THE UNIVERSITY OF CHICAGO

PRINCIPLES FOR COARSE-GRAINING IN BIOLOGICAL SYSTEMS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY

ADAM GORDON KLINE

CHICAGO, ILLINOIS
MARCH 2025

For my parents and grandparents

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I owe a great deal to a great number of people and despite my best attempts to thank them here I will fail. To begin, I would like to thank my mentors and friends at Purdue. In my first years of research, Adam Wasserman, Matthias Liepe, and R. Graham Cooks were great role models both as scientists and people. They were all invested in my growth and success well beyond the scope of our work together. I also would like to extend deepest thanks Tim Cook and Dan Hall for their friendship and guidance. The Society of Physics Students at Purdue became a new home for me, and I am so grateful for all of the close friends I made through it. All of these points of support and encouragement were instrumental in my decision to pursue a doctorate.

At UChicago, Heinrich Jaeger graciously supported me through my first year of work and taught me a lot about how to carry out and present research, especially impressing upon me the importance of clarity in scientific communication. The Jaeger lab was a wonderful place to begin my research, and I have to thank Kieran, Leah, and Melody in particular for being such great friends. I am grateful too for the warmth and companionship of those in my cohort and the physics department more generally. There are too many names to list. We tended to band up in large gatherings, collectively turning our shoulders to whichever problem sets we were procrastinating on. I'll never forget grilling out at the point, game nights, tennis and soccer matches, Valois, the prep for FTY, and the attempts we made to stay connected over Zoom and Discord once COVID arrived.

My committee has been wonderful. In the early stages of my work, when I began piecing together ideas about IB and RG, I was nervous about how this topic would be received. Arvind Murugan, Jason MacLean, Peter Littlewood, and Vincenzo Vitelli all encouraged me to continue chasing down these ideas and were instrumental in helping me to get over this anxiety. In particular, I would like to extend my thanks to Peter for supporting and advocating for my work many times over the years. It was an incredible privilege to have

this group of people rooting for me and supporting me.

I am deeply grateful to Alekandra Walczak and Thierry Mora for hosting me in Paris during the summers of 2022 and 2023. I extend my deepest thanks to Mirna, Paul, Xiaowen, Andrea, Antoine, María, Maria-Francesca, Natanael, Victor, Antonio, and everyone else for their friendship and efforts to make me feel so welcome.

I am deeply, incredibly grateful to my advisor Stephanie Palmer for her mentorship and companionship. It is obvious that the Ph.D. process requires one to develop intellectually, but what is perhaps less well-advertised is that it also requires a great deal of personal, emotional, and social growth. Stephanie helped me to see that the scientific endeavor is fundamentally human, and in training to become a scientist one needs to understand how to work with others, how to work with oneself, and how not everything is about work. Stephanie is an incredible scientist and thinker, and challenged me intellectually from the start. She also supported my growth as a person at every turn, and I was always certain that she had my best interest in mind. In any domain, very few are fortunate to be backed by someone who is at once so deeply compassionate and stunningly bright.

I owe many of my best experiences from the last six years to my friends. These people have kept me sane through no small effort, and essentially every good (or at least interesting) idea I have had can be traced in some way back to them. To Carlos, Mason, Kyle, Cheyne, Peter, Ben Lyo, Daine, Jeri, Alex, Umang, Jon, Jordan, Ted, Melody, Jason, and Tom, thank you for everything.

Finally, to my family, I don't know how to express sufficient gratitude. Absolutely none of this would have been possible without you. Thank you my brothers and cousins for the jam sessions, great conversations, and truly unjustifiable quantities of pizza shared over the years. To my parents, thank you for being so unconditionally supportive and caring, for always welcoming me home, and all of the work you put in to allow me to chase my dreams. To my grandparents, thank you all you have provided for me and the family throughout the years, and thank you for teaching me to seek with compassion.

ABSTRACT

Life presents us with bewildering phenomenology, from robust self-replication and morphogenesis, to learning, adaptation, and evolution. These phenomena are collective and typically carry no precedent in the inanimate world, which presents an exciting opportunity in the search for new emergent physics. Modern experimental methods offer increasingly high-resolution and large-scale data on the inner workings of biology, making conceivable the establishment of precise theories about the organization of living systems. As first steps towards this lofty and long-term goal, we need principles to define the most biologically relevant features in high-dimensional datasets, as well as tools to identify them and constrain theory. Motivated by the successes of many-body theory in physics, we examine how the renormalization group (RG) and information bottleneck (IB), two paradigmatic coarse-graining approaches, may contribute to the search for simplifying structure in large-scale biology. First, to explore how top-down and bottom-up coarse-graining might be reconciled, we demonstrate a formal theoretical connection between IB and RG. We find that the choice of relevance variable in IB determines the collective variables and their order of elimination in the RG scheme, suggesting IB can be used to select a notion of 'large scale' structure, such as a particular biological function. One especially exciting application of this idea may be to use it as a bridge, revealing how effective models in the language of RG can be read out of IB analyses on biological data. Next, prompted by multi-functionality in biology, we examine the idea that a given system might have multiple notions of large scale structure. We construct a simple model which exhibits this property, in that its correlations are best described in terms of two coexisting but independent RG flows, each coarse-graining with respect to a different collective basis. The failure of RG to predict large-scale correlations in a certain phase of this model points to a potential pitfall in interpreting maximum-entropy models of biological data as many-body systems. Finally, we turn our attention to real biological systems, and use IB to coarse-grain data taken from a large population of neurons in the salamander retina. By optimizing collective variables to predict future responses, we identify features in the retinal code that may enable downstream visual prediction. We also reveal that the encoding of predictive information in the retina becomes increasingly collective at long times, accompanied by a shift in the most relevant collective variables. This highlights the importance of adapting coarse-graining to function in biology.

CHAPTER 1

INTRODUCTION

In the last few decades, biophysics and quantitative biology have grown massively in scope (bio [2022]). High-throughput experimental and computational techniques are revealing the inner workings of brains (Hulse et al., Meshulam and Bialek [2024]), immune systems (Six et al. [2013]), and ecosystems (Liu and Good [2023]) in increasing detail, but assigning meaning to these data is a major theoretical challenge. When dealing with high-dimensional data, the issue of defining useful features and finding the models that govern their behavior is both conceptually and practically difficult. Similarly, it is difficult to extract predictions from highly detailed microscopic models. One hope is that not all of the details matter for describing all of the phenomena. Perhaps, with the appropriate coarse-graining methods, we may develop a view of biology which is more normative and less descriptive, bringing theory closer in line with the style of inquiry already employed with great success in physics (Bialek [2012]). On the other hand, biology is full of diversity, and we should be careful not to disavow descriptive approaches entirely. What, then, are the general principles that guide large, interacting systems towards their simplest descriptions, and how do they intersect with the questions of biology?

Part of the answer to this question may come from our understanding of effective models in physics. In our everyday experience of the world, the fact that atoms have internal structure is not immediately apparent. Most people get along just fine with an understanding in terms of the macroscopic. You learn to walk, handle objects, carry out conversations, and pay your taxes, and none of these things actually require an understanding of subatomic physics. But why not? Since the vast majority of matter that we interact with is comprised of atoms, shouldn't it be of great practical importance to know that it all comes down to neutrons, protons, and electrons bound up in myriad ways? A fascinating aspect of the natural world is that the laws governing reality at the smallest scales have limited impact on the phenomena at larger scales. New laws emerge from interactions between composite parts, and the possible ways of achieving this are effectively innumerable. Structure therefore arises not only from the laws obeyed by elementary constituents, but also the ways these constituents can be arranged. As Phil Anderson famously articulated, "More is Different" (Anderson [1972]).

Precisely how important are the small-scale rules in determining large-scale phenomena? Some connection must certainly be present, since we know that atomic and chemical makeup can affect the strength of a material, the viscosity of a fluid, or the electrical properties of a semiconductor. Benzene and water look very different microscopically, but when present in large enough quantities, they move and flow according to the Navier-Stokes equations, albeit with different parameters. Still, it may seem strange that the net effect of such microscopic differences is simply to shift parameters around in the same macroscopic theory.

One answer to this question is provided by the renormalization group (RG) approach. In the standard presentation of Wilsonian RG (Wilson [1971a,b], Wilson and Fisher [1972], Wilson [1974], Fisher [1995]), one describes the physics of a theory up to some smallest scale (typically length or time), called the cutoff. Then, by changing this cutoff scale one removes information about small-scale features in the system. The averaged activities of now-eliminated degrees of freedom cause shifts the effective parameters and interactions governing those which are still resolvable, and in this way, sweeping the cutoff from small to large scales yields dynamics in the space of theories. Fixed points in these dynamics correspond to macroscopic phases of matter, together with effective theories for how they behave. Critical points, for example, are theories at unstable fixed points in the RG flow. We also get a precise explanation for why some microscopic parameters affect macroscopic properties and why some don't. In the vicinity of a fixed point, attractive directions represent model parameters which are in a precise sense "irrelevant," since they die out at long RG times (large scales) while the repulsive directions represent "relevant" parameters, since their effects grow.

Probably the most lasting and revolutionary aspect of RG is not that it explains criticality and phase transitions, although this was a primary reason for Wilson's Nobel prize. Rather, it is the fact that RG justifies the use of effective theories in physics (Goldenfeld [2023]). A number of research directions have abstracted RG reasoning out of physical contexts to show that effective models can be justified analogously in areas such as machine learning and biology(Brown and Sethna [2003], Raju et al. [2018], Quinn et al. [2022], Berman et al. [2023], Berman and Klinger [2022], Bradde and Bialek [2017], Jona-Lasinio [1975]). These lines of reasoning are especially important for understanding how RG might contribute to coarse-graining in biological systems.

The most common way to describe detailed biological data in the language of statistical physics is by asserting that the states exhibited by the system are distributed according to a Boltzmann distribution (Schneidman et al. [2006], Tkačik et al. [2013, 2014], Mora et al. [2010]). To infer the corresponding Hamiltonian, one widely-used approach is known as the maximum-entropy method (Jaynes [1957]). Here, one introduces just enough parameters to reproduce a set of desired expectation values, while otherwise maximizing the model entropy. This always selects a Hamiltonian which is linear in the fitting parameters and the feature functions whose expectations are constrained. For example, by requiring a model to capture only the mean and pairwise statistics of binarized activity in a population of neurons, one arrives at a disordered Ising model. In the vertebrate retina, this pairwise model effectively captures higher-order correlations (Schneidman et al. [2006]), indicating that the corresponding higher-order interaction terms are not needed to explain the data. In some cases, this framing actually yields signatures of criticality (Mora and Bialek [2011], Mora et al. [2015], Tkačik et al. [2015]), though these claims remain controversial (Schwab et al. [2014], Morrell et al. [2021, 2024], Ngampruetikorn et al. [2023]). Regardless, since it is apt to think of certain biological systems in analogy with statistical physics, it may further be true that insights from the renormalization group can be applied. Further, whether any given criticality hypothesis turns out to be true, the statistical physics perspective has proved useful in quantifying collective phenomena in systems such as the brain and immune system.

A central difficulty in attempting to address biology using many-body theory is its complexity, which is present in multiple senses. First, given that living things and their parts evolve to fulfill multiple functions, there are many distinct collective phenomena one can model. For example, in *e coli*, one might care to know how an organism responds to heat shock, how often it is reproducing, or the dynamics of its run-and-tumble. Each of these systemic responses are meaningful from a functional perspective, and whether one is more important than another is not obvious. Second, complexity arises due widespread variation in microscopic organization, which in turn is present because biology is organized top-down. Natural selection acts directly on features at a high level, propagating genes of organisms with beneficial emergent, functional properties. In effect, some microscopic properties can be explained by the large-scale features they need to produce. This contrasts the bottom-up way we think about systems in physics, where all the organizing principles are specified at a microscopic level. Third, microscopic models of biological systems are frequently complicated. However, this complexity may be of the sort that is also present in physical systems which admit simple effective descriptions under coarseg, so it is a less serious concern (Brown and Sethna [2003], Machta et al. [2013]).

One coarse-graining framework which has found some success in biology, and can potentially address both of these major concerns about complexity is the Information Bottleneck (IB) (Tishby et al. [2000]). In this approach, one specifies a signal they wish to coarse-grain, (the "input") and a so-called "relevance variable", which has correlations with various features of the signal. IB then defines a coarse-grained representation of the input that retains as much information about the relevance variable as possible, at a given level of distortion. As an example, compressing the current state of a system to predict its future reveals the hierarchical structure of *drosophila* behavior (Berman et al. [2016]) and provides optimized collective variables for protein dynamics (Wang et al. [2019]). Predictive IB has also been connected to model reduction in dynamical systems more generally (Schmitt et al. [2023]). Another valid choice of relevance variable is a subset of all degrees of freedom at equal time, as has been applied to the hippocampus (Ramirez and Bialek [2021]) and various lattice models (Koch-Janusz and Ringel [2018], Gordon et al. [2021], Lenggenhager et al. [2020], Gökmen et al. [2021]). The relevance variable additionally does not need to have a physical interpretation such as degrees of freedom at other points in space or time. It could, for example, be the class label in a deep learning classification problem (Tishby and Zaslavsky [2015]).

The information bottleneck and renormalization group seem to represent two opposing viewpoints on coarse-graining. Whereas RG is bottom-up and is used primarily as an analytical tool, IB is top-down and in most cases is solved computationally due to its formal intractability. Yet, both offer fundamental views on the nature of simplification and have attracted widespread, overlapping interest. In Chapter 2, we present a formal mapping between these two frameworks, outlining a way to interpret one type of analysis in terms of the other. To understand how the constraint of IB optimality affects an RG cutoff scheme in detail, we examine the exactly solvable case of jointly Gaussian statistics. Under this constraint, the cutoff scheme associated to an IB solution is closely related to the so-called Litim regulator (Litim [2000, 2001]), an optimized cutoff scheme which is ubiquitous in the non-perturbative RG (NPRG) literature. Crucially, the choice of relevance variable (given some input variable) on the IB side determines the definition of collective variables and the order in which they are eliminated from the theory. This addresses the complexities outlined above, since the IB relevance variable can be chosen to represent different aspects of biological function (e.g. heat shock response vs run-and-tumble statistics) at will, and the collective variables are defined top-down, in terms of this function.

The IB-RG connection maps a choice of IB relevance variable to a notion of RG scale, but in field theories there typically is a clear choice of scale. This raises the question: are some relevance variables better than others? Put differently, are there systems which admit multiple notions of scale? In Chapter 3, we show that some mixture models can be thought of as having multiple coexisting, exactly independent RG flows, each with its own notion of scale, and define this property as multi-relevance. We then apply the vertex expansion approach from NPRG to a multi-relevant model, where each of two notions of scale corresponds to the state of an unobserved latent variable. In the regime where this latent variable can be inferred using a linear classifier, the RG analysis can be applied successfully but will give different answers depending on the choice of expansion point in state space. Next, in the regime where linear estimation of the latent state fails, we find that the RG flow for PCA-like coarse-graining becomes very complicated. In particular, accurate computation of IR quantities (expectation values) requires keeping track of a large number of parameters with irrelevant engineering dimensions and large initial values (UV). This breakdown in the predictive power of RG arises both from the essentially non-polynomial nature of the multi-relevant model, as well as the fact that in such a model, the collective basis which diagonalizes the two-point function does not agree with either notion of scale. This suggests that RG schemes derived from maximum-entropy models of data from multi-relevant distributions may completely misidentify the relevant collective variables and lead to poor predictions. Moreover, given the ubiquity of mixture distributions in biology, multi-relevance as defined here may be a generally important consideration in understanding how RG techniques should be applied.

Finally, in Chapter 4 we implement an IB coarse-graining scheme to reveal the functionally relevant collective variables in the response data of salamander eye output cells viewing naturalistic movie data. The vertebrate retina has been shown to perform predictive computation on incoming visual signals (Palmer et al. [2015], Sederberg et al. [2018]), and it has been hypothesized that further prediction occurs at each successive layer of the visual stream. We take the viewpoint of hypothetical downstream predictor neurons and search for features which optimally encode information about future responses. Across stimuli and prediction intervals, we find that all such information is compressible into a few (less than 10) linear collective variables of the present retinal output state. By leveraging variational inference and repeated stimulus trials in our dataset, we find that this predictive information is collectively encoded; it is mostly carried by correlations between neurons. At short timescales, individual effects matter more and noise autocorrelations contribute significantly, while at later timescales predictive features are highly collective and stimulus-induced correlations dominate. Our analysis demonstrates the feasibility of uncovering biologically relevant correlation structure in high-dimensional data using variational inference and basic machine learning tools.

For the sake of completeness, we also include work from an experimental project in soft matter physics. This project examined contact electrification of dielectric grains, which forms the basis for a myriad of physical phenomena. Despite its ubiquity, even the basic aspects of collisional charging between grains are still unclear. To address this, we develop a new experimental method, based on acoustic levitation, which allows us to controllably and repeatedly collide two sub-millimeter grains and measure the evolution of their electric charges. This is therefore the first tribocharging experiment to provide complete electric isolation for the grain-grain system from its surroundings. We use this method to measure collisional charging rates between pairs of grains for three different material combinations: polyethylene-polyethylene, polystyrene-polystyrene, and polystyrene-sulfonated polystyrene. The ability to directly and noninvasively collide particles of different constituent materials, chemical functionality, size, and shape opens the door to detailed studies of collisional charging in granular materials.

