Coarse graining systems on inhomogeneous graphs using contrastive learning

Doruk Efe Gökmen
Institute for Theoretical Physics
ETH Zurich
8093 Zurich, Switzerland
dgoekmen@ethz.ch

Sounak Biswas
Institut für Theoretische Physik und Astrophysik
Universität Würzburg
97074 Würzburg, Germany

Sebastian D. Huber
Institute for Theoretical Physics
ETH Zurich
8093 Zurich, Switzerland

Zohar Ringel
Racah Institute of Physics
The Hebrew University of Jerusalem
Jerusalem, 9190401, Israel

Felix Flicker
School of Physics and Astronomy
Cardiff University
Cardiff, CF24 3AA, United Kingdom

Maciej Koch-Janusz
Haiqu Inc. 95 Third Street,
San Francisco, California 94103, USA
maciej@haiqu.ai

Abstract

Understanding and characterizing the emergent behavior of systems with numerous interacting components is typically difficult. This is especially the case when these interactions occur on an inhomogeneous graph, a situation relevant to many systems in bio- and statistical physics. Here we showcase a data driven approach, aimed at optimally compressing the system's information based on an information-theoretic principle. We develop an efficient numerical algorithm applicable to systems on arbitrary static graphs which employs variational estimators of mutual information to find optimal compression. We demonstrate that the optimal compression maps interpretably extract physically relevant local degrees of freedom. This enables us to construct an effective theory of a strongly correlated system on a quasicrystal.

1 Introduction

Data from experiments and simulations are critical to the study of complex systems. A glut of raw data does not, however, equate understanding, particularly when its processing easily exceeds our computational resources. The aim of physicists is to distill data into a concise theory using appropriate collective variables which capture the essence of the system. Renormalization group (RG) approaches provide a systematic path towards that goal [32, 16], but identifying the relevant degrees of freedom (DOFs) and deriving their effective theory [23, 13] is challenging in inhomogeneous systems, which are ubiquitous in biology [5, 1, 4] and in physics of disordered materials[3, 27].

Here we address this by framing the RG for inhomogeneous systems as a lossy compression of information on a graph [29, 10], enabling a geometry-independent approach. We use contrastive learning [2, 26, 31] to execute our approach in high-dimensional data inherent in such problems. Our algorithm explicitly constructs new effective DOFs $\mathcal{H}^i$ summarizing the configurations of the original DOFs on local subgraphs $\mathcal{V}^i$. This is achieved by a coarse graining transformation $\mathbf{\Lambda}^i : \mathcal{V}^i \mapsto \mathcal{H}^i$

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

locally chosen in region i to maximise the real-space mutual information (RSMI) [7, 20, 18]

$$\mathbf{\Lambda}^i = \arg\max I(\mathcal{H}^i : \mathcal{E}^i), \tag{1}$$

where the environment $\mathcal{E}^i$ is defined using the graph distance. This objective is the basis of the real-space mutual information neural estimation (RSMI-NE) algorithm [7]. The compression constraint can be enforced by limiting the information capacity of $\mathcal{H}^i$ using a predetermined number of encoding bits, thereby providing an approximation to the information bottleneck problem (IB) [10].

The variational principle in Eq. 1 provides a powerful substitute for intuition: effective DOFs are locally determined by the statistics of their individual environments. This is essential for moving beyond translation invariant systems. While coarse graining using $\mathbf{\Lambda}^i$ erases microscopic fluctuations, its local optimisation can extract distinct large scale properties, as reflected in differing cardinalities of compressed variables $\mathcal{H}^i$ that emerge in different spatial regions across an inhomogeneous system. The main contribution of this paper is to extend RSMI-NE to extract the relevant collective variables on arbitrary static graphs using this insight.

Identification of coarse-grained lattice is another challenge in RG on disordered systems [13]. We show that the correct super-lattice can be revealed by analysing correlations of the compressed DOFs. This provides an operational definition of data-driven coarse graining in inhomogeneous systems.

