3] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress, Rep. Prog. Phys. **81**, 074001 (2018).

[4] S. D. Sarma, D.-L. Deng, and L.-M. Duan, Machine learning meets quantum physics, Phys. Today **72**, No. 3, 48 (2019).

[5] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, Definitions, methods, and applications in interpretable machine learning, Proc. Natl. Acad. Sci. U.S.A. **116**, 22071 (2019).

[6] J. Preskill, Quantum information and physics: Some future directions, J. Mod. Opt. **47**, 127 (2000).

[7] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, England, 2002).

[8] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Inc., New York, USA, 2009).

[9] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[10] Y. Blau and T. Michaeli, Rethinking Lossy compression: The rate-distortion-perception tradeoff, in *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, Proceedings of Machine Learning Research (PMLR, 2019), pp. 675–685, Vol. 97.

[11] N. Tishby, F. C. Pereira, and W. Bialek, The information bottleneck method, *Proceedings of the 37th Allerton Conference on Communication, Control and Computation* (Univ. of Illinois, Illinois, 2001), Vol. 49.

[12] A. A. Alemi, I. Fischer, J. V. Dillon, and K. Murphy, Deep variational information bottleneck, arXiv:1612.00410.

[13] K. G. Wilson and J. Kogut, The renormalization group and the ϵ expansion, Phys. Rep. **12**, 75 (1974).

[14] K. G. Wilson, The renormalization group: Critical phenomena and the Kondo problem, Rev. Mod. Phys. **47**, 773 (1975).

[15] M. E. Fisher, Renormalization group theory: Its basis and formulation in statistical physics, Rev. Mod. Phys. **70**, 653 (1998).

[16] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, Infinite conformal symmetry of critical fluctuations in two dimensions, J. Stat. Phys. **34**, 763 (1984).

[17] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, Infinite conformal symmetry in two-dimensional quantum field theory, Nucl. Phys. **B241**, 333 (1984).

[18] D. Friedan, Z. Qiu, and S. Shenker, Conformal Invariance, Unitarity, and Critical Exponents in Two Dimensions, Phys. Rev. Lett. **52**, 1575 (1984).

[19] P. Di Francesco, P. Mathieu, and D. Snchal, *Conformal Field Theory*, Graduate Texts in Contemporary Physics (Springer, New York, NY, 1997).

[20] J. L Cardy, *Scaling and Renormalization in Statistical Physics*, Cambridge Lectrue Notes in Physics (Cambridge University Press, Cambridge, 1996).

[21] C. Itzykson, H. Saleur, and J.-B. Zuber, *Conformal Invariance and Applications to Statistical Mechanics* (World Scientific, Singapore, 1998).

[22] D. Poland, S. Rychkov, and A. Vichi, The conformal bootstrap: Theory, numerical techniques, and applications, Rev. Mod. Phys. **91**, 015002 (2019).

[23] A. B. Zamolodchikov, Irreversibility of the flux of the renormalization group in a 2D field theory, JETP Lett. **43**, 730 (1986).

[24] J. Gaite and D. O'Connor, Field theory entropy, the h theorem, and the renormalization group, Phys. Rev. D **54**, 5163 (1996).

[25] H. Casini and M. Huerta, A c-theorem for entanglement entropy, J. Phys. A **40**, 7031 (2007).

[26] S. M. Apenko, Information theory and renormalization group flows, Physica (Amsterdam) **391A**, 62 (2012).

[27] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, Parameter space compression underlies emergent theories and predictive models, Science **342**, 604 (2013).

[28] V. Balasubramanian, J. J. Heckman, and A. Maloney, Relative entropy and proximity of quantum field theories, J. High Energy Phys. 05 (2015) 104.

[29] C. Bény and T. J. Osborne, The renormalization group via statistical inference, New J. Phys. **17**, 083005 (2015).

[30] C. Bény and T. J. Osborne, Information-geometric approach to the renormalization group, Phys. Rev. A **92**, 022330 (2015).

[31] C. Bény, Coarse-grained distinguishability of field interactions, Quantum **2**, 67 (2018).

[32] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, Mutual Information Neural Estimation (PMLR, 2018), pp. 531–540, arXiv:1801.04062.

[33] B. Poole, S. Ozair, A. Van Den Oord, A. Alemi, and G. Tucker, On Variational Bounds of Mutual Information (PMLR, 2019), pp. 5171–5180, arXiv:1905.06922.

[34] D. Efe Gökmen, Z. Ringel, S. D. Huber, and M. Koch-Janusz, Statistical physics through the lens of real-space mutual information, arXiv:2101.11633.

[35] D. E. Gökmen, Z. Ringel, S. D. Huber, and M. Koch-Janusz, Phase diagrams with real-space mutual information neural estimation, arXiv:2103.16887.