CHAPTER 2

GAUSSIAN INFORMATION BOTTLENECK AND THE NON-PERTURBATIVE RENORMALIZATION GROUP

AGK[1] and Stephanie E. Palmer[1,2]

1. *Department of Physics, The University of Chicago, Chicago IL 60637*

2. *Department of Organismal Biology and Anatomy, The University of Chicago, Chicago IL 60637*

2.1 Abstract

The renormalization group (RG) is a class of theoretical techniques used to explain the collective physics of interacting, many-body systems. It has been suggested that the RG formalism may be useful in finding and interpreting emergent low-dimensional structure in complex systems outside of the traditional physics context, such as in biology or computer science. In such contexts, one common dimensionality-reduction framework already in use is information bottleneck (IB), in which the goal is to compress an "input" signal X while maximizing its mutual information with some stochastic "relevance" variable Y. IB has been applied in the vertebrate and invertebrate processing systems to characterize optimal encoding of the future motion of the external world. Other recent work has shown that the RG scheme for the dimer model could be "discovered" by a neural network attempting to solve an IB-like problem. This manuscript explores whether IB and any existing formulation of RG are formally equivalent. A class of soft-cutoff non-perturbative RG techniques are defined by families of non-deterministic coarsening maps, and hence can be formally mapped onto IB, and vice versa. For concreteness, this discussion is limited entirely to Gaussian statistics

(GIB), for which IB has exact, closed-form solutions. Under this constraint, GIB has a semigroup structure, in which successive transformations remain IB-optimal. Further, the RG cutoff scheme associated with GIB can be identified. Our results suggest that IB can be used to *impose* a notion of "large scale" structure, such as biological function, on an RG procedure.

2.2 Introduction

An overarching theme in the study of complex systems is effective low-dimensionality. We are content, for example, with the existence of laws of fluid dynamics whose few phenomenological parameters accurately account for the macroscopic behavior of many completely different fluids. We are also confident that the laws are insensitive to the particular microscopic configuration of a fluid at any given time. These are connected, but different notions of low-dimensionality; the first deals with simplification in model space, while the second refers to the emergence of collective modes, of which relatively few, when compared to the total number of degrees of freedom, will be important. A central result of Wilson's renormalization group (RG) formulation is that an effective low-dimensional model of a system may be found through repeated coarsening of the microscopic or "bare" model. In other terms, by successively removing dynamical degrees of freedom from the system description, the effective model "flows" towards a description involving very few parameters. In general, there are many strategies which can be used to simplify the description of a high-dimensional system, and RG methods, though vast in breadth, form only a subset of these. An altogether different dimensionality reduction framework is the information bottleneck (IB), which attempts to compress (or more accurately *coarsen*) a signal while keeping as much information about some pre-defined "relevance" variable as possible (Tishby et al. [2000]). Both IB and RG have been applied in theoretical neuroscience (Creutzig et al. [2009], Chalk et al. [2018], Palmer et al. [2015], Meshulam et al. [2019]), computer science (Mehta and Schwab [2014], Tishby and Zaslavsky [2015], Shwartz-Ziv and Tishby [2017], Alemi et al. [2019], Kolchinsky et al. [2019], Saxe et al. [2019]), and other frontier areas of applied statistical physics (Koch-Janusz and Ringel [2018], Gordon et al. [2021]). Given the ubiquitous need to find simplifying structure in complex models and data, a synthesis of the ideas present in IB and RG could yield powerful new analysis methods and theoretical insight.

IB and RG are motivated by different problems—IB is typically thought of as a data analysis tool, while RG is an analytical tool. Yet both describe coarsening schemes which, as we demonstrate here, can be exactly identified with each other. What purpose does this connection serve? In RG, not all coarsening schemes are created equal. For a given notion of emergent structure, such as the patterning of the degrees of freedom within a particular phase, one needs to properly define the RG procedure to capture that structure. Frequently, this emergent property is complicated, and one does not know *a priori* how it arises from the microscopic degrees of freedom, or indeed how to coarsen them to reveal the appropriate low-dimensional structure. IB can be thought of as a way of picking an RG scheme in which the notion of relevance is *defined* by the practitioner, and collective modes pertinent to that structure are *discovered* and not guessed at. In a sense, IB formalizes the problem of picking an RG scheme that preserves relevant collective behavior, while also generalizing what is meant by relevance.

Probability-theoretic investigations of renormalization group methods are not a recent development (Jona-Lasinio [2001]). One early paper by Jona-Lasinio used limit theorems from probability theory to argue the equivalence of older, field-theoretic RG formalism due to Gell-Mann and Low with the modern view due to Kadanoff and Wilson (Jona-Lasinio [1975]). Recent work (Apenko [2012], Machta et al. [2013], Bény and Osborne [2015], Raju et al. [2018], Lenggenhager et al. [2020], Gordon et al. [2021]) has focused on connections of RG to information theory. Since the general goal in RG is to remove information about some modes or system states through coarsening, an effective characterization of RG explains how the information loss due to coarsening generates the RG flow or relates to existing notions of emergence. Moreover, like the probabilistic viewpoint promoted by Jona-Lasinio, the information-theoretic viewpoint enjoys a healthy separation from physical context. The hope is that, by removing assumptions about the particular organization or interpretation of the degrees of freedom in the system, RG methods can be generalized and made applicable to problems outside of a traditional physics setting (Meshulam et al. [2019], Bradde and Bialek [2017]). This viewpoint also has the potential to enrich traditional RG applications, as Koch-Janusz et al. point out (Koch-Janusz and Ringel [2018]). Their neural-network implementation of an IB-like coarsening scheme was able to "discover" the relevant, large-scale modes of the dimer model, whose thermodynamics are completely entropic, and whose collective modes do not resemble the initial degrees of freedom. More recently, Gordon et al. built upon this scheme to formally connect notions of "relevance" between IB and RG (Gordon et al. [2021]).

In contrast to most RG formulations which require an explicit notion of how the modes of the system should be ordered, the information bottleneck approach defines the relevance of a feature by the information it carries about a specified relevance variable. To be concrete, let X be a random variable, called the "input," which we wish to coarsen. Then, let Y be another random variable, called the "relevance variable," which has some statistical interaction with the input X. IB defines a non-deterministically coarsened version of X, $\tilde{X}$, which is optimal in the sense that the mutual information (MI) between $\tilde{X}$ and Y is maximized. Because $\tilde{X}$ is defined as a non-deterministic coarsening of X, an exact correspondence between RG and IB demands that the RG scheme uses what is known as a "soft" cutoff. This means, for example, that the ubiquitous perturbative momentum shell approach put forth by Wilson cannot be mapped exactly onto IB under the interpretation of $\tilde{X}$ as some coarse-grained variable. The trade-off between degree of coarsening, indicated by $I(\tilde{X};X)$ and the amount of relevant information retained $I(\tilde{X};Y)$ is controlled by a continuous variable, denoted β.

Formally, the non-deterministic map which yields $\tilde{X}$ from X is found by optimizing the IB objective function,

$$P_\beta(\tilde{x}|x) = \underset{P(\tilde{x}|x)}{\arg\min}\, I(X;\tilde{X}) - \beta I(\tilde{X};Y). \tag{2.1}$$

For large values of β, the compressed representation $\tilde{x}$ is more detailed and retains a greater deal of predictive information about Y. Conversely, for smaller β, relatively few features are kept, in favor of reducing $I(\tilde{X};X)$ (increasing compression/coarsening). The formalism investigated here is the one originally laid out in 2000 by Tisbhy et al. (Tishby et al. [2000]), but since then a number of thematically similar IB schemes have been proposed (Slonim et al. [2006], Slonim and Tishby [2000], Strouse and Schwab [2017]). IB methods have been employed extensively in computer science, specifically towards artificial neural networks and machine learning (Mehta and Schwab [2014], Tishby and Zaslavsky [2015], Shwartz-Ziv and Tishby [2017], Alemi et al. [2019], Kolchinsky et al. [2019], Saxe et al. [2019]). In theoretical neuroscience, Palmer et al. have demonstrated using IB that the retina optimally encodes the future state of some time-correlated stimuli, suggesting that prediction is a biological function instantiated early on in the visual stream (Palmer et al. [2015], Salisbury and Palmer [2016]). IB has also been applied in studies of other complex systems, for instance to efficiently discover important reaction coordinates in large MD simulations (Wang et al. [2019]), and to rigorously demonstrate hierarchical structure in the behavior of Drosophila over long timescales (Berman et al. [2016]).

From a broad perspective, there are some basic similarities between RG and IB. Both frameworks entail a coarsening procedure by which the irrelevant aspects of a system description are discarded in order to generate a lower-dimensional, “effective” picture. Further, the Lagrange multiplier β in IB, which parameterizes the level of detail retained, can be seen as roughly analogous to the scale cutoff present in some implementations of RG. As a first guess, one might imagine that X in IB roughly corresponds to the (fluctuating) bare state of a system we are interested in renormalizing, and its compressed representation $\tilde{X}$

is a coarsened dynamical field akin to a fluctuating "local" order parameter. However, it is not difficult to find implementations of RG which do not map to IB in this way, and vice versa. For example, in Wilsonian RG schemes with a hard momentum cutoff, the decimation step represents a deterministic map from bare to coarsened system state. Together with our provisional interpretation, this contradicts the original formulation of IB, in which the coarsening is non-deterministic [1].

Another, more serious discrepancy is due to the expected use cases of these two theoretical frameworks. Generically, the fixed point description of criticality offered by RG is legitimate due to the presence of infinitely many interacting degrees of freedom, otherwise the coarsened model cannot be mapped back into the original model space. In IB, the random variables X is finite-dimensional, such as a finite lattice of continuous spins, and "dimensional reduction" does not refer to the convergence towards a low-dimensional critical manifold in model space, but instead the actual removal of dimensions from the coarsened representation of X. Finally, and perhaps most dauntingly, there is not an obvious equivalent of the IB relevance variable Y in RG. It seems counterintuitive that one would want more control over the collective mode basis used to describe a system, when for the vast majority of RG applications, length or energy scale works perfectly well as a cutoff.

Despite these apparent mismatches, there are some significant structural similarities between IB for continuous variables and a class of RG implementations involving soft cutoffs. For concreteness, we restrict our discussion of the correspondence to Gaussian statistics. While this precludes the analysis of non-Gaussian criticality, it allows all of the results to be expressed analytically and makes connections more transparently. This can also serve as a basis for later investigations involving non-Gaussian statistics and interacting systems. To begin, we show that Gaussian information bottleneck (GIB) (Chechik et al. [2005]) exhibits

[1] Alternative IB frameworks have been proposed which result in deterministic mappings (Slonim and Tishby [2000], Strouse and Schwab [2017]), and these could conceivably be connected to hard-cutoff RG schemes, though some issues occur for continuous random variables. We restrict our present discussion to the original framework.

a semi-group structure in which successive IB coarsenings compose larger IB coarsenings. This structure is summarized in an explicit function of the Lagrange multiplier β which simply multiplies under semigroup action and is therefore analogous to the length scale in canonical RG. Next we explore how the coarsening map $P(\tilde{x}|x)$ provided by IB defines an infra-red regulator which serves as a soft cutoff in several non-perturbative renormalization group (NPRG) schemes. This relation shows that the freedom inherent in choosing a cutoff scheme maps directly to the choice of Y-statistics in IB. Finally, we use a Gaussian field theory as a toy model to explore the physical significance of this fact. One result is that the RG scheme provided by IB can select a collective mode basis which is not Fourier, and hence impose a cutoff which cannot be interpreted as a wavenumber. Additionally, in whichever collective mode basis is chosen, the shape of this IB cutoff scheme is closely related to the Litim regulator which is ubiquitous in NPRG literature (Litim [2001]).

2.3 Semigroup structure in Gaussian Information Bottleneck

Every IB problem begins with the distribution $P(x, y)$, which specifies the statistical dependencies linking the input variable X to the relevance variable Y. Gaussian information bottleneck (GIB) refers to the subset of IB problems in which $P(x, y)$ is jointly Gaussian. Under this constraint, a family of coarsening maps $P_\beta(\tilde{x}|x)$ can be found exactly for all β. Chechik et al. (Chechik et al. [2005]) showed this by explicitly parameterizing the coarsening map, then minimizing the IB objective function with respect to these parameters. Their parameterization consists of two matrices A and Σ_ξ, which are used to define the compressed representation $\tilde{X}$ as a linear projection of the input plus a Gaussian "noise" variable ξ. Explicitly, $\tilde{X} = AX + \xi$ with $\xi \sim \mathcal{N}(0, \Sigma_\xi)$. Under this parameterization, one exact solution is given by

$$\begin{cases} \Sigma_\xi = I \\ A(\beta) = \operatorname{diag}\{\alpha_i(\beta)\}\, V^T \\ \alpha_i(\beta) = \left[\frac{\beta(1-\lambda_i)-1}{\lambda_i s_i}\right]^{1/2} \Theta\left(\beta - \frac{1}{1-\lambda_i}\right), \end{cases} \tag{2.2}$$

where Θ is the Heaviside step function and $s_i = [V^T \Sigma_X V]_{ii}$. The matrix V represents a set of eigenvectors with corresponding eigenvalues λ_i in the following way,

$$\Sigma_X^{-1} \Sigma_{X|Y} V = V \operatorname{diag}\{\lambda_i\}\ , \tag{2.3}$$

where the matrices Σ_X and $\Sigma_{X|Y}$ are the covariances of the distributions $P(x)$ and $P(x|y)$, respectively. These distributions are assumed to be known since they are determined by $P(x, y)$. The matrix $\Sigma_X^{-1}\Sigma_{X|Y}$ used above also appears in canonical correlation analysis and we therefore refer to it as the "canonical correlation matrix". Note that since it is not generally symmetric, the eigenvector matrix V is not generally orthogonal. An important property of the canonical correlation matrix is that its eigenvalues are real and lie within the unit interval; that is, $\lambda_i \in (0, 1)$ for all i.

The GIB solution (2.2) is not unique. At a cursory level, this follows from the IB objective function (2.1), which is a function only of mutual information terms and hence invariant to all invertible transformations on X, $\tilde{X}$, and Y. However, not all invertible transformations $X \to f(X)$ will leave the joint distributions $P(x, y)$ and $P_\beta(x, \tilde{x})$ Gaussian. It is specifically invertible linear transformations $X \to LX$ (and analogous transformations for $\tilde{X}$ and Y) which preserve IB optimality and leave all joint distributions Gaussian. One consequence of this is that $\tilde{X} \to L\tilde{X}$ changes the coarsening parameters $(A, \Sigma_\xi) \to (LA, L\Sigma_\xi L^T) = (A', \Sigma'_\xi)$. If L is invertible, then these new parameters also solve GIB. When testing whether a given parameter combination (A, Σ_ξ) is GIB-optimal, it is therefore useful to consider the quantity $V^{-1}A^T\Sigma_\xi^{-1}AV^{-T}$, which is invariant to all invertible linear transformations on $X, \tilde{X}$, and

Y.

In this section, we show that solutions to GIB have an exact semigroup structure, wherein two GIB solutions "chained together" compose a larger solution which is still optimal. To be more precise, let $P(x, y)$ be jointly Gaussian, then suppose $P_{\beta_1}(\tilde{x}_1|x)$ is IB optimal. Because $P_{\beta_1}(\tilde{x}_1|x)$ is Gaussian under the parameterization $\tilde{X}_1 = A_1 X + \xi_1$, it must also be that $P(\tilde{x}_1, y)$ is jointly Gaussian and thus a valid starting point for a new GIB problem. Taking $\tilde{X}_1$ to be the new input variable, let the second optimal coarsening map be $P_{\beta_2}(\tilde{x}_2|x)$, and parameterize it the same way: $\tilde{X}_2 = A_2 \tilde{X}_1 + \xi_2$. Then, we claim, the composition of these two coarsening maps, obtained by integrating the expression $P_{\beta_2}(\tilde{x}_2|\tilde{x}_1)P_{\beta_1}(\tilde{x}_1|x)$ over $\tilde{x}_1$, is also given by a single IB-optimal coarsening $P_\beta(\tilde{x}|x)$ for some $\beta = \beta_2 \circ \beta_1$, where $\circ$ is a binary operator whose explicit form will be provided shortly. We represent this composition schematically with the Markov chain,

$$Y \leftrightarrow X \xrightarrow{\beta_1} \tilde{X}_1 \xrightarrow{\beta_2} \tilde{X}_2. \tag{2.4}$$

To simplify the analysis, we begin by redefining [2] the input variable X by projecting it onto the eigenvectors of the canonical correlation matrix. Assuming that V is full-rank,

$$X \to V^T X \tag{2.5}$$

is an invertible linear transformation. Invertibility guarantees that the objective function is unaffected, while linearity guarantees that $P(y, x)$ remains Gaussian. Additionally, after the first compression to $\tilde{X}_1$, the new analogous quantities, e.g. $\Sigma_{\tilde{X}_1}$, $\Sigma_{\tilde{X}_1|Y}$, and A_2 will remain diagonal. For the transformation matrices A_1 and A_2, this fact can be seen by inspecting (2.2), while Lemma B.1. in (Chechik et al. [2005]) proves that Σ_X and $\Sigma_{X|Y}$ are diagonal.

[2] To clarify notation: by $X \to LX$, we mean that each instance of LX should be replaced by X.