1.1 Related work

Recently machine learning has inspired a number of new RG approaches [21, 12]. The RSMI-NE algorithm [7, 8] maximises a variational lower-bound of mutual information using contrastive learning [2, 26, 31] to implement Eq. 1. Such estimators have also been used to calculate entropies in physical systems [25], and quantum information theoretic extensions have recently been proposed [28, 9, 19]. RSMI-NE algorithm has similarities with previous approaches representation learning using mutual information maximisation [11] with the difference that the former bounds the entropy of representations from above. This compression sets an IB problem [30], whose connections to RG has been recently exposed [10, 17], enabling the physical interpretability of representations as relevant operators. Here we extend the RSMI-NE algorithm to arbitrary static graphs.

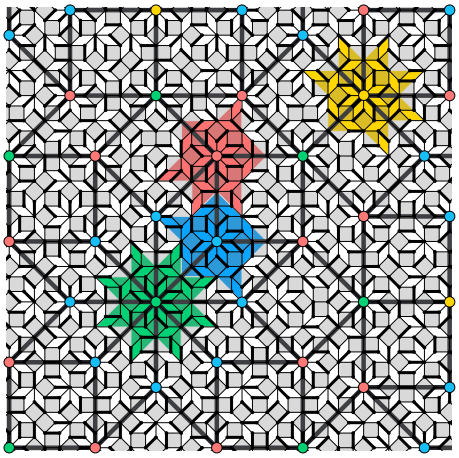

Figure 1: A microscopic dimer configuration (black links) on the AB tiling's edges, with an overlaid AB *super-lattice*, self-similar to the microscopic one. The effective DOF at a super-vertex of a given (colour coded) valence will be obtained by coarse graining the dimer configuration in the surrounding region $\mathcal{V}$ of a shape dictated by the inflation rules and shown as a polygon of a matching colour.

2 Real-space mutual information based coarse graining on a graph

Consider a system of microscopic degrees of freedom $\mathcal{X}$, distributed according to a joint probability distribution $P(\mathcal{X})$. We define a coarse graining (CG) map from $\mathcal{X} = \bigcup_i \mathcal{V}^i$ into a new set of compressed variables $\mathcal{X}' = \bigcup_i \mathcal{H}^i$ as a conditional probability $P(\mathcal{X}'|\mathcal{X}) = \prod_i P_{\mathbf{\Lambda}^i}(\mathcal{H}^i|\mathcal{V}^i)$. Each local factor defines a distinct compression map $\mathbf{\Lambda}^i : \mathcal{V}^i \mapsto \mathcal{H}^i$. In our work, $\mathcal{V}^i$ can be configurations on any local subgraph, *e.g.* given by topological balls on the graph, or another set dictated by the problem structure. Refs. [20, 10] showed that extracting long-range physics (in the form of RG-relevant operators) in translation-invariant systems can be cast as finding $\mathbf{\Lambda}$ maximizing a compression theoretic objective

$$I_{\mathbf{\Lambda}}(\mathcal{H} : \mathcal{E}) = \mathbb{E}_{(\mathcal{H},\mathcal{E})}\left[\log P(\mathcal{H},\mathcal{E}) - \log P(\mathcal{H})P(\mathcal{E})\right], \tag{2}$$

where the environment $\mathcal{E}$ of $\mathcal{V}$ is spatially separated. We note that objective 2 is geometry-independent, and can serve to define individual maps $\mathbf{\Lambda}^i$ for $\mathcal{V}^i$ on a graph. Furthermore, also the coarse-grained graph topology can be *determined by the correlations* of the new variables $\mathcal{H}^i$ thus resolving the key problems in constructing a well-defined coarse graining of DOFs interacting on an arbitrary graph.

We parametrize the coarse graining $P_{\mathbf{\Lambda}^i}(\mathcal{H}^i|\mathcal{V}^i)$ with an inner-product ansatz $\mathcal{H}^i_k := \mathrm{sign}\left(\mathbf{\Lambda}^i_{kj}\mathcal{V}^i_j\right)$, with linear NNs $\mathbf{\Lambda}^i = (\Lambda^i_k)$, though general NNs can also be used. The index k labels the components of a coarse-grained variable, j refers to the spatial positions in patch $\mathcal{V}^i$, where the indexing is defined with respect to a fixed labelling of vertices in the graph. The sign function discretization provides a compressive constraint. To allow for gradient-based training, it is smoothened using the Gumbel-softmax reparametrization [24, 15], which is annealed during training.