[36] M. Koch-Janusz and Z. Ringel, Mutual information, neural networks and the renormalization group, Nat. Phys. **14**, 578 (2018).

[37] P. M. Lenggenhager, D. E. Gökmen, Z. Ringel, S. D. Huber, and M. Koch-Janusz, Optimal renormalization group transformation from information theory, Phys. Rev. X **10**, 011037 (2020).

[38] S. Hassanpour, D. Wuebben, and A. Dekorsy, Overview and investigation of algorithms for the information Bottleneck method, in *Proceedings of the 11th International ITG Conference on Systems, Communications and Coding, SCC 2017* (VDE (Verlag GmbH), Berlin, 2017), pp. 1–6.

[39] A. E. Parker, T. Gedeon, and A. G. Dimitrov, Annealing and the rate distortion problem, in *Proceedings of the 15th International Conference on Neural Information Processing Systems*, NIPS02 (MIT Press, Cambridge, MA, USA, 2002), pp. 993–976.

[40] T. Gedeon, A. E. Parker, and A. .G Dimitrov, The mathematical structure of information Bottleneck methods, Entropy **14**, 456 (2012).

[41] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, Information Bottleneck for Gaussian variables, in *Advances in Neural Information Processing Systems*, edited by S. Thrun, L. K. Saul, and B. Schölkopf (MIT Press, Cambridge, 2004), Vol. 16, pp. 1213–1220.

[42] H. A. Kramers and G. H. Wannier, Statistics of the two-dimensional ferromagnet. Part I, Phys. Rev. **60**, 252 (1941).

[43] L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition, Phys. Rev. **65**, 117 (1944).

[44] P. Nightingale, Finite-size scaling and phenomenological renormalization, J. Appl. Phys. **53**, 7927 (1982).

[45] B. Derrida and L. De Seze, Application of the phenomenological renormalization to percolation and lattice animals in dimension 2, J. Phys. **43**, 475 (1982).

[46] J. L. Cardy, Conformal invariance and universality in finite-size scaling, J. Phys. A **17**, L385 (1984).

[47] J. L. Cardy, Operator content of two-dimensional conformally invariant theories, Nucl. Phys. **B270**, 186 (1986).

[48] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.126.240601 for further discussion and details, which includes [49–58].

[49] E. Schneidman, N. Slonim, N. Tishby, R. R. de Ruyter van Steveninck, and W. Bialek, Analyzing neural codes using the information Bottleneck method (2001), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.3263.

[50] F. Creutzig and H. Sprekeler, Predictive coding and the slowness principle: An information-theoretic approach, Neural Comput. **20**, 1026 (2008).

[51] L. Buesing and W. Maass, A spiking neuron as information Bottleneck, Neural Comput. **22**, 1961 (2010).

[52] N. Slonim and N. Tishby, Document clustering using word clusters via the information Bottleneck method, in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00 (Association for Computing Machinery, New York, 2000), pp. 208–215.

[53] S. Still, W. Bialek, and L. Bottou, Geometric clustering using the information Bottleneck method, in *Advances in Neural Information Processing Systems*, edited by S. Thrun, L. Saul, and B. Schölkopf (MIT Press, Cambridge, 2004), Vol. 16, pp. 1165–1172.

[54] D. J. Strouse and D. J. Schwab, The information Bottleneck and geometric clustering, Neural Comput. **31**, 596 (2019).

[55] F. Creutzig, A. Globerson, and N. Tishby, Past-future information bottleneck in dynamical systems, Phys. Rev. E **79**, 041925 (2009).

[56] S. Still, Information Bottleneck approach to predictive inference, Entropy **16**, 968 (2014).

[57] S. Agmon, E. Benger, O. Ordentlich, and N. Tishby, Critical slowing down near topological transitions in rate-distortion problems, arXiv:2103.02646.

[58] N. Slonim, The information Bottleneck: Theory and applications, Ph.D. thesis, The Hebrew University of Jerusalem, 2002.
[59] A. Banerjee and Z. Ringel, Information bottleneck and Gaussian field theory (to be published).
[60] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, International Series of Monographs on Physics (Clarendon Press, Oxford, 1989).
[61] C. N. Yang and S. C. Zhang, SO_4 symmetry in a Hubbard model, Mod. Phys. Lett. B **04**, 759 (1990).
[62] T. Senthil, A. Vishwanath, L. Balents, S. Sachdev, and M. P. A. Fisher, Deconfined quantum critical points, Science **303**, 1490 (2004).
[63] R. Bondesan and A. Lamacraft, Learning symmetries of classical integrable systems, arXiv:1906.04645.
[64] A. Nir, E. Sela, R. Beck, and Y. Bar-Sinai, Machine-learning iterative calculation of entropy for physical systems, Proc. Natl. Acad. Sci. U.S.A. **117**, 30234 (2020).