In this new basis, they are given by

$$(\Sigma_X)_{ij} = s_i \delta_{ij} \,, \tag{2.6}$$

$$(\Sigma_{X|Y})_{ij} = s_i \lambda_i \delta_{ij} \,. \tag{2.7}$$

We now show that successively applied GIB compression as portrayed in (2.4) composes GIB transformations of greater compression. A more detailed treatment is given in appendix 2.7. Suppose that A and Σ_ξ describe a non-deterministic map $AX + \xi$. From Lemma A.1. in (Chechik et al. [2005]), this map (A, Σ_ξ) is IB-optimal if there exists some β such that

$$[A^T \Sigma_\xi^{-1} A]_{ij} = \alpha_i^2(\beta)\delta_{ij} \,, \tag{2.8}$$

where α_i is as given in (2.2).

Consider two successive maps with bottleneck parameters β_1 and β_2, each with unit noise. The composition of these transformations is represented by the pair $(A, \Sigma_\xi) = (A_2 A_1, A_2 A_2^T + I)$. Both A_1 and A_2 can be computed explicitly using (2.2), though A_2 is initially given in terms of the statistics $P(\tilde{x}_1, y)$. Using $\tilde{X}_1 = A_1 X + \xi_1$, we thus re-write A_2 in terms of the original relevance variable-input variable statistics $P(x, y)$. After this substitution, direct evaluation of $A^T \Sigma_\xi^{-1} A$ shows that $(A_2 A_1, A_2 A_2^T + I)$ is IB optimal.

$$[A_1 A_2 (A_2^2 + I)^{-1} A_2 A_1]_{ij} = \alpha_i^2(\beta_2 \circ \beta_1)\delta_{ij} \,, \tag{2.9}$$

where $\beta_2 \circ \beta_1$ is the bottleneck parameter of the full, 1-step compression,

$$\beta_2 \circ \beta_1 = \frac{\beta_2 \beta_1}{\beta_2 + \beta_1 - 1} \,. \tag{2.10}$$

It is important to note that this computation *defines* the binary operator $\circ$. If GIB did not have a semigroup structure, it would not be possible to identify $\circ$ in this manner. Direct computations show that this operator satisfies closure and associativity, and thus furnishes the space in which β values live, that is $\mathbb{R} > 1$, with a semigroup structure. As bonuses, if we consider $\beta = \infty$ to be an element, we see that it is the identity element. This aligns with the fact that in the limit $\beta \to \infty$, the IB objective (2.1) becomes insensitive to the encoding cost $I(X; \tilde{X})$ and hence no coarsening occurs; $\tilde{X}$ becomes a deterministic function of every component of X which contains information about Y. Further, $\circ$ is symmetric. One should be careful to note, however, that the maps $P_{\beta_1 \circ \beta_2}(\tilde{x}|x)$ and $P_{\beta_2 \circ \beta_1}(\tilde{x}|x)$ need only agree in the overall level of compression achieved, and may otherwise differ since $\tilde{X} \to L\tilde{X}$ is a symmetry.

2.3.1 What is the significance of this structure?

A broad goal of this paper is to explore structural similarities between IB and RG. The semigroup structure present in Wilsonian RG is crucial to its explanation of scaling phenomena, so its presence in GIB is a promising sign. The traditional picture is this: consider the RG transformations $\mathcal{R}_{b_1}$ and $\mathcal{R}_{b_2}$ which rescale length by factors b_1 and b_2, respectively. Then a fundamental property of $\mathcal{R}$ is that $\mathcal{R}_{b_1}\mathcal{R}_{b_2} = \mathcal{R}_{b_1 b_2}$. This structure imposes a strong constraint on the behavior of the flow near a fixed point. If σ represents an eigenvector of the Jacobian matrix at the fixed point, then its associated eigenvalue λ_σ will scale as b^{y_σ}(Goldenfeld [2018]). In short, the semigroup typically allows one to define the critical exponent y_σ.

The operator $\circ$ we introduced does not immediately lend itself to this sort of analysis. However, we can introduce a function $b(\beta)$ which satisfies $b(\beta_2 \circ \beta_1) = b(\beta_2)b(\beta_1)$. By inspection, this function is given by

$$b(\beta) = \frac{\beta}{\beta - 1}\,. \tag{2.11}$$

This quantity is interesting because it is analogous to the length-rescaling factor found in typical Wilsonian or Kadanoff RG schemes, yet in IB there is no need for a notion of space, and hence rescaling length generally means nothing. Compare this with, for example, a momentum-shell decimation scheme. One identifies the rescaling factor by comparing the new and old UV cutoffs, and so it acquires the meaning of a length-rescaling factor. Here, b is determined entirely by the Lagrange multiplier β and the structure of GIB, both of which are defined without deference to an existing notion of spatial extent. As discussed in the introduction, connecting IB to RG is attractive, in part, precisely because IB is an information-theoretic framework and does not rely on physical interpretations. Hence this rescaling factor b should be considered an information-theoretic quantity in the same way as β.

Can $b(\beta)$ as defined above be used in the same way as the rescaling factor b is used in RG? First, limits of the IB problem involving extremal values of β should match intuition about b in an RG context. Indeed, for $\beta \to \infty$, the zero-coarsening limit, $b(\beta) \to 1$. Next, by the data processing inequality, at $\beta = 1$ the optimal IB solution is degenerate with complete coarsening, i.e. $\tilde{X}$ becomes independent of X. Correspondingly, the limit $\beta \to 1$ gives $b(\beta) \to \infty$. Next, let us recall the scope of the GIB framework. GIB makes statements only about completely Gaussian statistics, so no anomalous scaling will appear, and thus a discussion of critical exponents is hard to motivate. Second, GIB is defined for finite-dimensional X and Y, so we cannot simply connect it to, say, momentum-shell Wilsonian RG, which only makes statements about infinite systems. Finally, and related to the last point, we have not identified yet what the analogous "model space" is in the context of IB, or how an optimal GIB map could represent an RG transformation in that space. This will be the subject of the next section, where we show that the non-deterministic nature of IB coarsening aligns exactly with existing soft-cutoff RG methods.

Whether or not this analysis helps to formally connect IB and RG, it is interesting to ask whether other IB problems exhibit semigroup structure. One could imagine, for example, that a series of high-β compression steps (low-compression limit) might be easier than one large compression step. If this is the case, IB problems with semigroup structure may benefit from an iterative chaining scheme similar to the one we present here. One possible application of this structure is the construction of feed-forward neural networks with IB objectives. If the IB problem in question has semigroup structure, then the task of training the entire network can be reduced to training the layers one-by-one on smaller (higher-compression) IB problems. This has benefits in biological systems, such as biochemical and neural networks, where processing is often hierarchical, likely as a result of underlying evolutionary and developmental constraints. Biological systems are also shaped by their output behavior, which sets a natural relevance variable in the arc from sensation to action.

2.4 Structural similarities between IB and NPRG

2.4.1 Soft-cutoff NPRG is a theory of non-deterministic coarsening

The renormalization group is not a single coherent framework, but rather a collection of theories, computational tools, and loosely-defined motifs. As such, it is probably not possible to succinctly define RG on the whole. A common theme, at least, is that RG techniques describe how the effective model of a given system changes as degrees of freedom are added or removed. The modern view of RG theory, which is largely due to Wilson (Wilson [1971a,b, 1974], Wilson and Fisher [1972], Fisher [1998]) and Kadanoff (Kadanoff [1966]), concerns itself with the removal of degrees of freedom through a process known as decimation, in which a thermodynamic quantity (typically the partition function) is re-written by performing a configurational sum or integral over a subset of the original modes. Here, even before discussing rescaling and renormalization, we must make procedural choices. To begin, one must specify the subset of degrees of freedom which are to be coarsened off. In theories where modes are labelled by wavenumber or momentum, one typically establishes a cutoff and decimates all modes with momentum above it. As a result, those modes are completely removed from the system description, and their statistics are incorporated into the couplings which parameterize the new effective theory. Another consideration is the practicality of carrying out such a procedure. If the model in consideration can be expanded in a perturbation series about a Gaussian model, and if the non-Gaussian operators are irrelevant or marginal under the flow, then this analysis is amenable to perturbative RG. However, this is often not the case, for example in systems far from their critical dimension, or in non-equilibrium phase transitions, where there may not even be critical dimensions (Canet et al. [2004, 2005]).

In non-perturbative RG (NPRG) approaches, the need for a perturbative treatment is removed by working from a formally exact flow equation at the outset. The first such treatment was put forth in 1973 by Wegner and Houghton, who used Wilson's idea of an infinitesimal momentum-shell integration to derive an exact flow equation for the full coarse-grained Hamiltonian (Wegner and Houghton [1973]). Because this equation describes the evolution of the Hamiltonian for every field configuration, this and other NPRG flow equations are called integro-differential equations, and the NPRG is sometimes referred to as the functional renormalization group (FRG). Later, Wilson and Kogut (Wilson [1974]), as well as Polchinski (Polchinski [1984]), proposed new NPRG flow equations in which the cutoff was not described explicitly through a literal demarcation between included and excluded modes, but instead through non-deterministic coarsening, so that the effective Hamiltonian satisfies a functional generalization of a diffusion equation [3]. These approaches were introduced, at least in part, as a response to difficulties [4] that arise from the sharp cutoff in the

[3] This interpretation is explicitly presented in the Wilson-Kogut paper. Given that their approach is formally equivalent with the one put forth by Polchinski, the interpretation should apply to that framework as well.

[4] Under the sharp cutoff construction, some issues include the generation of non-local position-space interactions (Wilson [1974]), unphysical nonanalyticities in correlation functions, and the need to evaluate ambiguous expressions such as $\delta(x)f(\Theta(x))$ where the function f is not known(Kopietz et al. [2010]).

Wegner-Houghton construction. Correspondingly, the Wilson-Polchinski FRG approach can be thought to give a soft cutoff, where modes can be “partially coarsened".

The most common NPRG approach in use today was first described in 1993 by C. Wetterich (Wetterich [1993b]). Like the Wilson-Polchinski NPRG, the Wetterich approach uses a soft cutoff, but the objects computed by this framework are fundamentally different. Instead of computing the effective Hamiltonian of modes which are below the cutoff, the Wetterich framework computes the effective free energy of modes *above* the cutoff. For this reason, we say that the Wilson-Polchinski framework is UV-regulated and Wetterich is IR-regulated. Yet, despite this difference in perspective, the Wetterich formalism still describes the flow effective models make from their microscopic to macroscopic pictures. In this section, we will explore how the soft-cutoff construction is related to a notion of non-deterministic coarsening, and in turn, the information bottleneck framework. An in-depth discussion of the philosophy and implementation of NPRG techniques would be distracting, so we instead refer the reader to a number of good references on the topic (Bagnuls and Bervillier [2001], Berges et al. [2002], Delamotte [2012], Kopietz et al. [2010]).

So far we have not explained how one actually imposes a soft-cutoff scheme. We begin by examining the Wetterich setup, in which one writes the effective (Helmholtz) free energy at cutoff k,

$$\begin{aligned} W_k[J] &= \log \int \mathcal{D}\chi \exp \Bigg[- S[\chi] - \Delta S_k[\chi] \\ &\qquad + \sum_a \int \mathrm{d}^d x \, J_a(x) \chi_a(x) \Bigg] . \end{aligned} \tag{2.12}$$

The bare action, given by S, is the microscopic theory which is known. The source J allows us to take (functional) derivatives of this object to obtain cumulants (connected Green’s functions). The remaining term $\Delta S_k[\chi]$ is known as the deformation, and it is this term which enforces the cutoff. It is written as a bilinear in χ,

$$\Delta S_k[\chi] = \frac{1}{2} \sum_{ab} \int \mathrm{d}^d x \, \mathrm{d}^d y \, [R_k]_{ab}(x, y) \chi_a(x) \chi_b(y) \, . \tag{2.13}$$

For compactness, we will often resort to a condensed notation and express integrals instead as contraction over suppressed continuous indices. For example, the deformation may be re-written

$$\Delta S_k[\chi] = \frac{1}{2} \chi^\dagger R_k \chi \, . \tag{2.14}$$

The kernel (matrix) R is known as the regulator, and it controls the "shape" of the cutoff. Almost always, it is chosen to be diagonal in Fourier basis so that the cutoff k has the interpretation of a wavenumber or momentum. The resulting Fourier-transformed regulator $R_k(q)$ has some freedom in its definition, but it must satisfy the following properties (Litim [2001]):

1. $\lim_{q^2/k^2 \to 0} R_k(q) > 0$
2. $\lim_{k^2/q^2 \to 0} R_k(q) = 0$
3. $R_k(q) \to \infty \quad \forall q \quad \text{as} \quad k \to \infty$

These constraints guarantee that the deformation acts as an IR cutoff. The first condition increases the effective mass of low-momentum modes and suppresses their contribution to the effective free energy. The second ensures that modes with high momentum ($q^2 > k^2$) are left relatively unaffected, and contribute more fully to W_k. The third condition ensures that the so-called "effective action," defined as

$$\Gamma_k[\varphi] = J^\dagger \varphi - W_k[J] - \Delta S_k[\varphi] \, , \tag{2.15}$$

approaches the bare action (or Hamiltonian, as the case may be) in the limit $k \to \infty$. Here, the order parameter φ is given by $\delta W_k[J]/\delta J^\dagger$. Because of this construction, the second regulator property also ensures that in the limit $k \to 0$, the deformation ΔS_k disappears, and the effective action Γ_k becomes the Legendre transform of $W[J]$. This functional $\Gamma_{k=0}$ is known in many-body theory as the 1PI generating functional, and in statistical mechanics as the Gibbs free energy. In the Wetterich formalism, one is generally interested in computing the flow of Γ_k because of these useful boundary conditions.

To see how this approach is related to non-deterministic coarsening, we will connect it to a soft-cutoff UV-regulated approach, also put forth by Wetterich, which is formally equivalent to the Wilson-Polchinski framework. We begin with the following expression defining the average action $\Gamma_k^{\text{av}}[\tilde{\chi}]$, taken directly from the paper (Wetterich [1993a]), with only a slight change in notation

$$-\Gamma_k^{\text{av}}[\tilde{\chi}] = \log \int \mathcal{D}\chi \, P_k[\tilde{\chi}|\chi] \exp(-S[\chi]) \,, \tag{2.16}$$

where we refer to this functional $P_k[\tilde{\chi}|\chi]$ as the coarsening map. If we were interested in performing deterministic coarsening, i.e. one involving a hard cutoff, the coarsening map would be something like a delta-function $\delta(\tilde{\chi} - \Phi_k[\chi])$ for some functional Φ_k. However, in all soft-cutoff UV-regulated approaches, this distribution is Gaussian in $\tilde{\chi}$

$$P_k[\tilde{\chi}|\chi] = \exp\left[-\frac{1}{2}\left(\tilde{\chi} - A_k\chi\right)^{\dagger} \Delta_k^{-1} \left(\tilde{\chi} - A_k\chi\right) - C_k\right] . \tag{2.17}$$

In principle, given the coarsening parameters A_k and Δ_k for all k, the exact flow equation for Γ_k^{av} is determined. Wetterich gives explicit choices for these parameters, while Wilson and Polchinski independently give their own (though in slightly different fashion). The term C_k is a normalizing constant which is essentially unimportant to the remainder of our discussion.

Now we connect the IR and UV approaches to show that they are complementary, and in some sense, equivalent. In particular, suppose we know $P_k[\tilde{\chi}|\chi]$ for all k. Then, from this single object, one can construct both the IR-regulated and UV-regulated flows. This should make intuitive sense; the IR-regulated part tracks the thermodynamics of the already-integrated modes, while the UV-regulated part tracks the model of the unintegrated modes. This can all be seen clearly by writing out the full sourced partition function $Z[J]$ and invoking the normalization of the coarsening map:

$$\begin{aligned} Z[J] &= \int \mathcal{D}\chi \, \exp\left(-S[\chi] + J^\dagger \chi\right) \\ &= \int \mathcal{D}\tilde{\chi}\, \mathcal{D}\chi \, P_k[\tilde{\chi}|\chi] \exp\left(-S[\chi] + J^\dagger \chi\right) \\ &\sim \int \mathcal{D}\tilde{\chi} \, \exp\left(-\frac{1}{2}\tilde{\chi}^\dagger \Delta_k^{-1} \tilde{\chi} + W_k[J + \tilde{J}[\tilde{\chi}]]\right) \end{aligned} \tag{2.18}$$

In the final expression, the normalizing constant C_k has been dropped. Readers familiar with the Polchinski formulation will immediately recognize $W_k[\tilde{J}[\tilde{\chi}]]$ as the effective interaction potential. However, the argument to this potential is shifted by the source J, which therefore enters nonlinearly, unlike in Polchinski's approach. This difference is due to the fact that we define a flow for each initial source configuration, instead of adding a linear source term to the vacuum flow.

To arrive at (2.18) above, we had to define the effective field-dependent source $\tilde{J}$ and identify a suitable deformation term in $P_k[\tilde{\chi}|\chi]$. By directly substituting (2.17), one can see that

$$\tilde{J}[\tilde{\chi}] = A_k^\dagger \Delta_k^{-1} \tilde{\chi}\,, \tag{2.19}$$

and

$$\Delta S_k[\chi] = \frac{1}{2}\chi^\dagger A_k^\dagger \Delta_k^{-1} A_k \chi\,. \tag{2.20}$$

As promised, the existence of a family of distributions $P_k[\tilde{\chi}|\chi]$ with a known parameterization (A_k, Δ_k) allows us to define an IR regulator scheme, and therefore compute the NPRG flow both above and below the cutoff. The deformation term ΔS_k ultimately came from the χ^2 term present in the coarsening map, which could be interpreted as a free energy. We also identify immediately that the IR regulator R_k corresponding to a given choice of coarsening map is given by $A_k^\dagger \Delta_k^{-1} A_k$.