The main idea in contrastive learning is to train a critic function that maximises the embedding distance between similar and dissimilar samples. To estimate the mutual information using contrastive learning, the outputs $\mathcal{H}^i$ of the networks $\mathbf{\Lambda}^i$ are combined with samples of DOFs in $\mathcal{E}^i$ and fed to a critic function $f(\mathcal{H}^i, \mathcal{E}^i)$, which is parametrised by a deep neural network. Intiutively, the critic function learns to maximise the embedding distance between correlated pairs drawn from $p(\mathcal{H}^i, \mathcal{E}^i)$ and uncorrelated sample pairs drawn from $p(\mathcal{H}^i)p(\mathcal{E}^i)$. The parameters of the deep critic network f and the linear coarse graining network $\mathbf{\Lambda}^i$ are trained simultaneusly using stochastic gradient descent to maximise a tight lower-bound of mutual information as in [21, 12].

3 Results: solving a statistical physics conjecture

To showcase our algorithm and the interpretability of its outputs we apply it to a dimer model on a quasiperiodic Ammann-Beenker (AB) graph, conjectured to exhibit discrete scale invariance [6]. Dimer models are defined by configurations of binary variables on the edges obeying local constraints: at each vertex, exactly one incident edge is occupied. AB graphs have a scale-invariant self-similar structure under a scale transformation σ – the microscopic lattice and the self-similar *super*lattice (with bold links) are shown in Fig. 1. Note the difference between trivial scale-invariance of the graph, and that of the *physical system* on the graph. We generate dimer configurations, which are the inputs, using Monte Carlo sampling [6].

To coarse grain we specify the spatial partition into $\mathcal{V}^i$. In the AB tiling the hierarchical structure gives a natural choice [14], spanned by four distinct classes of blocks [22], shown in Fig. 1, named after the coordination of the corresponding vertices in the super-lattice. The variational compression maps $\mathbf{\Lambda}^i$ assign to Monte Carlo dimer configurations in $\mathcal{V}^i$ a short binary code $\mathcal{H}^i$ (Fig. 2a), the bits being set by applying individual components Λ_k to $\mathcal{V}^i$ (itself a long bit-string of dimer occupations in the block). Code lengths are found by sequentially increasing the number of components in $\mathbf{\Lambda}$ and re-training the compression to maximize objective 2 (Fig. 2b,f)

Optimal linear maps on the space of dimer configurations on $\mathcal{V}^i$ are shown for classes 8 and 3 in Figs.2c and g. They reveal the emergent DOF to be $\mathbb{Z}_n$ *clock variable*, with n the connectivity in the superlattice. This is manifested in the code statistics: we first note that the 4-bit codes form a closed 8-cycle, with neighbours differing by a single bit-flip, and each code having exactly two 1-bit distant neighbours (Fig. 2e). The uniform frequencies and the cyclic structure of the code reveal a symmetry. Indeed, a class-8 patch $\mathcal{V}$ of the AB *lattice* is locally symmetric under $\pi/4$ rotations. We observe that under such rotations the components of the optimal $\mathbf{\Lambda}$ map (Fig. 2c) change as $\mathsf{C}_8 : \ (\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4) \to (\Lambda_4, -\Lambda_3, -\Lambda_1, -\Lambda_2)$, which is a representation of a generator of the cyclic group C_8. In other words, th compression map, and consequently the emergent DOF, carry a representation of what is *a priori* a (local) symmetry of the AB lattice.

Similar analysis can be performed for other classes of $\mathcal{V}$, which have a mirror symmetry. In particular, under its action for the class-3 patch in Fig. 2g we have $(\Lambda_1, \Lambda_2) \to (\Lambda_2, \Lambda_1)$, explaining equal frequency of the **01** and **10** codes. Hence, we conclude that, rather than becoming continuous, the emergent DOFs of the dimer system at larger scales remain discrete, and mimic the local symmetry of the underlying super-lattice. This holds at even scales, as we demonstrate for σ^2 and σ^4 scale transformations, providing the first explicit indication of invariance of dimer physics under *discrete* scale transformations, or discrete scale invariance (DSI).