We will next use this viewpoint to introduce information bottleneck into the discussion. In particular, we will associate the coarsening map $P_k[\tilde{\chi}|\chi]$ with the IB coarsening map $P_\beta(\tilde{x}|x)$ and examine some consequences. This discussion comes with some restrictions. Firstly, one should note that all soft-cutoff NPRG frameworks, regardless of the structure of the microscopic action, assume a Gaussian coarsening map. With a non-Gaussian $P_k[\tilde{\chi}|\chi]$, the flow may still be defined, but it will not, in general, satisfy any known exact flow equations. This is easiest to see in the IR Wetterich formalism, since a non-Gaussian P_k would yield a $\Delta S_k[\chi]$ which is no longer bilinear in χ, and hence one could not write the flow equation in terms of the exact effective propagator, as it usually is. Indeed, the more general $\Delta S_k[\chi]$ could have terms at arbitrarily high order in χ, and thus require arbitrarily high-order derivatives of Γ_k in the flow equation. So, while it is not impossible to seriously consider non-Gaussian $P_k[\tilde{\chi}|\chi]$, it is certainly inadvisable without good reason.

With this in mind, we must also note that IB has an exact solution involving Gaussian $P_\beta(\tilde{x}|x)$, but only when the variables X and Y are jointly Gaussian. By analogy, this restricts us to discussing theories where the bare action $S[\chi]$, or perhaps more accurately, the bare Hamiltonian $\mathcal{H}[\chi]$ contains only linear and bilinear terms in χ. While everything presented above holds for general S, everything that follows will be totally Gaussian so that IB optimality can be exactly satisfied. Finally, note that IB may not be well-defined for infinite-dimensional random variables such as fields, so our scope is further limited to finite-dimensional multivariate Gaussian distributions of classical variables.

2.4.2 The Gaussian IB regulator scheme

In the last section, we briefly introduced soft-cutoff NPRG approaches and argued that both UV- and IR-regulated flows can be defined given a family of Gaussian coarsening maps $P_k[\tilde{\chi}|\chi]$. Broadly, we aim to show that IB and RG can be connected by identifying this map with the IB-optimal coarsening map $P_\beta(\tilde{x}|x)$. By this we do not mean to say that the family of maps produced by IB are the correct starting point for NPRG. Instead, we simply note that IB-optimality is a constraint one could impose on the coarse-graining scheme. Assuming we do so, what characteristics does the IB-RG scheme carry?

By working within the constraints of GIB, we can write down an exact, IB-optimal Gaussian coarsening map. Then, because the NPRG coarsening scheme is also Gaussian, we can identify the structure in our IB solution which represents an IR cutoff scheme. To begin, we recall that the role of the regulator is to deform the microscopic theory through the addition of a mass-like term which "freezes out" the most relevant modes,

$$\Delta S_k[\chi] = \frac{1}{2}\chi^\dagger R_k \chi. \tag{2.21}$$

When the bare variable, x, is finite-dimensional, this is written as

$$\Delta S_\beta(x) = \frac{1}{2}x^\dagger R_\beta x \tag{2.22}$$

with R_β a positive semi-definite matrix. Following the argument in section 2.4.1, we can identify the deformation produced by a Gaussian coarsening $\tilde{X} = AX + \xi$

$$\Delta S(x) = \frac{1}{2}x^T A^T \Sigma_\xi^{-1} A x. \tag{2.23}$$

Now, by imposing IB-optimality (2.48) on (A, Σ_ξ), we find that

$$\begin{aligned} R_\beta^{(\mathrm{IB})} &= A(\beta)^T \Sigma_\xi(\beta)^{-1} A(\beta) \\ &= V \ \mathrm{diag}(\alpha_i^2(\beta)) \ V^T, \end{aligned} \tag{2.24}$$

with

$$\alpha_i^2(\beta) = \frac{\beta(1-\lambda_i)-1}{\lambda_i s_i}\Theta\left(\beta - \frac{1}{1-\lambda_i}\right) . \tag{2.25}$$

After the substitution $\beta_i = (1-\lambda_i)^{-1}$, we can write

$$\left[R_\beta^{(\mathrm{IB})}\right]_{ij} = \sum_u V_{iu}\frac{\beta-\beta_u}{s_u(\beta_u-1)}\Theta(\beta-\beta_u)V_{uj}. \tag{2.26}$$

Finally, we project the degrees of freedom into the canonical correlation basis obtained by diagonalizing the canonical correlation matrix $\Sigma_X^{-1}\Sigma_{X|Y}$, as discussed in section 2.3. That is, we take $X \to V^T X$ which takes $R \to V^{-1}RV^{-T}$, which makes R diagonal,

$$\left[R_\beta^{(\mathrm{IB})}\right]_{ij} = \frac{\beta-\beta_i}{s_i(\beta_i-1)}\Theta(\beta-\beta_i)\delta_{ij}. \tag{2.27}$$

Here the β_i are critical bottleneck values. These values β_i are given by $(1-\lambda_i)^{-1}$, where λ_i are the eigenvalues of the canonical correlation matrix. Intuitively, one can think of each eigenvalue λ_i as characterizing the amount of information that the i^{th} collective mode carries about the relevance variable Y. Smaller values of λ_i indicate modes with more information. Because the critical β_i value monotonically increases with λ_i, this same intuition holds, though $\beta_i \in (1,\infty)$, whereas $\lambda_i \in (0,1)$. The values s_i are given by $[V^T\Sigma_X V]_{ii}$, and represent the variances of the now-independent modes $\tilde{x}$. Θ denotes the Heaviside step function.

In the typical context, R is diagonalized by a Fourier transform, and thus it represents a cutoff in wavevector or momentum. Here this notion is generalized, and instead of identifying a cutoff wavenumber k, we should consider the cutoff to be of information-theoretic origin, and fundamentally defined by β. Consequently, the degree to which the mode labelled by i is coarsened should be found by comparing its corresponding critical value β_i to the cutoff β. As such, we can essentially make the replacements $k^2 \to \beta$ and $q^2 \to \beta_i$, with the caveat that β and β_i should approach unity as k^2 and q^2 go to zero.

In Figure 2.1, we plot $R^{(\mathrm{IB})}$ obtained from the first toy model presented in section 2.5.2 and compare it against the well-known Litim regulator (Litim [2001]), denoted $R^{(\mathrm{L})}$ and given in Eq. (2.28). Ignoring for now the particulars of the model, we point out that the IB and Litim regulators appear qualitatively similar, and for fixed parameters t and η, all limits involving q and k satisfy the regulator scheme requirements. Moreover, we see that the NPRG and IB notions of mode relevance are in agreement. Smaller canonical correlation eigenvalues λ (top plot) correspond to collective modes which get integrated out later in the flow. This is reflected in the structure of the soft cutoff, which increasingly suppresses fluctuations as $q \to 0$.

Is it okay to take (2.27) seriously as an IR regulator scheme? Let us attempt to compare with the conditions outlined in the last section. The typical interpretation of the first requirement on R is that the lowest energy modes should be given extra mass by the regulator so that they are “frozen out” of the configurational integral. In fewer words, there should not be soft modes in intermediate stages of the flow. By analogy, it must be true that $(R_\beta)_{11} > 0$, where we take $\beta_1 = \min_i \beta_i$ to represent the most “relevant” mode (in the IB sense). Indeed, for all β, this is satisfied by (2.27). Next, R must vanish for the i^{th} mode when the cutoff β is taken sufficiently far below β_i. Because of the step function, this is satisfied. Finally, each diagonal component $(R_\beta)_{ii}$ should diverge as $\beta \to \infty$ so that at zero compression, only the saddle point configuration of the microscopic theory contributes to the generating function, or whichever thermodynamic potential we are interested in. If β_i are all finite, then this limit holds as well [5].

Because it satisfies all of the properties required of a typical regulator in a soft-cutoff scheme, we call (2.27) the “IB regulator" and denote it $R^{(\mathrm{IB})}$. This identification has some interesting consequences, which will be explored in the coming section. One particularly

[5] Two edge cases may appear important; $\beta_i \to 1$ and $\beta_i \to \infty$. Because $\beta_i = (1-\lambda_i)^{-1}$, these correspond to limits where a mode $[V^T X]_i$ is deterministically related to Y or completely independent of Y, respectively. Complete deterministic dependence between X and Y should not be considered without modification, and in the case of independence, those modes may be removed as a formality.

striking feature is that the cutoff scheme is now parameterized by the family of distributions $P(x, y)$. In IB theory, these distributions formalize the notion of “important features” of X implicitly through its correlations with Y. This means that the RG scheme selected by a given set of IB solutions will *not* favor, for instance, “long distance modes” unless $P(x, y)$ is chosen to enforce that. Instead, the analogue of long distance modes are those modes which have the most information about Y. In section 2.5.2 we will attempt to clarify this by calculating the IB regulator explicitly in a simple, familiar context.

2.5 Consequences and interpretations of the correspondence

2.5.1 Correspondence in the non-Gaussian case

So far, we have demonstrated that the application of a soft cutoff in NPRG is equivalent to a non-deterministic coarsening of the bare system. In turn, the formal correspondence between IB and NPRG amounts to imposing IB-optimality on this coarsening map for some specified choice of a relevance variable, Y, and $P(x, y)$. This relationship does not depend on X or Y being Gaussian, nor on the semigroup structure of IB, which may or may not hold beyond Gaussian IB.

The correspondence between IB and NPRG in a non-Gaussian context is difficult to formalize for several reasons. Firstly, as pointed out in Sec. 2.4.1, virtually all soft-cutoff NPRG approaches impose a coarsening scheme which is Gaussian. That is, the renormalized degrees of freedom $\tilde{\chi}$ are some linear projection of the bare variables χ plus some uncorrelated, Gaussian noise which has fixed variance and mean. Non-Gaussian coarsening maps are not forbidden, however they are less convenient to work with, and it is probably for this reason that they are not used. Therefore, even if one had the solution to a non-Gaussian IB problem, the corresponding NPRG structure would look unfamiliar and be difficult to compare with known cutoff schemes. The second complication is that non-Gaussian IB is generally intractable from an analytical standpoint. Whereas non-Gaussian coarsening maps are problematic to NPRG merely because they are unconventional, analytic solutions to IB in systems with non-Gaussian $P(x, y)$ are not known, at least in problems relevant to RG theory.

We can, however, rely on the fact that IB has the correct limiting behaviors to allow us to make a useful connection to the notion of scale in NPRG. In the high-compression limit, $\beta \to \beta_1$ in IB, and $P_\beta(\tilde{x}|x)$ maps all degrees of freedom in x into a single degree of freedom in $\tilde{x}$. For β values below β_1, and in particular at $\beta = 1$, all modes of the system have been integrated out. This can be seen in the fact that the coarsening map degenerates to a trivial mapping where the coarsened state $\tilde{x}$ is entirely noise, and has no mutual information with x. In NPRG at this limit, the free energy contains the exact thermodynamic information because the regulator is set to zero. In the Gaussian case, we see this as the disappearance of $R_\beta^{(\mathrm{IB})}$ as β dips below β_1. On the other hand, in the low-compression limit where β is large, $P_\beta(\tilde{x}|x) \to \delta(\tilde{x} - x)$, and there is no coarsening. Comparisons at intermediate β values, which are most useful as an application of IB to data, are possible to make to NPRG when the coarsening map in RG is IB-optimal.

Moreover, the semigroup structure of GIB outlined in section 2.3 is not a necessary condition for IB-RG correspondence in a non-Gaussian context. One simple way to observe this is that semigroup structure is not necessary to define NPRG. For contrast, consider the traditional Wilsonian picture: a bare theory is given, then it is coarsened by a small amount by integrating out an infinitesimal fraction of the total collective modes in the system. After this integration, one rescales the state space and renormalizes the degrees of freedom to place the new effective theory in the original theory space. Then, by iterating this process, one defines the RG flow. Within this construction, the flow arises from a coarsening procedure which manifestly has semigroup structure. In NPRG, on the other hand, one begins with a cutoff scheme, in other words a family of coarsening maps, parameterized by scale. Then, formally exact flow equations are obtained by varying the cutoff scale infinitesimally. Importantly, no semigroup structure is ever imposed on the coarsening procedure; the only requirements are differentiability of the cutoff scheme with respect to scale and some limiting behaviors which ensure that the initial conditions of the flow are known and the final conditions leave no trace of design choices made in the RG procedure. As we discuss above, the IB-RG correspondence holds in the non-Gaussian case because the RG scheme specified by a differentiable family of IB-optimal coarsening maps satisfies these constraints.

When imposing IB-optimality on an NPRG procedure, it is important to recognize that not all IB solutions give rise to meaningful NPRG coarsening schemes. At each β, the optimal coarsening map $P_\beta(\tilde{x}|x)$ can be morphed through arbitrary invertible transformations on $\tilde{x}$ and x independently without affecting optimality. This seems at odds with a flow-equation description provided by a typical NPRG setup, in which $P_\beta(\tilde{x}|x)$ is chosen explicitly to be a differentiable function of β. One way to resolve this, given a family of solutions to an IB problem, would be to use this freedom to perform an invertible mapping and explicitly enforce continuity with respect to β. On the other hand, one could take seriously the notion that IB solutions are not fundamentally governed by a set of flow equations, but rather by a self-consistent updating scheme known as the Blahut-Arimoto (BA) procedure (Tishby et al. [2000]). This scheme iteratively re-expresses the coarsening map $P(\tilde{x}|x)$ in terms of its previous estimate, the initial statistics $P(x, y)$, and β, until convergence is reached. In appendix 2.8, we explore whether these self-consistent updates could practically replace a flow equation approach as an analytical tool. Even under Gaussian constraints, we find that parameterizing the update map in terms of the objects one would typically keep track of along an NPRG flow does not yield a simple framework. We conclude that without significant advances in IB theory, the BA map treatment is not likely to be applicable for IB-optimal NPRG analysis.

In Eq. (2.27), we present a soft cutoff scheme which arises from the constraint of GIB-optimality, but it is given in terms of quantities which have no physical context, and so it is hard to say *a priori* how it relates to existing cutoff schemes structurally. In the next section, we consider a toy model which provides this physical context and therefore affords us a glimpse into how IB-optimal NPRG schemes differ from those already employed.

2.5.2 Collective modes are not always Fourier: a minimal example

In the Wetterich NPRG, the cutoff is enforced through a deformation $\Delta S_k[\chi] = \frac{1}{2}\chi^\dagger R_k \chi$ added to the bare action or Hamiltonian. In section 2.4.2, we identified this structure as the free energy of a Gaussian coarsening map from the bare degrees of freedom χ to some compressed representation $\tilde{\chi}$. We then defined the IB regulator through the deformation produced by the map solves the Gaussian Information Bottleneck problem, and showed that it satisfies the various "design" constraints traditionally placed upon it. An immediate consequence of this construction is that the regulator design space is now parameterized by the joint distributions $P(x, y)$ which define the starting point of IB, and for many such distributions, the preferred basis selected by IB will look nothing like Fourier modes. Of course, for finite systems not organized in a lattice, this is unsurprising; the Fourier basis will not exist in any familiar sense. However, for practitioners of NPRG, it may cause discomfort to consider a regulator $R_{v(\beta)}(u)$ in which the numbers v and u do not represent radii in momentum space. In contrast, for the majority of applications, the standard cutoff scheme is provided by the Litim regulator

$$R_k^{(\mathrm{L})}(q, q') = \delta^d(q - q')(k^2 - q^2)\Theta(k^2 - q^2)\,, \tag{2.28}$$

which should be interpreted as a soft momentum-space cutoff. The Litim regulator sees widespread use both because it is optimized to give good convergence properties in certain contexts [6], and because its simple form often leads to analytically expressible flow equations (after appropriate truncation procedures) (Litim [2000, 2001]).

The IB regulator $R_\beta^{(\mathrm{IB})}$ given in (2.27) does not manifestly have any such nice qualities, and in the general case may be difficult to interpret. In this section, we calculate $R_\beta^{(\mathrm{IB})}$ explicitly in a trivial statistical field theory problem to explore its structure in a familiar context and address some of its non-intuitive features. For our model, we consider a real scalar field $\chi(x)$ in d dimensions at equilibrium and finite temperature $k_B T = 1$. This fluctuating field will serve as the "input variable" X. We also add a disordered source field $h(x)$ which will serve as the "relevance variable" Y. Together we have

$$\mathcal{H}[\chi|h] = \int \mathrm{d}^d x \left\{ \frac{1}{2}\chi(x)(t - \nabla^2)\chi(x) - h(x)\chi(x) \right\}. \tag{2.29}$$

We also give Gaussian statistics to the disorder,

$$\begin{aligned} \overline{\mathcal{A}[h]} \quad &= \quad \det(2\pi H)^{-1/2} \int \mathcal{D}h\, \mathcal{A}[h] \times \\ &\quad \exp\left(-\frac{1}{2} \int \mathrm{d}^d x_1 \mathrm{d}^d x_2\, h(x_1)[H^{-1}](x_1, x_2) h(x_2) \right). \end{aligned} \tag{2.30}$$

In our condensed notation, the above equations are re-expressed as

$$\mathcal{H}[\chi|h] \quad = \quad \frac{1}{2}\chi^T G_0^{-1}\chi - h^T\chi \tag{2.31}$$

$$\log P[h] \quad \sim \quad -\frac{1}{2} h^T H^{-1} h. \tag{2.32}$$

Together, the Boltzmann weight $\mathcal{H}[\chi|h]$ and the distribution $P[h]$ describing the disor-

[6] The Litim regulator was introduced in the context of NPRG analysis of the $O(N)$ model. Its favorable or "optimal" characteristics are manifested through improved convergence properties of so-called "threshold functions," which constitute a frequently encountered class of momentum integrals involving the cutoff.

der statistics constitute a joint distribution $P[\chi, h]$ which is jointly Gaussian and thus—momentarily casting aside worries about the continuously infinite-dimensional random variables—a valid starting point for GIB. From the IB standpoint, the goal would usually be to construct a coarsened field $\tilde{\chi}(x)$ which discards some information about χ while encoding as much as possible about the statistics of h. However, the goal here is not to discuss $\tilde{\chi}$, but rather to better understand the NPRG cutoff scheme that IB imposes as a consequence of this starting point. Since we have assumed a canonical form for the bare Green's function G_0^{-1} and the source term is $h \cdot \chi$, the only remaining control over $P[\chi, h]$ is the two-point correlation of h,

$$\overline{h(x_1)h(x_2)} = H(x_1, x_2). \tag{2.33}$$

To explore different forms of $R_\beta^{(\mathrm{IB})}$, we therefore consider three different constructions of H. First, we choose h to be totally uncorrelated at different points, with a constant variance at each point. Second, we choose H diagonal in Fourier basis, but with some dispersion that adds position-space correlations. In both of these first examples, we will arrive at regulators with momentum-space cutoffs. It is the goal of the third case to present an H which is not diagonal in momentum basis, thereby introducing a non-momentum cutoff structure.