Having found the emergent DOFs in each class of $\mathcal{V}^i$ individually, we turn to their correlations. To this end, we simultaneously coarse grain dimer configurations in multiple patches, which collectively form an AB superlattice as in Fig. 1a We use the optimised compression maps (Fig. 2c,g). Since we found that distribution of each state's frequencies reflects the underlying superlattice symmetry, these internal DOFs can be identified with spatial orientations along the superlattice edges. For example,

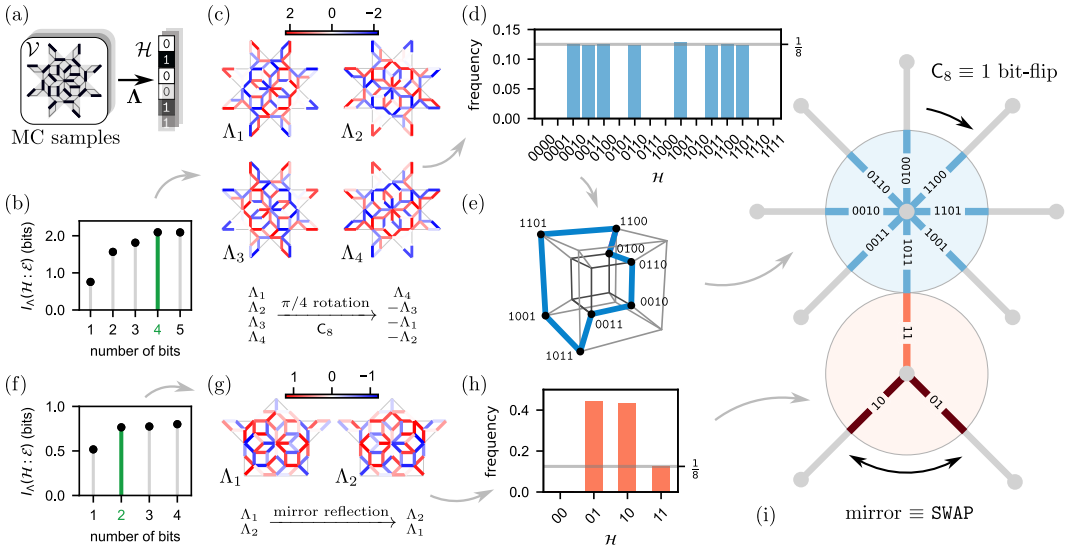

Figure 2: (a) Coarse graining $\mathbf{\Lambda}$ maps Monte Carlo configurations on $\mathcal{V}$ into bit-strings $\mathcal{H}$ on super-vertices of σ^{-2} deflated tiling. (b, f) Length of the bit-string $\mathcal{H}^{8(3)}$ is determined by saturation of mutual information at 4 (2) bits at 8 (3) super-vertices. The respective optimal filters $\mathbf{\Lambda}$ in (c, g) carry a representation of the local C_8 and mirror symmetries of corresponding super-vertices. (d, h) The code statistics indicate $\mathcal{H}^{8(3)}$) are clock variables with 8(3)-states. (e) In particular, states of $\mathcal{H}^8$ form a closed 8-loop, each state having exactly two neighbours with Hamming-distance 1. (i) The local symmetries induce transitions between adjacent clock-states, enabling to identify abstract clock-states with spatial directions along the links of the quasiperiodic lattice (see main text).

since mirror symmetry w.r.t. the axis connecting the 8- and 3-vertices in Fig. 2i relates the 3-vertex codes **01** and **10** (Fig. 2h), the remaining state **11** is the one pointing towards the 8-vertex.