IB regulator when disorder correlations are diagonal in momentum space

In the first and simplest case, we take H to be a δ-function multiplied by some constant factor η. Since the Fourier transform $\mathcal{F}$ is unitary [7], the momentum-space representation of

[7] We choose the formalism in which $[\mathcal{F}](k, x)$ includes a factor of $(2\pi)^{-d/2}$ so that $\mathcal{F}^\dagger \mathcal{F} = I$. Further, the appearance of factors $\delta^d(q - q')$ as opposed to the more traditional $\delta^d(q + q')$ is a consequence of our decision to conjugate with respect to $\mathcal{F}^\dagger \cdot \mathcal{F}$, instead of $\mathcal{F} \cdot \mathcal{F}$. This appeals more to the matrix multiplication shorthand.

H is unchanged from its position-space form, such that

$$\begin{aligned} H(x_1, x_2) &= \eta\delta^d(x_1 - x_2)\,; && (2.34) \\ \tilde{H}(q_1, q_2) &= [\mathcal{F}H\mathcal{F}^\dagger](q_1, q_2) && (2.35) \\ &= \eta\delta^d(q_1 - q_2)\,. \end{aligned}$$

The first step in GIB analysis is constructing the canonical correlation matrix $\Sigma_X^{-1}\Sigma_{X|Y}$, where we have chosen $X \leftrightarrow \chi$ and $Y \leftrightarrow h$. After a calculation involving only Gaussian integral identities and our definition of $P[\chi, h]$, we obtain

$$\Sigma_\chi^{-1}\Sigma_{\chi|h} = (I + HG_0)^{-1}. \quad (2.36)$$

Next, we find the right eigenfunctions $V(x, u)$ and corresponding eigenvalues $\lambda(u)$ of the correlation. For our current construction of H,

$$\begin{aligned} V(x, q) &= \mathcal{F}^\dagger(x, q) = (2\pi)^{-d/2}e^{iq\cdot x} && (2.37) \\ \lambda(q) &= (1 + \eta\tilde{G}_0(q))^{-1}\,, && (2.38) \end{aligned}$$

where $\tilde{G}_0(q) = 1/(t + q^2)$ is obtained after Fourier transform of G_0. To finally obtain the IB regulator in a familiar form, we would like to find a way to express it completely in terms of q, k, and the various other parameters introduced in this application. However, equation (2.27) gives us $R^{(\mathrm{IB})}$ in terms of the bottleneck parameter β, which has not been defined yet in this application.

The crucial insight is to note that β serves essentially the same role as k in the typical theory. To find the explicit map between the two, we use the fact that critical bottleneck values $\beta(q)$ are defined in terms of the canonical correlation eigenvalues $\lambda(q)$ through $\beta(q) =$

$(1-\lambda(q))^{-1}$. In this model, the critical bottleneck values are

$$\beta(q) = \frac{1}{\eta \tilde{G}_0(q)} + 1 \,. \tag{2.39}$$

Using this map, we can replace β with $\beta(k)$, where k is the usual momentum cutoff. Doing so, we find that the IB regulator can be neatly expressed in terms of the Litim regulator,

$$\begin{aligned} R_k^{(\mathrm{IB})}(q) &= \frac{t+q^2}{t+q^2+\eta}(k^2-q^2)\Theta(k^2-q^2) \\ &= \lambda(q) R_k^{(L)}(q) \,. \end{aligned} \tag{2.40}$$

In particular, the limit $\eta \to 0$ gives $R^{(\mathrm{IB})} \to R^{(\mathrm{L})}$. It is interesting that the Litim regulator appears in this expression, since its derivation invokes optimality principles which are not obviously connected to information bottleneck.

Momentum-space IB regulator with dispersion in disorder correlations

Without changing our decision to make H diagonal in Fourier basis, we can also add q-dependence to η. In this case, the steps taken above are essentially unchanged, and we end up with a slightly different regulator,

$$\begin{aligned} R_k^{(\mathrm{IB})}(q) &= \lambda(q)\left(\frac{\eta(q)}{\eta(k)}\tilde{G}_0^{-1}(k) - \tilde{G}_0^{-1}(q)\right) \times \\ &\quad \Theta\left(\frac{\eta(q)}{\eta(k)}\tilde{G}_0^{-1}(k) - \tilde{G}_0^{-1}(q)\right) . \end{aligned} \tag{2.41}$$

With some manipulations, one could re-write this in terms of $(1-x)\Theta(1-x)$ in order to appeal to the Litim description once again.

A new feature appears in the regulator scheme when η is given q-dependence. For extreme choices of η, the ordering of modes can actually be reversed. To see how this is possible, note that fundamentally it is the IB parameter β which sets the cutoff, while the critical values $\beta(q)$ define the mapping to q. Therefore, by picking, e.g., $\eta(q) \sim \tilde{G}_0^{-2}(q)$, one achieves a $\beta(q)$ which monotonically decreases with respect to q, meaning longer wavelength modes (lower q) actually get integrated out *before* shorter ones. However, this construction presents some pathologies and is hard to interpret in the truly continuous case, so we will not explore it further here.

2.5.3 Explicit form of the IB regulator in a more general case

In the last section we assumed a form of H which was diagonal in Fourier basis. This assumption led us to a regulator scheme which could be interpreted as a soft cutoff in momentum space. In this section we explore an example in which the disorder correlator is not diagonal in the Fourier basis. Physically, this means that the disorder statistics are no longer translationally invariant, and hence the collective modes selected by the IB procedure will not be Fourier in nature. We define H such that this translation invariance is broken but the collective modes can still be solved for exactly,

$$H = \eta\, \mathcal{F}^{\dagger} \tilde{G}_0^{-1/2} \mathcal{F}_{\alpha} \tilde{G}_0 \mathcal{F}_{\alpha}^{\dagger} \tilde{G}_0^{-1/2} \mathcal{F}, \tag{2.42}$$

where $\mathcal{F}_{\alpha}$ is the fractional Fourier transform through angle α and η is a constant that represents the strength of disorder. Under this definition, we can again compute Σ_{χ} and find the spectrum of $\Sigma_{\chi}^{-1}\Sigma_{\chi|h}$. This yields eigenfunctions analogous to the plane wave solutions in last section, but indexed by a new parameter u which can neither be interpreted as position nor wavenumber,

$$\begin{aligned} V^{\dagger}[\cdot](u) &= \left\{ \tilde{G}_0^{1/2} \mathcal{F}_{\alpha}^{\dagger} \tilde{G}_0^{-1/2} \mathcal{F} \right\} [\cdot](u) & (2.43) \\ \lambda(u) &= (1 + \eta \tilde{G}_0(u))^{-1}. & (2.44) \end{aligned}$$

Here, the notation $[\cdot]$ indicates that $V^\dagger$ is best conceptualized as a functional parameterized by u, where for instance the collective modes of $\chi(x)$ would be given by $V^\dagger[\chi](u)$. Stated differently, the leftmost operator $\tilde{G}_0^{1/2}$ is evaluated at u, and the rightmost is a Fourier transform over the integrand $[\cdot]$. Unfortunately, this solution is only formal, and cannot be visualized in the same manner as plane waves. In a true field theory, even with the trivial Gaussian setup, both $H(x_1, x_2)$ and $V(x, u)$ are poorly behaved when written as functions of x. When written as an integral in q, V diverges when $|q_{\max}| \to \infty$, and is discontinuous in both x and u. One way to conceptualize this is by comparison with G_0^{-1}, which includes ∇^2 and thus cannot be written as elementary functions of x. After Fourier transform, we can replace the operator description with a simple function of the continuous variables q. Similarly, although we cannot express H and V as functions of x, the various operators we are interested in can be written simply in the non-orthogonal basis defined by V:

$$\left[V^{-1} H V^{-\dagger}\right](u_1, u_2) = \eta \delta^d(u_1 - u_2) \tag{2.45}$$

$$\left[V^\dagger G_0 V\right](u_1, u_2) = \tilde{G}_0(u_1) \delta^d(u_1 - u_2) \tag{2.46}$$

It is hard to say what the label u physically represents beyond being a parameter that defines and orders collective modes $\chi'(u) = V^\dagger[\chi](u)$ in the system. Despite this, the regulator maintains its simple form,

$$R_v^{(\mathrm{IB})}(u) = \lambda(u) R_v^{(\mathrm{L})}(u), \tag{2.47}$$

where now v takes the role of the cutoff, replacing k as u has replaced q. That is, the collective modes labelled by u are ordered in terms of their predictiveness about the disordered source field h. GIB then imposes a soft-cutoff scheme at a scale v, which is a proxy for the bottleneck parameter β, as k was in the Fourier case. We stress that these labels v and u are defined by the correlation structure of $P[\chi, h]$ and have no simple intrinsic physical meaning. Without significantly more effort, all we can say is that a mode labelled u_1 carries more information about the disorder h than a mode labelled u_2 if $u_1 < u_2$.

Many of the difficulties present in this discussion, such as the poorly-behaved character of collective modes $V(x, u)$ and disorder correlator $H(x_1, x_2)$, as well as the non-intuitive nature of the mode labels u and v, stem from a common cause. IB is only suited to analysis of systems with finitely many degrees of freedom, and field theories have infinitely many. The calculations above were nonetheless performed in this context to demonstrate that IB defines collective modes of a system and establishes a cutoff scheme which, in general, differs from traditional notions of relevance, as represented by the Fourier basis and momentum cutoff. This idea could be crucial to understanding collective behavior in systems without clear notions of locality or organization. Such problems abound in, for example, the brain where long-distance connections between brain areas are common and important for computation while information is also spread across many areas and recombined for important, multi-modal tasks. The recurrent, highly interconnected, and still computationally efficient structure in the brain renders the simple notion of physical distance between cells rather limiting.

2.5.4 The relevance variable Y can have many physical interpretations

Gaussian IB begins with a choice of joint distribution $P(x, y)$. As we have discussed, this distribution gives a constrained parameterization of a cutoff scheme which is analogous to the one employed in Wetterich NPRG. In the last section, we showed that not all choices $P(x, y)$ lead to collective modes $V^T X$ which have a canonical interpretation such as Fourier modes. That discussion was carried out under the assumption that the relevance variable Y pertains to a source field with some disorder statistics. Generally speaking, this is only one way of constructing Y. Even within the constraint of $P(x, y)$ being jointly Gaussian, the physical interpretation of x and y can vary. Here we briefly discuss some of these alternative interpretations.

First, Y may represent the environment of a set of variables X. This scenario is analogous to the one presented by Koch-Janusz et al. (Koch-Janusz and Ringel [2018]). Consider a collection of spins on a lattice, and choose some enclosed region. Let X be the state of the spins in that region and let Y denote the state of those outside. In the case that these spins have Gaussian statistics, this is a valid starting point for GIB. With this setup, we expect that the most relevant collective modes would be relatively slowly varying in position. In fact, Gordon et al. recently formalized this idea for field theories not restricted to Gaussian statistics (Gordon et al. [2021]). They consider a "buffer" zone between X and Y whose size is taken to infinity. In this limit, the first collective variables encoded by IB at strong compression (low β in our notation) correspond to the operators with the smallest scaling dimensions, and hence the most relevant operators in the RG sense. Their approach is therefore promising for the analysis of systems with local interactions whose order parameter is not already given. More fundamentally, they have shown that Y and X can be chosen to enforce a traditional, "physical" definition of relevance.

Second, consider a stationary stochastic process with Gaussian statistics both in time and across variable index. We could choose X to represent the current state of the system while Y represents the future. Here, the most relevant modes are those projections of X which vary the slowest. In fact, if we suppose that time has been properly discretized, this interpretation of the GIB problem is equivalent to a certain class of slow feature analysis problems (Creutzig and Sprekeler [2008]).

Third, we can imagine another dynamical system in which variables X which are driven by a stochastic signal Y such that the joint distribution is Gaussian and stationary. Now, the features of X which are most relevant are no longer simply the slowest-varying components. The cutoff scheme we find will depend on the statistics which generate Y, the manner in which Y couples to X, the internal dynamics of X, and whether we take Y to be in the past, future, or present.

Together with the example from last section, in which Y fulfilled the role of a disordered source field, these examples span a number of physically interesting scenarios. Certainly, more are possible. Any valid interpretation will generally consist of a set of random variables $\{Z_i\}$ that obeys a Gaussian joint distribution, which is then partitioned into two or three disjoint sets. The first is $\{X_n\}$, the second is $\{Y_m\}$, and the third, which is optional, is a dummy set containing every Z_i which we don't care to include in the model. In the case that these sets aren't disjoint, it is possible to have X and Y become deterministically related which is an invalid starting point for GIB. Finally, we note that while this framework allows for some discussion of systems involving dynamics, it is poorly suited for application to general stochastic processes as the distribution $P(X, Y)$ must be stationary. This also means that the connections drawn here between GIB and NPRG are *not* meant to cover the more general, dynamical NPRG framework often seen in nonequilibrium statistical mechanics literature (Kopietz et al. [2010], Canet et al. [2011], Haga [2019]). However, given the importance of both IB and the dynamical NPRG to applications in nonequilibrium settings, we believe that a more general framework is in demand.

2.6 Conclusion

In this manuscript, we have examined structural similarities between the Gaussian information bottleneck problem and a class of RG techniques involving soft cutoffs. Our main result is to identify that the crucial connection between the two is a non-deterministic coarsening map. In NPRG, this map defines both the UV-regulated coarse-grained Hamiltonian of the Wilson-Polchinski picture, as well as the IR-regulated free energy used in the Wetterich approach. Therefore, one can rigorously connect IB to RG by requiring that this coarsening map solves a particular IB problem. In doing so, one parameterizes a space of soft cutoff schemes in terms of IB relevance variable statistics $P(x, y)$. Additionally, one can identify the structures in an IB problem which are analogous to UV- and IR-cutoffs in RG.

While we believe that this connection holds for more general IB problems, we limited our discussion to Gaussian statistics for two main reasons. First, NPRG coarsening maps are always Gaussian, since this leads to simpler flow equations with physical interpretations. Second, in order to be compatible with this first consideration, we studied only the GIB problem which has exactly known solutions that are Gaussian (Chechik et al. [2005]).

Another result was to show that the GIB coarsening map satisfies a semigroup property. In particular, we identify an explicit function $b(\beta)$ which multiplies under composition of coarsening maps in a manner analogous to the length scale in a traditional RG setting. Given that the typical role of semigroup structure in RG theory is the identification of anomalous exponents, it is not within the scope of this manuscript to assign a similar task to $b(\beta)$. More immediately, the presence of this structure within GIB raises the question of whether it may be present in IB schemes more generally. If so, would an iterative coarse-graining scheme consisting of repeated low-compression transformations be advantageous as an analysis technique?

By explicitly comparing the set of GIB solutions provided by Chechik et al. with a generic NPRG scheme, we identified the IR cutoff scheme present in GIB (2.27). A similar analysis can be carried out to identify the UV cutoff, but doing so involves a discussion about reparameterization which we felt would distract from the main points. Direct computations on a toy model showed that the IB regulator has some characteristics which are similar to the ubiquitous Litim regulator (Litim [2001]). An important generalization is that IB selects the collective mode basis according to which features of the system state X will be most informative about Y, whatever it is chosen to be. We gave a simple example in which this collective mode basis could not be interpreted as a Fourier basis. In general, this will be the case, though depending on how Y is defined, one may still arrive at collective modes which are essentially Fourier in nature. One bit of analysis we did not carry out is the connection of IB to the dynamical NPRG, though for non-equilibrium problems involving IB—such as the predictive coding problem—this may be a fruitful avenue for further work.

Next, we note that IB is generally extremely difficult to solve, so restricting an NPRG scheme to a family of exact IB solutions is completely unrealistic without significant advances in IB theory. One avenue of attack is to find better ways of solving IB. As outlined in appendix 2.8, a more general parametric Blahut-Arimoto scheme would be very powerful in this context since it could essentially replace the flow-equation description with a self-consistency scheme at each cutoff value. However, given that the exact Gaussian form we derive is complicated, this seems unlikely to work. A more realistic approach to practical IB-RG implementation is to relax the IB-optimality constraint. We suggest that even in a non-Gaussian setting, one could directly calculate the IB regulator (2.27) proposed here and use the NPRG flow equations in exactly the same way. While the resulting statistics would no longer be exactly IB-optimal, this procedure is no more difficult than any other NPRG implementation, and may produce qualitatively similar results to an exact IB solution.

We reiterate that not all IB problems will benefit from the RG connections presented here, and vice versa. Ideally, the problem in question involves a system with a large, but finite, number of degrees of freedom X statistically coupled to a similarly large number of random variables Y. Finiteness is required by IB, but because of the construction of the NPRG, this is not an issue. The flow is defined exactly even in the absence of a traditional rescaling step, which would be illegal in a finite system since it adds more modes. Biophysics systems, for example, may be particularly well-suited to IB-RG analysis, because Y can be chosen to have biological relevance, and the cutoff scheme will define and prioritize collective modes that are most informative about that function. Biological systems all have size and energy constraints that make the efficient compression of inputs from the external world critical for survival. Balancing that, and just as important for function, organisms also have clear preference for what is relevant in that external signal, namely which aspects can be used to drive behavior that confers a fitness benefit. The IB framework helps cast behavioral relevance as the prime mover in input compression, while the RG can help show how this kind of computation is achieved. Uniting these theories can provide a way to pull together normative notions of relevance with their mechanistic implementation.

2.7 Detailed derivation of GIB semigroup structure

A map (A, Σ_ξ) representing $\tilde{X} = AX + \xi$ solves the GIB problem if it satisfies:

$$[V^{-1}A^T\Sigma_\xi^{-1}AV^{-T}]_{ij} = \frac{\beta(1-\lambda_i)-1}{s_i\lambda_i}\Theta\left(\beta - \frac{1}{1-\lambda_i}\right)\delta_{ij}$$

for some β. To show that the composition of two GIB maps is IB-optimal, we explicitly compute the above expression for the map (A, Σ_ξ) arrived at by sequential coarsening. The individual maps are,

$$\begin{aligned} \tilde{X}_1 &= A_1X + \xi_1 &(2.48)\\ \tilde{X}_2 &= A_2\tilde{X}_1 + \xi_2\,. &(2.49) \end{aligned}$$

This construction gives

$$\begin{aligned} \tilde{X}_2 &= A_2A_1X + A_2\xi_1 + \xi_2 \\ &= AX + \xi &(2.50) \end{aligned}$$

So we have that, for $\Sigma_{\xi_1} = \Sigma_{\xi_2} = I$,

$$(A, \Sigma_\xi) = (A_2A_1, A_2A_2^T + I)\,. \tag{2.51}$$

In order to ensure that both A_1 and A_2 are diagonal, we project X into canonical correlation basis with the replacement $X \to V^TX$. Note that A_2 is actually automatically diagonal because the first compressed representation $\tilde{X}_1 = A_1 X + \xi_1$ is already in canonical correlation basis. After this transformation, the optimality condition (2.48) is simplified because the V^{-1} matrices have been absorbed into the definition of X. The new condition is:

$$\begin{aligned}[A^T \quad & \Sigma_\xi^{-1} A]_{ij} = \\ & \frac{\beta(1-\lambda_i)-1}{s_i\lambda_i}\Theta\left(\beta - \frac{1}{1-\lambda_i}\right)\delta_{ij}\end{aligned} \tag{2.52}$$

Now we explicitly compute A_1 and A_2. From (2.2) we have:

$$[A_1]_{ij} = \left[\frac{\beta_1(1-\lambda_i)-1}{s_i\lambda_i}\right]^{1/2}\Theta\left(\beta_1 - \frac{1}{1-\lambda_i}\right)\delta_{ij} \tag{2.53}$$

$$[A_2]_{ij} = \left[\frac{\beta_2(1-\lambda'_i)-1}{s'_i\lambda'_i}\right]^{1/2}\Theta\left(\beta_2 - \frac{1}{1-\lambda'_i}\right)\delta_{ij} \tag{2.54}$$

where

$$[\Sigma_{X|Y}]_{ij} = s_i\lambda_i\delta_{ij} \tag{2.55}$$

$$[\Sigma_X]_{ij} = s_i\delta_{ij} \tag{2.56}$$

$$[\Sigma_{\tilde{X}_1|Y}]_{ij} = s'_i\lambda'_i\delta_{ij} \tag{2.57}$$

$$[\Sigma_{\tilde{X}_1}]_{ij} = s'_i\delta_{ij} \tag{2.58}$$

The latter two equations must be re-expressed in terms of the original $X - Y$ statistics, represented by λ_i and s_i.

$$\begin{aligned}
\Sigma_{\tilde{X}_1|Y} &= A_1\Sigma_{X|Y}A_1^T + I \\
&\Rightarrow\ s_i'\lambda_i' = \lambda_i s_i[A_1]_{ii}^2 + 1 \qquad (2.59)\\
\Sigma_{\tilde{X}_1} &= A_1\Sigma_X A_1^T + I \\
&\Rightarrow\ s_i' = s_i[A_1]_{ii}^2 + 1 \qquad (2.60)
\end{aligned}$$

Solving for λ', we have:

$$\lambda_i' = \frac{\lambda_i s_i[A_1]_{ii}^2 + 1}{s_i[A_1]_{ii}^2 + 1} \qquad (2.61)$$

Now, directly evaluating A_1 and s_i, we get the following for λ' and s':

$$\begin{aligned}
\lambda_i' &= \min\left\{\frac{\beta_1}{\beta_1-1}\lambda_i,\, 1\right\} \qquad (2.62)\\
s_i' &= \frac{\beta_1(1-\lambda_i)-1}{\lambda_i}\Theta\left(\beta_1 - \frac{1}{1-\lambda_i}\right) + 1 \qquad (2.63)
\end{aligned}$$

Using these last two expressions, A_2 can be expressed directly in terms of s and λ. By direct substitution, we can now check whether the composite scheme (A, Σ_ξ) satisfies the GIB optimality condition (2.52):

$$\begin{aligned}
[A\Sigma_\xi^{-1}A^T]_{ij} &= [A_2A_1(A_2^2+I)^{-1}A_1A_2]_{ij} \\
&= \frac{(\beta_1(1-\lambda_i)-1)(\beta_2(1-\frac{\beta_1}{\beta_1-1}\lambda_i)-1)}{s_i\lambda_i(\beta_1(1-\lambda_i)+\beta_2(1-\frac{\beta_1}{\beta_1-1}\lambda_i)-1)}\Theta\left(\beta_2 - \frac{1}{1-\min\left\{1,\, \frac{\beta_1}{\beta_1-1}\lambda_i\right\}}\right)\delta_{ij} \\
&= \frac{(\beta_2\circ\beta_1)(1-\lambda_i)-1}{s_i\lambda_i}\Theta\left(\beta_2\circ\beta_1 - \frac{1}{1-\lambda_i}\right)\delta_{ij} \qquad (2.64)
\end{aligned}$$

where the binary operator $\circ$ is given by

$$\beta_2\circ\beta_1 = \frac{\beta_2\beta_1}{\beta_2+\beta_1-1}\,. \qquad (2.65)$$

By identifying $\beta_2 \circ \beta_1$ with a single value β, we find that the GIB optimality condition (2.48) is satisfied. It is important to note that this operator maps the space of valid β values $\mathbb{R} > 1$ to itself. That is,

$$\circ : \mathbb{R} > 1 \times \mathbb{R} > 1 \rightarrow \mathbb{R} > 1 \tag{2.66}$$

Which means that $\beta_2 \circ \beta_1$ really can be identified as a bottleneck parameter. Along with associativity, this means that $(\mathbb{R} > 1, \circ)$ is a semigroup representing sequential GIB coarsening.

2.8 Practicality and significance of the Blahut-Arimoto algorithm

2.8.1 Is the Blahut-Arimoto procedure a useful analytical tool in IB-optimal NPRG?

The apparent goal of Information Bottleneck is to identify the coarsening map $P_\beta(\tilde{x}|x)$ for some set of β values. This seems to align poorly with the problem statement and goals of NPRG, in which the coarsening map $P_k[\tilde{\chi}|\chi]$ is taken as the starting point and used to derive the flow equations. Is it really true that solving IB only gets us to the starting point of an RG scheme, after which we still need to "do the RG part?" In this section, we investigate one way to resolve this dissonance by noting that the quantities one would usually consider to be the results of the NPRG flow can be used to parameterize $P_\beta(\tilde{x}|x)$ itself. From this viewpoint, one may organize the computation around a set of self-consistent update equations instead of a set of flow equations.

The general IB problem can be solved, in principle, by iterating what is known as the Blahut-Arimoto procedure, which is borrowed from rate distortion theory in a more general context (Tishby et al. [2000]). This procedure relies on the fact that when $P_\beta(\tilde{x}|x)$ is IB optimal, it satisfies the following condition

$$P_\beta(\tilde{x}|x) = Z_\beta(x)^{-1} P_\beta(\tilde{x}) \exp\left(-\beta D_{\mathrm{KL}}[P(y|x)||P_\beta(y|\tilde{x})]\right) , \tag{2.67}$$

where everything on the RHS is to be considered a function of $P_\beta(\tilde{x}|x)$ through

$$\begin{aligned} P_\beta(\tilde{x}) &= \int \mathrm{d}x\, P_\beta(\tilde{x}|x) P(x) \,, & (2.68) \\ P_\beta(y|\tilde{x}) &= \frac{1}{P_\beta(\tilde{x})} \int \mathrm{d}x\, P(y|x) P_\beta(\tilde{x}|x) P(x) \,. & (2.69) \end{aligned}$$

The function $Z_\beta(x)$ normalizes $P_\beta(\tilde{x}|x)$ and therefore also depends on $P_\beta(\tilde{x}|x)$ through the above equations.

In brief, the BA procedure entails taking an estimate for $P_\beta(\tilde{x}|x)$, plugging it into the IB optimality criterion above, then iterating until satisfactory convergence. In this way, we say that $P_\beta(\tilde{x}|x)$ is self-determined. This procedure is practically very difficult—if not impossible—for distributions of multivariate continuous variables in general. However, in the case of GIB, we can parameterize the distributions then use Gaussian integral identities to update these parameters exactly. Chechik et al. (Chechik et al. [2005]) carry out this procedure in terms of the matrices A and Σ_ξ, used to define $\tilde{X} = AX + \xi$. We repeat this computation but instead parameterize the update equation using $\Sigma_{\tilde{X}}$, $\Sigma_{X|\tilde{X}}$, and $\Sigma_{X\tilde{X}}$. The first two of these represent objects of interest in the UV- and IR-regulated parts of the NPRG scheme, respectively. The third quantity, $\Sigma_{X\tilde{X}}$ carries information about how the IR degrees of freedom $\tilde{X}$ are coupled to the original, UV variables X. In a very condensed form, the BA update equations in this parameterization read:

$$\begin{aligned} \Sigma'_{X|\tilde{X}} &= [\Sigma_X^{-1} + \beta^2 B^T \Sigma'_{\tilde{X}|X} B]^{-1} \,, & (2.70) \\ \Sigma'_{\tilde{X}} &= [\Sigma'^{-1}_{\tilde{X}|X} - \beta^2 B \Sigma_{X|\tilde{X}} B^T]^{-1} \,, & (2.71) \\ \Sigma'_{X\tilde{X}} &= \beta \Sigma'_{X|\tilde{X}} B^T \Sigma'_{\tilde{X}} \,, & (2.72) \end{aligned}$$

where both B and $\Sigma'_{\tilde{X}|X}$ can be expressed in terms of β, $P(x,y)$ and the current estimate for the parameterization of $P(\tilde{x},x)$. The full expressions are complicated and given fully in appendix 2.8.2. Note that $\Sigma_{X|\tilde{X}}$ represents the IR-regulated flow; it is directly analogous to the effective propagator G_k in the Wetterich formalism. In other words, given that we are only looking at Gaussian statistics, the function $W_\beta(J)$ (or Γ_k) can be simply reconstructed from $\Sigma_{X|\tilde{X}}$. Next, $\Sigma_{\tilde{X}}$ represents the UV-regulated part, since the probability distribution describing $\tilde{X}$ can be reconstructed from it.

We reiterate that this self-consistent updating scheme comes from IB optimality, written in terms of objects we would usually calculate in NPRG. The idea of a self-consistent updating scheme which determines the IR-regulated statistics and UV-regulated dynamics simultaneously is interesting. In addition to essentially replacing the flow-equation description, it is very non-perturbative in nature. However, it seems wrong that imposing a constraint on $P(\tilde{x}|x)$ should make anything easier, especially given the fact that IB enforces a goal which is only sometimes aligned with the typical goals of RG analysis. A natural question, then, is whether IB has actually provided any new leverage. More precisely, if we really have given up the flow equation in favor of a self-consistency scheme, does this new scheme actually help to calculate the objects of interest as the flow equation usually would? If so, why would IB optimality be necessary?

In the case of general, i.e. non-Gaussian $P(x,y)$, the integration

$$\int \mathrm{d}x\, P(y|x)P(x|\tilde{x}) \tag{2.73}$$

can't be carried out directly. This is equivalent to the statement that at (and below) intermediate values of k in NPRG, $W_k[J]$ can't be directly computed from its integral representation. The whole point of Wilsonian RG is to get around this integration step by connecting W_k to $W_{k\to\infty} = 0$ by invoking a known flow equation. So, to answer our question, the IB update scheme may actually provide the same leverage, but only if *(1.)* we can represent the BA procedure parametrically, and *(2.)* the derivation of that parametric representation does not require the explicit marginalization of x to obtain $P(y|\tilde{x})$. The updates we present above for the fully Gaussian problem satisfy the first requirement, but fail the second since we explicitly carried out Gaussian integrals over x in the derivation. It is therefore unclear at this point whether some structure in IB could allow us to estimate $P(y|\tilde{x})$ parametrically, which seems to be a prerequisite for the utility of a more general IB-RG framework in which IB is exactly enforced. Finally, we note that these conditions are necessary, but not sufficient, since further integration steps may be required to complete the BA update, for example in computing $D_{KL}[P(y|x)||P(y|\tilde{x})]$ and going from an updated $P(x,\tilde{x})$ back to the moments of $P(x|\tilde{x})$.

2.8.2 Blahut-Arimoto update scheme for GIB in terms of NPRG objects

Eqs. (2.70) depict the Blahut-Arimoto updates for $\Sigma_{X|\tilde{X}}$, $\Sigma_{\tilde{X}}$, and $\Sigma_{X\tilde{X}}$ at a schematic level. Written as expectations, these matrices are:

$$\begin{aligned} [\Sigma_{X|\tilde{X}}]_{ab} &= \mathbb{E}_{X|\tilde{X}=\tilde{x}}\left\{(X-\mu_{X|\tilde{X}}(\tilde{x}))_a(X-\mu_{X|\tilde{X}}(\tilde{x}))_b\right\} && (2.74)\\ [\Sigma_{\tilde{X}}]_{ab} &= \mathbb{E}_{\tilde{X}}\left\{(\tilde{X}-\mu_{\tilde{X}})_a(\tilde{X}-\mu_{\tilde{X}})_b\right\} && (2.75)\\ [\Sigma_{X\tilde{X}}]_{ab} &= \mathbb{E}_{X,\tilde{X}}\left\{(X-\mu_X)_a(\tilde{X}-\mu_{\tilde{X}})_b\right\} && (2.76)\end{aligned}$$

As described in the main text, the BA procedure can be thought of as an iterative procedure wherein an estimate for $P(\tilde{x}|x)$, or equivalently $P(x,\tilde{x})$, is plugged into a known functional representing the consistency condition required by optimality. Schematically,

$$\begin{aligned} P'(x,\tilde{x}) &= \mathrm{BA}\left[P(x,\tilde{x})\right]\\ &= \frac{1}{Z_\beta(x)}P(x)P(\tilde{x})\exp\left[-\beta D_{\mathrm{KL}}\left[P(y|x)||P(y|\tilde{x})\right]\right] && (2.77)\end{aligned}$$

where D_{KL} is the Kullback-Leibler divergence, defined for two distributions P and Q of the same variable as

$$D_{\mathrm{KL}}\left[P||Q\right] = \int \mathrm{d}y\, P(y) \log \frac{P(y)}{Q(y)} \tag{2.78}$$

and the RHS can be seen as a functional of $P(x, \tilde{x})$ through the expressions:

$$\begin{aligned} P(\tilde{x}) &= \int \mathrm{d}x\, P(x, \tilde{x}) && (2.79) \\ P(y|\tilde{x}) &= \frac{1}{P(\tilde{x})} \int \mathrm{d}x\, P(y|x) P(x, \tilde{x}) && (2.80) \\ Z_\beta(x) &= \int \mathrm{d}\tilde{x}\, P(\tilde{x}) \exp\left[-\beta D_{\mathrm{KL}}\left[P(y|x)||P(y|\tilde{x})\right]\right] . && (2.81) \end{aligned}$$