We then probe the correlations by conditioning on the state of one of the vertices. In Fig. 3, superlattice fragments are shown with the state of the conditioning clock variable, identified with a direction, in orange, while the conditional distribution of DOFs at the other vertices in grey-scale. Remarkably, this distribution effectively forces occupation of some states, and excluding others. To wit, when the 3-vertex DOF points towards the 8-vertex, the distribution $P(\mathcal{H}|\mathcal{H}^3)$ of the latter is sharply peaked in the matching direction, while *no other neighbour* of the 3-vertex points towards it (allowing, for example, the identification of the 8-vertex code **1011** with a spatial orientation in Fig. 2i). Conversely, when the 3-vertex DOF points towards one of its other neighbours, it is "matched" by the latter, while the 8-vertex DOF distribution has zero weight *only* in the direction towards that 3-vertex.

Examining all correlations we find that the effective DOFs in $\mathcal{V}^i$'s across the lattice are paired with *one and only one* of their neighbours into emergent "super-dimers" on the edges of the superlattice. The exclusion of certain clock variable orientations in Figs. 2a-e is a precise reflection of the defining dimer-constraints, which these super-dimers obey. Moreover, comparison of further correlations to those of the microscopic dimers in Fig. 3a reveals that not just the local dimer constraints, but also longer-range correlations are reproduced. The physics of the microscopic dimer model on the AB lattice is thus replicated at larger scales, allowing to explicitly confirm the DSI conjecture.

4 Conclusions

We have demonstrated that relevant collective DOFs and their effective theory can be extracted from the structure of information inherent in raw high-dimensional data in a computationally efficient and interpretable manner. This approach excels in systems with irregular geometry and has formal connections to RG via the theory of lossy information compression. These characteristics are essential in *e.g.* complex biological models of tissues [5, 1, 4], and in disordered materials [3, 27], and we envisage the application of our tools in these domains.

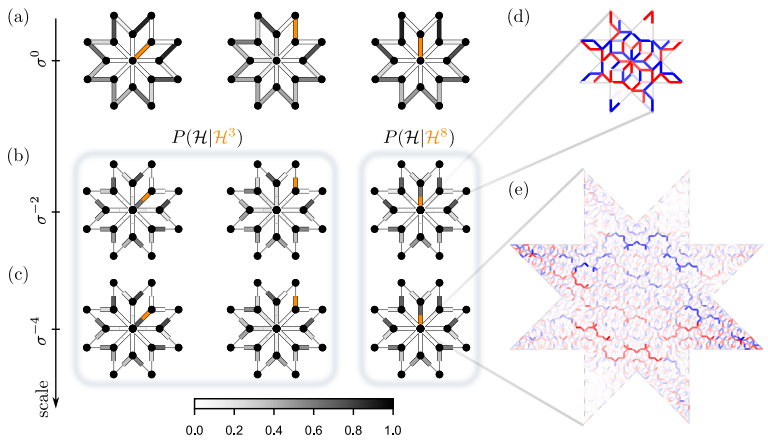

Figure 3: (a) Probability distribution of microscopic (*i.e.* σ^0) dimer occupations (grey-scale) on an AB graph patch, conditioned on one of the links (orange) hosting a dimer. (b, c) First two columns: the probabilities $P(\mathcal{H}|\mathcal{H}^3)$ of the constructed *emergent* clock variables on the σ^2 and σ^4 superlattice (in grey-scale), conditioned on two distinct states of one of the 3-clocks (in orange). The third column shows distributions conditioned on a state of the central 8-clock. Sharply peaked conditional distributions reveal binding of clocks into super-dimers at both σ^2 and σ^4 scales, and reproduction of microscopic dimer correlations at higher scales. (d, e) One component of the coarse graining maps for the 8-state clock variable at scales σ^2 and σ^4 (a structured function of 2760 local variables).

References

[1] Silvanus Alt, Poulami Ganguly, and Guillaume Salbreux. Vertex models: from cell mechanics to tissue morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1720):20150520, 2017.

[2] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. 80:531–540, 10–15 Jul 2018.

[3] J. D. Bernal. Geometry of the structure of monatomic liquids. *Nature*, 185(4706):68–70, jan 1960.

[4] Reza Farhadifar, Jens-Christian Röper, Benoit Aigouy, Suzanne Eaton, and Frank Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17(24):2095–2104, 2007.