In this appendix, we derive the equations (2.70) using the explicit form of the BA map presented above. The goal is to express "updates" for the matrices $\Sigma_{X|\tilde{X}}$, $\Sigma_{\tilde{X}}$, and $\Sigma_{X\tilde{X}}$ in terms of their current estimates. In general, quantities describing the updated joint distribution $P'(x, \tilde{x})$ will be primed. To begin, we evaluate $P(y|\tilde{x})$ using elementary properties of Gaussian variables. Next, we evaluate the divergence D_{KL}, and the partition function $Z_\beta(x)$. Finally, we combine these elements and read off the updated parameters. Suppose $a = b + c$, with $b \sim \mathcal{N}(\mu_b, \Sigma_b)$ and $c \sim \mathcal{N}(\mu_c, \Sigma_c)$. Then

$$a \sim \mathcal{N}(\mu_a, \Sigma_a) \quad \text{with} \quad \mu_a = \mu_b + \mu_c\,, \quad \Sigma_a = \Sigma_b + \Sigma_c \tag{2.82}$$

Therefore consider $y = Wx + z$ with $z \sim \mathcal{N}(0, \Sigma_{Y|X})$ and suppose that $P(x|\tilde{x}) = \mathcal{N}(\mu_{X|\tilde{X}}, \Sigma_{X|\tilde{X}})$. Then

$$\mu_{Y|\tilde{X}} = W\mu_{X|\tilde{X}} \tag{2.83}$$

$$\Sigma_{Y|\tilde{X}} = W\Sigma_{X|\tilde{X}}W^T + \Sigma_{Y|X} \tag{2.84}$$

Now, consider jointly Gaussian variables (a, b). Then

$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_b^{-1}(b - \mu_b) \tag{2.85}$$

Hence, assuming without loss of generality that $\mu_Y = 0$, $\mu_X = 0$, and $\mu_{\tilde{X}} = 0$,

$$\mu_{Y|X} = Wx \quad \Rightarrow \quad W = \Sigma_{YX}\Sigma_X^{-1} \tag{2.86}$$

and

$$\mu_{X|\tilde{X}} = \Sigma_{X\tilde{X}}\Sigma_{\tilde{X}}^{-1}\tilde{x} \tag{2.87}$$

so finally,

$$\begin{aligned} \Sigma_{Y\tilde{X}} &= \Sigma_{YX}\Sigma_X^{-1}\Sigma_{X\tilde{X}} & (2.88) \\ \Sigma_{Y|\tilde{X}} &= \Sigma_{YX}\Sigma_X^{-1}\Sigma_{X|\tilde{X}}\Sigma_X^{-1}\Sigma_{XY} + \Sigma_{Y|X} & (2.89) \end{aligned}$$

These matrices allow us to construct $P(y|\tilde{x})$ and thereby calculate D_{KL}. For Gaussian distributions, the KL divergence has a standard form. In this context, we care only about the terms which carry x and $\tilde{x}$ dependence. We have then,

$$\begin{aligned} D_{\mathrm{KL}}[P(y|x)||P(y|\tilde{x})] &\sim \frac{1}{2}(\mu_{Y|X} - \mu_{Y|\tilde{X}})^T\Sigma_{Y|\tilde{X}}^{-1}(\mu_{Y|X} - \mu_{Y|\tilde{X}}) \\ &= \frac{1}{2}\tilde{x}^T\Sigma_{\tilde{X}}^{-1}\Sigma_{\tilde{X}Y}\Sigma_{Y|\tilde{X}}^{-1}\Sigma_{Y\tilde{X}}\Sigma_{\tilde{X}}^{-1}\tilde{x} - x^T\Sigma_X^{-1}\Sigma_{XY}\Sigma_{Y|\tilde{X}}^{-1}\Sigma_{Y\tilde{X}}\Sigma_{\tilde{X}}^{-1}\tilde{x} + \left\{x^2\right\} \\ &= \frac{1}{2}\tilde{x}^T\Sigma_{\tilde{X}}^{-1}\Sigma_{\tilde{X}Y}\Sigma_{Y|\tilde{X}}^{-1}\Sigma_{Y\tilde{X}}\Sigma_{\tilde{X}}^{-1}\tilde{x} - x^TB^T\tilde{x} + \left\{x^2\right\}, \qquad (2.90) \end{aligned}$$

where $\sim$ denotes "up to addition of a constant." The matrix B describing the coupling between x and $\tilde{x}$ has been introduced for convenience. Note also that there is a pure-x term in this quantity, denoted $\left\{x^2\right\}$, which will cancel with the partition function $Z_\beta(x)$ that normalizes $P(\tilde{x}|x)$ in the BA map. In addition to this trivial x-dependence, $Z_\beta(x)$ also contributes a new x^2 term, which needs to be included.

$$\begin{aligned}
Z_\beta(x) &= \int \mathrm{d}\tilde{x}\, P(\tilde{x}) \exp\left(-\beta D_{\mathrm{KL}}\left[P(y|x)||P(y|\tilde{x})\right]\right) \\
&= \int \mathrm{d}\tilde{x}\, \exp\left(-\frac{1}{2}\tilde{x}^T\left[\Sigma_{\tilde{X}}^{-1} + \beta\Sigma_{\tilde{X}}^{-1}\Sigma_{\tilde{X}Y}\Sigma_{Y|\tilde{X}}^{-1}\Sigma_{Y\tilde{X}}\Sigma_{\tilde{X}}^{-1}\right]\tilde{x} + \beta x B^T\tilde{x} + \left\{x^2\right\} + \text{consts.}\right) \\
&= \int \mathrm{d}\tilde{x}\, \exp\left(-\frac{1}{2}\tilde{x}^T\Sigma'^{-1}_{\tilde{X}|X}\tilde{x} + \beta x^T B^T\tilde{x} + \left\{x^2\right\} + \text{consts.}\right) \\
&\sim \exp\left(\frac{1}{2}\beta^2 x^T B^T \Sigma'_{\tilde{X}|X} B x + \left\{x^2\right\}\right) \qquad (2.91)
\end{aligned}$$

Here we have introduced $\Sigma'_{X\tilde{|}X}$ to further clean up notation. Now it is straightforward to obtain $P'(x,\tilde{x})$ from the BA map by direct evaluation.

$$\begin{aligned}
P'(x,\tilde{x}) &= Z_\beta(x)^{-1}P(x)P(\tilde{x})\exp\left[-\beta D_{\mathrm{KL}}\left[P(y|x)||P(y|\tilde{x})\right]\right] \\
&= \exp\left(-\frac{1}{2}x^T\Sigma_X^{-1}x - \frac{1}{2}\beta^2 x^T B^T\Sigma'_{\tilde{X}|X}Bx - \frac{1}{2}\tilde{x}^T\Sigma'^{-1}_{\tilde{X}|X}\tilde{x} + \beta x^T B^T\tilde{x}\right) \quad (2.92)
\end{aligned}$$

Now, finally, all that remains is to complete the square and put the distribution in the form

$$P'(x,\tilde{x}) \sim \exp\left(-\frac{1}{2}(x-\mu'_{X|\tilde{X}})^T\Sigma'^{-1}_{X|\tilde{X}}(x-\mu'_{X|\tilde{X}}) - \frac{1}{2}\tilde{x}^T\Sigma'^{-1}_{\tilde{X}}\tilde{x}\right)\,, \qquad (2.93)$$

where

$$\mu_{X|\tilde{X}} = \Sigma'_{X\tilde{X}}\Sigma'^{-1}_{\tilde{X}}\tilde{x}\,. \qquad (2.94)$$

After completing the square, the updated matrices $\Sigma'_{X|\tilde{X}}$, $\Sigma'_{\tilde{X}}$, and $\Sigma'_{X\tilde{X}}$ can be read off,

$$\begin{aligned}
\Sigma'^{-1}_{X|\tilde{X}} &= \Sigma_X + \beta^2 B^T\Sigma'_{\tilde{X}|X}B & (2.95) \\
\Sigma'^{-1}_{\tilde{X}} &= \Sigma'^{-1}_{\tilde{X}|X} - \beta^2 B\Sigma'_{X|\tilde{X}}B^T & (2.96) \\
\Sigma'_{X\tilde{X}} &= \beta\Sigma'_{X|\tilde{X}}B^T\Sigma'_{\tilde{X}}\,. & (2.97)
\end{aligned}$$

These can be written entirely in terms of the old estimates through the substitutions,

$$\begin{aligned}
B^T &= \Sigma_X^{-1}\Sigma_{XY}\Sigma_{Y|\tilde{X}}^{-1}\Sigma_{Y\tilde{X}}\Sigma_{\tilde{X}}^{-1} && (2.98)\\
\Sigma'^{-1}_{\tilde{X}|X} &= \Sigma_{\tilde{X}}^{-1} + \beta\Sigma_{\tilde{X}}^{-1}\Sigma_{\tilde{X}Y}\Sigma_{Y|\tilde{X}}^{-1}\Sigma_{Y\tilde{X}}\Sigma_{\tilde{X}}^{-1} && (2.99)\\
\Sigma_{Y\tilde{X}} &= \Sigma_{YX}\Sigma_X^{-1}\Sigma_{X\tilde{X}} && (2.100)\\
\Sigma_{Y|\tilde{X}} &= \Sigma_{Y|X} + \Sigma_{YX}\Sigma_X^{-1}\Sigma_{X|\tilde{X}}\Sigma_X^{-1}\Sigma_{XY}\,. && (2.101)
\end{aligned}$$

Finally, we note that iteration of these equations does not guarantee the convergence of each matrix involved, since invertible linear transformations on the random variables are a symmetry of the objective function. The GIB-optimal solutions, which are described by the fixed points of this update scheme, are connected continuously by these symmetries. If one wishes to use these updates practically and ensure that all matrices converge to fixed values, it is necessary to break this reparameterization invariance by taking extra steps after each update. In the original GIB paper (Chechik et al. [2005]), the reparameterization-invariant quantities α_i are instead plotted over iteration of their BA scheme, because their convergence is guaranteed.

2.9 Selected computations for toy model

2.9.1 Canonical correlation Green's function

A central object in GIB is the canonical correlation matrix, $\Sigma_X^{-1}\Sigma_{X|Y}$. From this object, one obtains the eigenvector matrix V, which describes the linear transformation of X into its collective modes, and eigenvalues λ_i, which order these modes in terms of their information content about Y. In the toy model, we begin with physical definitions for the statistics in Eqs. (2.29) and (2.30). Then, by interpreting χ as the input variable X and the disorder h as the relevance variable Y, we ask what the structure of the resulting GIB-regularized

NPRG scheme looks like. Like any other GIB problem, we must first calculate the canonical correlation Green's function, $\Sigma_\chi^{-1}\Sigma_{\chi|h}$. Two Green's functions come directly from the definitions:

$$\Sigma_{\chi|h} = G_0\,, \qquad \Sigma_h = H \tag{2.102}$$

To find Σ_χ, we need $\Sigma_{\chi h}$, which we get through $\mu_{\chi|h}$:

$$\mu_{\chi|h} = \Sigma_{\chi h}\Sigma_h^{-1}h \tag{2.103}$$

Compute this mean by looking at the Hamiltonian for χ with frozen disorder h:

$$\begin{aligned} \mathcal{H}[\chi|h] &= \frac{1}{2}\chi^T\Sigma_{\chi|h}^{-1}\chi - h^T\chi \\ &= \frac{1}{2}(\chi-\mu_{\chi|h})^T\Sigma_{\chi|h}^{-1}(\chi-\mu_{\chi|h}) - \frac{1}{2}\mu_{\chi|h}^T\Sigma_{\chi|h}^{-1}\mu_{\chi|h} && (2.104) \\ \Rightarrow h &= \Sigma_{\chi|h}^{-1}\mu_{\chi|h} && (2.105) \end{aligned}$$

Hence we can identify $\Sigma_{\chi h}$:

$$\Sigma_{\chi h} = \Sigma_{\chi|h}\Sigma_h = G_0 H \tag{2.106}$$

Now, use the Schur complement formula to identify Σ_χ:

$$\begin{aligned} \Sigma_\chi &= \Sigma_{\chi|h} + \Sigma_{\chi h}\Sigma_h^{-1}\Sigma_{h\chi} \\ &= G_0 + G_0 H H^{-1} H G_0 \\ &= G_0 + G_0 H G_0 && (2.107) \end{aligned}$$

So finally, the canonical correlation Green's function is

$$\Sigma_\chi^{-1}\Sigma_{\chi|h} = (I + HG_0)^{-1} \tag{2.108}$$

2.9.2 Canonical correlation eigendecomposition calculations

Fourier collective basis

Once the canonical correlation Green's function is known, one calculates its eigenfunctions (or eigenvectors, in the usual, finite-dimensional case) and eigenvalues. In the main text, we consider three constructions of H which altogether yield two eigenbases: Fourier and non-Fourier. Let's first calculate the spectrum $\lambda(q)$ for case *2* in section 2.5.2, which also covers the analysis of case *1*.

$$\begin{aligned} H(x_1, x_2) &= \frac{1}{(2\pi)^d} \int \mathrm{d}^d q \, \eta(q) e^{iq\cdot(x_1 - x_2)} \\ &= [\mathcal{F}^\dagger \tilde{H} \mathcal{F}](x_1, x_2) \end{aligned} \tag{2.109}$$

Where $\tilde{H}$ represents a "diagonal" function, $\tilde{H}(q_1, q_2) = \eta(q_1)\delta^d(q_1 - q_2)$. The frozen disorder propagator G_0 is also diagonal in Fourier basis:

$$\begin{aligned} G_0(x_1, x_2) &= \delta^d(x_1 - x_2)[t - \nabla^2_{x_2}]^{-1} \\ &= \frac{1}{(2\pi)^d} \int \mathrm{d}^d q \, \frac{1}{t + q^2} e^{iq\cdot(x_1 - x_2)} \\ &= [\mathcal{F}^\dagger \tilde{G}_0 \mathcal{F}](x_1, x_2) \end{aligned} \tag{2.110}$$

We use $\tilde{G}_0(q)$ to represent both the function $(t + q^2)^{-1}$, and the diagonal kernel $(t + q^2)^{-1}\delta^d(q_1 - q_2)$ interchangeably, as needed. Using the expression for $\Sigma_\chi^{-1}\Sigma_{\chi|h}$ derived in the last section, we have:

$$\begin{aligned} \Sigma_\chi^{-1}\Sigma_{\chi|h} &= (I + HG_0)^{-1} \\ &= (I + \mathcal{F}^\dagger \tilde{H} \mathcal{F} \mathcal{F}^\dagger \tilde{G}_0 \mathcal{F})^{-1} \\ &= \mathcal{F}^\dagger (I + \tilde{H}\tilde{G}_0)^{-1} \mathcal{F} \\ &= V \Lambda V^{-1} \end{aligned} \tag{2.111}$$

Since $\mathcal{F}$ is unitary and both $\tilde{H}$ and $\tilde{G}_0$ are diagonal, we have:

$$V(x,q) = \mathcal{F}^\dagger(x,q) = \frac{1}{(2\pi)^{d/2}} e^{iq\cdot x}\,, \qquad \lambda(q) = \frac{1}{1+\eta(q)\tilde{G}_0(q)} \tag{2.112}$$

Non-Fourier collective basis

Next, we carry out the same computation for case *3*, in which H is not diagonal in Fourier basis, and so neither is the canonical correlation Green's function. Written formally, the disorder correlator is

$$H = \eta \mathcal{F}^\dagger \tilde{G}_0^{-1/2} \mathcal{F}_\alpha \tilde{G}_0 \mathcal{F}_\alpha^\dagger \tilde{G}_0^{-1/2} \mathcal{F}, \tag{2.113}$$

where $\mathcal{F}_\alpha$ is the d-dimensional fractional Fourier transform. The 1-dimensional version defined as:

$$\mathcal{F}_\alpha^{(1)}[f](u) = (2\pi i \sin\alpha)^{-1/2} \int_{\mathbb{R}} \mathrm{d}x\, f(x) \exp\left[-i\left(\csc(\alpha) ux - \frac{1}{2}\cot(\alpha)(x^2+u^2)\right)\right] \tag{2.114}$$

This transform is unitary, satisfies $\mathcal{F}_\alpha^{(1)} = \mathcal{F}_{-\alpha}^{(1)\dagger}$, and $\alpha = \pi/2$ gives the usual one-dimensional Fourier transform. To construct the d-dimensional version $\mathcal{F}_\alpha$, we simply take tensor products: $\mathcal{F}_\alpha = \mathcal{F}_\alpha^{(1)} \otimes \cdots \otimes \mathcal{F}_\alpha^{(1)}$ with d copies. Hence, $\mathcal{F}_\alpha$ has properties analogous to $\mathcal{F}_\alpha^{(1)}$, namely

$$\mathcal{F}_\alpha^\dagger = \mathcal{F}_{-\alpha} = \mathcal{F}_\alpha^{-1}, \qquad \text{and} \qquad \mathcal{F}_{\alpha=\pi/2} = \mathcal{F} \tag{2.115}$$

As in the Fourier case, we calculate the canonical correlation Green's function in terms of H and G_0, then write it in the form $V\Lambda V^{-1}$ with Λ diagonal.

$$\begin{aligned}
\Sigma_\chi^{-1}\Sigma_{\chi|h} &= (I + HG_0)^{-1} \\
&= (I + \eta\mathcal{F}^\dagger \tilde{G}_0^{-1/2} \mathcal{F}_\alpha \tilde{G}_0 \mathcal{F}_\alpha^\dagger \tilde{G}_0^{-1/2} \mathcal{F}\mathcal{F}^\dagger \tilde{G}_0 \mathcal{F})^{-1} \\
&= (I + \eta\mathcal{F}^\dagger \tilde{G}_0^{-1/2} \mathcal{F}_\alpha \tilde{G}_0 \mathcal{F}_\alpha^\dagger \tilde{G}_0^{1/2} \mathcal{F})^{-1} \\
&= (I + \eta(\mathcal{F}^\dagger \tilde{G}_0^{-1/2} \mathcal{F}_\alpha \tilde{G}_0^{1/2}) \tilde{G}_0 (\tilde{G}_0^{-1/2} \mathcal{F}_\alpha^\dagger \tilde{G}_0^{1/2} \mathcal{F}))^{-1} \\
&= (\tilde{G}_0^{-1/2} \mathcal{F}_\alpha^\dagger \tilde{G}_0^{1/2} \mathcal{F})^{-1} (I + \eta\tilde{G}_0)^{-1} (\mathcal{F}^\dagger \tilde{G}_0^{-1/2} \mathcal{F}_\alpha \tilde{G}_0^{1/2})^{-1} \\
&= V\Lambda V^{-1}
\end{aligned} \tag{2.116}$$

Hence, we arrive at the eigendecomposition:

$$V(x,u) = [\mathcal{F}^\dagger \tilde{G}_0^{-1/2} \mathcal{F}_\alpha \tilde{G}_0^{1/2}](x,u)\,, \qquad \lambda(u) = \frac{1}{1 + \eta\tilde{G}_0(u)} \tag{2.117}$$

In the main text, we refrain from writing $V^\dagger$ as a kernel $V^\dagger(u,x)$, because it is discontinuous and divergent. This is more evident when it is expressed in integral form,

$$\begin{aligned}
V^\dagger(u,x) &= [\tilde{G}_0^{1/2} \mathcal{F}_\alpha^\dagger \tilde{G}_0^{-1/2} \mathcal{F}](u,x) \\
&= (-2\pi i \sin\alpha)^{-d/2} (2\pi)^{-d/2} \sqrt{\frac{1}{1+u^2}} \\
&\quad \times \int \mathrm{d}^d q \, \sqrt{1+q^2} \exp\left[iq\cdot(u\csc\alpha - x) - \frac{i}{2}\cot(\alpha)(q^2+u^2) \right],
\end{aligned} \tag{2.118}$$

where, e.g., $u^2 = u\cdot u = u_1^2 + u_2^2 + ... + u_d^2$.