[5] Alexander G. Fletcher, Miriam Osterfield, Ruth E. Baker, and Stanislav Y. Shvartsman. Vertex models of epithelial morphogenesis. *Biophysical Journal*, 106(11):2291–2304, 2014.

[6] Felix Flicker, Steven H. Simon, and S.A. Parameswaran. Classical Dimers on Penrose Tilings. *Phys. Rev. X*, 10(1), Jan 2020.

[7] Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Statistical physics through the lens of real-space mutual information. *Phys. Rev. Lett.*, 127:240603, Dec 2021.

[8] Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Symmetries and phase diagrams with real-space mutual information neural estimation. *Phys. Rev. E*, 104:064106, Dec 2021.

[9] Ziv Goldfeld, Dhrumil Patel, Sreejith Sreekumar, and Mark M. Wilde. Quantum neural estimation of entropies, 2023.

[10] Amit Gordon, Aditya Banerjee, Maciej Koch-Janusz, and Zohar Ringel. Relevance in the renormalization group and in information theory. *Phys. Rev. Lett.*, 126:240601, Jun 2021.

[11] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. In *International Conference on Learning Representations*, 2019.

[12] Hong-Ye Hu, Shuo-Hui Li, Lei Wang, and Yi-Zhuang You. Machine learning holographic mapping by neural network renormalization group. *Phys. Rev. Res.*, 2:023369, Jun 2020.

[13] Ferenc Iglói and Cécile Monthus. Strong disorder rg approach of random systems. *Physics Reports*, 412(5):277–431, 2005.

[14] A. Jagannathan. Quantum spins and quasiperiodicity: A real space renormalization group approach. *Phys. Rev. Lett.*, 92:047202, Jan 2004.

[15] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax, 2017.

[16] Leo P. Kadanoff. Scaling laws for ising models near T_c. *Phys. Phys. Fiz.*, 2:263–272, Jun 1966.

[17] Adam G Kline and Stephanie E Palmer. Gaussian information bottleneck and the non-perturbative renormalization group. *New Journal of Physics*, 24(3):033007, mar 2022.

[18] Maciej Koch-Janusz and Zohar Ringel. Mutual information, neural networks and the renormalization group. *Nature Physics*, 14:578–582, 2018.

[19] Sangyun Lee, Hyukjoon Kwon, and Jae Sung Lee. Estimating entanglement entropy via variational quantum circuits with classical neural networks, 2023.

[20] Patrick M. Lenggenhager, Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Optimal renormalization group transformation from information theory. *Phys. Rev. X*, 10:011037, Feb 2020.

[21] Shuo-Hui Li and Lei Wang. Neural network renormalization group. *Phys. Rev. Lett.*, 121:260601, Dec 2018.

[22] Jerome Lloyd, Sounak Biswas, Steven H. Simon, S.A. Parameswaran, and Felix Flicker. Statistical mechanics of dimers on quasiperiodic Ammann-Beenker tilings. *Phys. Rev. B*, 106(9), sep 2022.

[23] Shang-keng Ma, Chandan Dasgupta, and Chin-kun Hu. Random antiferromagnetic chain. *Phys. Rev. Lett.*, 43:1434–1437, Nov 1979.

[24] Chris J. Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables, 2017.

[25] Amit Nir, Eran Sela, Roy Beck, and Yohai Bar-Sinai. Machine-learning iterative calculation of entropy for physical systems. *Proceedings of the National Academy of Science*, 117(48):30234–30240, December 2020.

[26] Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5171–5180. PMLR, 09–15 Jun 2019.

[27] H. Sheng, W. Luo, F. Alamgir, J. Bai, and E. Ma. Atomic packing and short-to-medium-range order in metallic glasses. *Nature*, 439, 1 2006.

[28] Myeongjin Shin, Junseo Lee, and Kabgyun Jeong. Estimating Quantum Mutual Information Through a Quantum Neural Network. 6 2023.

[29] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. In *Proceedings of the 37th Allerton Conference on Communication, Control and Computation*, volume 49, 07 2001.

[30] Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method, 2000.

[31] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding, 2019.

[32] Kenneth G. Wilson. Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture. *Phys. Rev. B*, 4:3174–3183, Nov 1971.