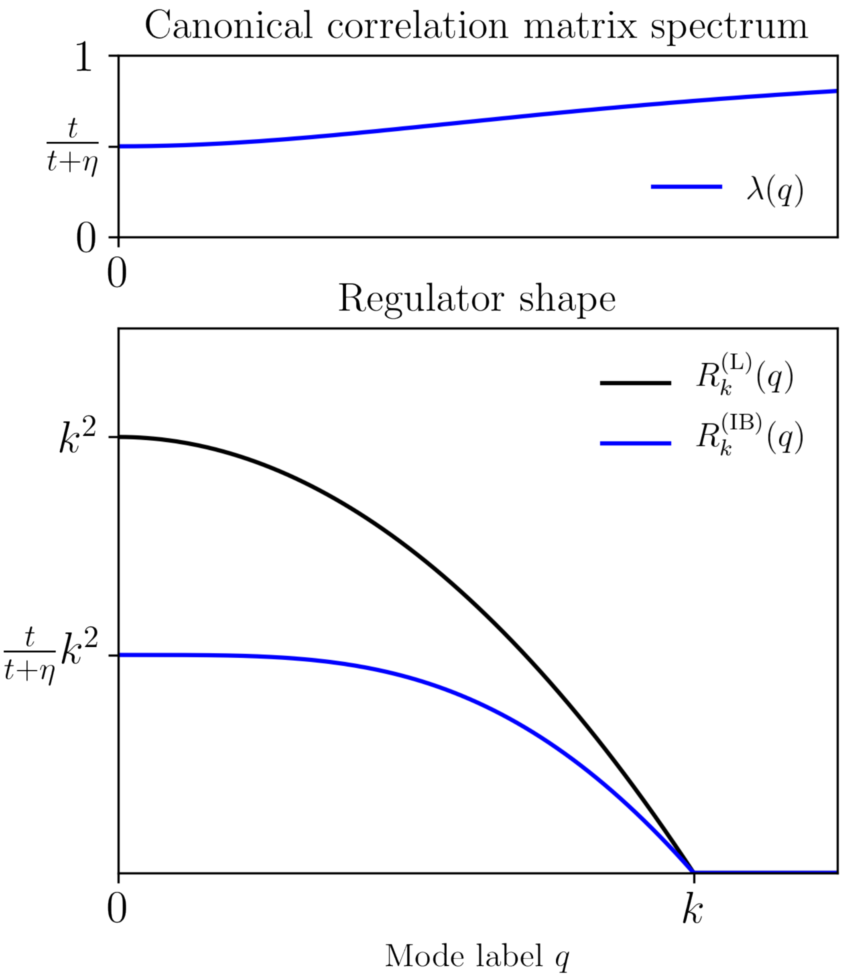

Figure 2.1: IR Regulators compared between Litim and IB schemes. The IB problem depicted here is from the toy model discussed in section 2.5.2 for the simple case where the collective modes selected by IB are Fourier and the disorder correlation has no dispersion (η is constant). *Top*: Eigenvalues of the canonical correlation matrix $\Sigma_X^{-1}\Sigma_{X|Y}$ as a function of label q, which may be interpreted as a wavevector magnitude. Modes with smaller eigenvalue can be thought to carry more information about Y. *Bottom*: Regulator values as a function of cutoff k and mode label q for the Litim scheme (2.28) and the IB scheme (black and blue, respectively).

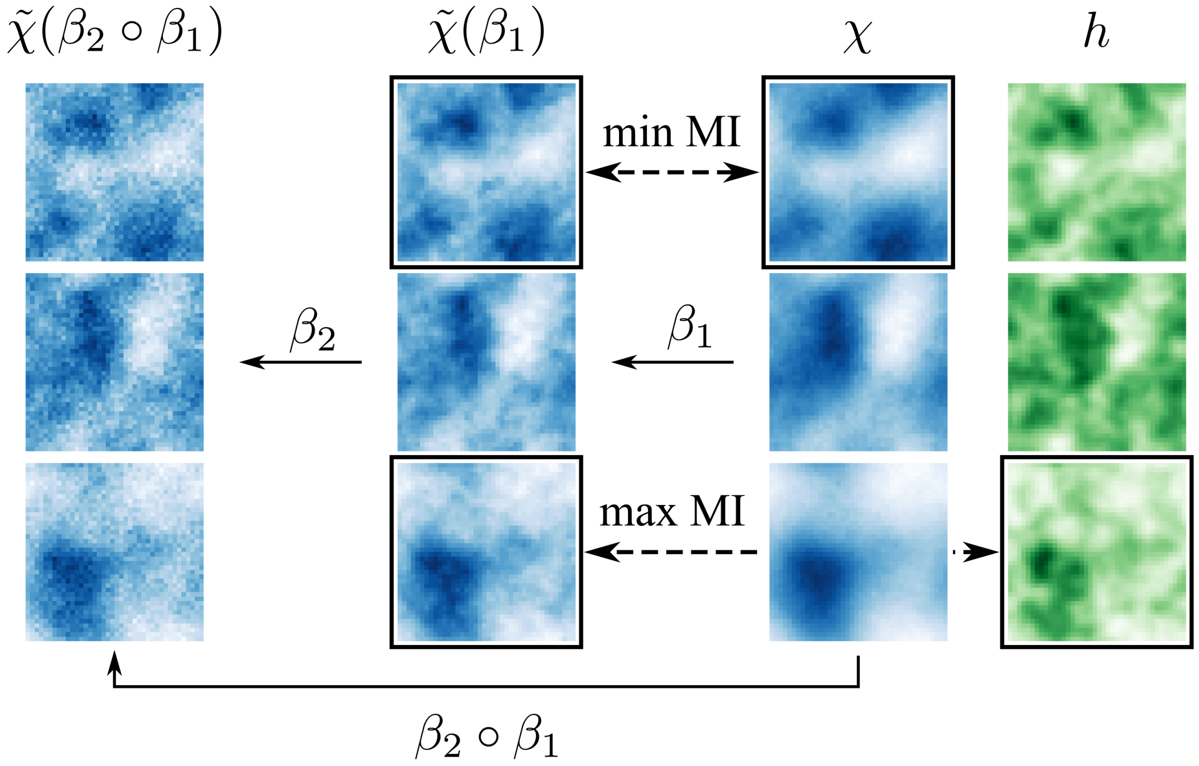

Figure 2.2: A depiction of the IB problem applied to a Gaussian field theory for $d = 2$, as described in Eq. (2.29). Each column represents a different random variable ($\tilde{X}$, X, or Y) in the IB problem, while each row depicts a sample drawn from the joint distribution between them. Using h as the relevance variable Y, the GIB-optimal coarsened field $\tilde{\chi}$ can be constructed through non-deterministic coarsening of χ, as depicted by the arrows. The Lagrange multiplier β_1 controls the trade-off between minimizing mutual information between $\tilde{\chi}(\beta_1)$ and χ while maximizing mutual information between $\tilde{\chi}(\beta_1)$ and h. According to the semigroup structure described in Sec. 2.3, this process can be repeated to generate $\tilde{\chi}(\beta_2 \circ \beta_1)$ through non-deterministic mapping from $\tilde{\chi}(\beta_1)$ with compression level β_2.

CHAPTER 3

MULTI-RELEVANCE: COEXISTING BUT DISTINCT NOTIONS OF SCALE IN LARGE SYSTEMS

AGK[1] and Stephanie E. Palmer[1,2]

1. *Department of Physics, The University of Chicago, Chicago IL 60637*

2. *Department of Organismal Biology and Anatomy, The University of Chicago, Chicago IL 60637*

Recently, renormalization group (RG) methods have been used in fields at the boundaries of traditional physics (Berman and Klinger [2022], Goldman et al. [2023], Erbin et al. [2022], Lahoche et al. [2021], Mehta and Schwab [2014], Koch-Janusz and Ringel [2018], Di Sante et al. [2022], Balog et al. [2022], Bény and Osborne [2014], Cotler and Rezchikov [2023a], Strandkvist et al. [2020], Pessoa and Caticha [2018], Jentsch and Lee [2023]). In the domain of theoretical biology, and in particular neuroscience, there have been attempts to apply RG methods to data (Bradde and Bialek [2017], Meshulam et al. [2019]) and models (Brinkman [2023], Tiberi et al. [2022]) in order to find simplifying structure through the language of many-body physics. Yet many biological systems are more complex than field theories, and when directly adopting RG techniques there is danger of oversimplifying, or of failing to define important collective variables correctly. This raises the question of how one should choose a cutoff scheme when applying RG in biological systems.

To frame the problem, let us consider the firing patterns of N neurons in a biological neural network. The joint distribution, $P(\sigma)$, is a probability distribution over $\sigma \in \{0,1\}^N$, and $-\log P(\sigma)$ is analogous to an energy for classical Ising spins. Models for this system have been fit to real data taken from the vertebrate retina, and the resulting energy landscapes have many local minima when constrained to a fixed total firing rate (Tkačik et al. [2014], Prentice et al. [2016]). Because each of these collective states is thought to encode a different set of features in the stimulus, we expect the definitions of important collective variables to change depending on the region of state space under consideration. The problem is that most RG approaches implement linear coarse-graining, meaning a single collective basis is applied globally. Using the non-perturbative RG (NPRG), these statements can be made precise for distributions over finitely many variables.

In the Wetterich construction of NPRG, the cutoff is enforced by a adding a regularization term to the Hamiltonian which limits fluctuations in IR variables. For continuous degrees of freedom ϕ, the Helmholtz energy is given by

$$W_k(J) = \log \int d\phi \, e^{-\mathcal{H}(\phi) - \frac{1}{2}\phi \cdot R_k \cdot \phi + J \cdot \phi}, \tag{3.1}$$

where, R_k adds mass to long-wavelength collective variables ($|q| < k$), suppressing their fluctuation contributions to the integral, but leaves short-wavelength modes ($|q| > k$) unaffected. As $k \to 0$, it is required that $R_k \to 0$, so that W_k becomes the exact connected generating function. In practice, the goal is to compute the effective average action

$$\Gamma_k(\varphi) = \max_J \{\varphi \cdot J - W_k(J)\} - \frac{1}{2} R_k \cdot \varphi^2 \,, \tag{3.2}$$

which obeys the exact flow equation (Wetterich [1993b])

$$\partial_k \Gamma_k(\varphi) = \frac{1}{2} \operatorname{Tr} \left\{ \partial_k R_k \cdot \left[R_k + \Gamma_k^{(2)}(\varphi) \right]^{-1} \right\} . \tag{3.3}$$

Here, $\varphi = \langle \phi \rangle$ is the flowing expectation value, defined as the variational derivative of $W_k(J)$ with respect to J. As $k \to 0$, Γ_k approaches the effective potential, denoted Γ, sometimes referred to as the Gibbs energy. NPRG methods are quite diverse in implementation and motivation (c.f. these reviews for broader background (Delamotte [2012], Kopietz et al. [2010], Bagnuls and Bervillier [2001], Dupuis et al. [2021])).

The fact that this RG procedure implements linear coarse-graining can be seen in the quadratic dependence of the regulator term on ϕ in Eq. (3.1). On the other hand, a coarse-graining scheme which captures biologically relevant features might use nonlinear coarse-graining (Kline and Palmer [2022]), but this would add ϕ dependence to the regulator R_k and invalidate the Wetterich Eq. (3.3). What is needed is some way to coarse-grain according to a number of different linear collective bases, without making the calculation intractable.

We provide a simple mechanism that allows for this kind of variability while leaving the essential flow equations unchanged. Some systems are most naturally modeled by a collection of coexisting, exactly independent RG flows, where each describes the collective physics according to a different notion of scale. We term systems with this property *multi-relevant.*

A system described by a state ϕ and Hamiltonian $\mathcal{H}(\phi)$ is multi-relevant when there exists a finite set of Hamiltonians $\{\mathcal{H}_z(\phi)\}$ which are polynomial in ϕ and which have finite partition functions such that

$$\exp\left(-\mathcal{H}(\phi)\right) = \sum_z \exp\left(-\mathcal{H}_z(\phi)\right) \, . \tag{3.4}$$

This expression, $\exp(-\mathcal{H})$, describes a mixture model whose components are normalizable and can be described by power-series expansions. Note that z can be interpreted as a label on a latent state that is not explicitly part of the system state. An immediate consequence of this condition is that the total partition function is simply a sum of the component partition functions [1]. This can be expressed in terms of the effective actions as $\exp \mathcal{L}[\Gamma](J) = \sum_z \exp \mathcal{L}[\Gamma_z](J)$, where $\mathcal{L}[F](J) = \max_\varphi \{\varphi \cdot J - F(\varphi)\}$ is the Legendre transform of F. By

[1] A similar mixture construction has been used to perform accurate modeling of complex systems in which metastable states have a significant impact on the overall thermodynamics (Liu et al. [2022, 2019]).

Γ_z we mean the effective potential obtained from $\mathcal{H}_z$ by running the RG flow (3.3) down to $k \to 0$. Because of this, each component model $\mathcal{H}_z$ can be renormalized independently subject to its own cutoff scheme $R_{z,k}$. The resulting effective potentials can be combined to give the full effective potential. In this sense, a multi-relevant system has multiple exactly independent RG flows.

We construct a toy model that is multi-relevant and has two mixture components, $\mathcal{H}_A$ and $\mathcal{H}_B$. Each is a finite, nonlocal analogue of scalar ϕ^4 theory, centered about a point s_z, with $\phi \in \mathbb{R}^N$ and N large. Explicitly,

$$\mathcal{H}_z(\phi + s_z) = \sum_{ab} \frac{1}{2}(K^z_{ab} + u^z_2 \delta_{ab})\phi_a \phi_b + \frac{u^z_4}{4!}\phi^2_a \phi^2_b \tag{3.5}$$

$$\mathcal{H}(\phi) = -\log\left[e^{-\mathcal{H}_A(\phi)} + e^{-\mathcal{H}_B(\phi)}\right]. \tag{3.6}$$

We refer to (3.6) as the "ϕ^4 mixture model". Here, the usual kinetic term has been replaced by a generic positive semi-definite matrix K^z_{ab} plus a "mass" part $u^z_2\delta_{ab}$, which is defined by requiring that the smallest eigenvalue of K^z is unity. The eigenvector matrix V^z of K^z is sampled from the Haar measure on $O(N)$, and s_z has a random orientation.

Although no notion of locality is present, we can still use RG, following (Bradde and Bialek [2017]). When analyzing the component energy labeled z, we take the collective degrees of freedom to be the eigenvectors of K^z. Assume that the K^z eigenvalues are well-described by a density $\rho_z(\lambda)$. In the collective basis, the scale-free part of the kinetic term has the form:

$$\sum_a \lambda_{z,a}\phi^2_a = \int d\lambda \rho_z(\lambda)\lambda\phi(\lambda)^2 \sim \int \frac{d^{D_z}q}{(2\pi)^{D_z/2}} q^2 |\phi(q)|^2$$

By assumption, ρ_z depends on λ as a power law, and we can identify the exponent with spatial dimensionality:

$$\rho_z(\lambda) \sim \lambda^{\alpha_z - 1} \Rightarrow \alpha_z = D_z/2$$

To be clear, we have not introduced anything like a spatial manifold in which these degrees of freedom are localized; this relation to dimension is only an analogy. The spectral density exponent α controls the scaling properties of couplings in the same way that dimensionality does in field theory. The UV scale Λ_z is the largest eigenvalue of K^z and scales as N^{1/α_z}.

For our analysis, we use the vertex expansion method with expansion points s_z. Although the NPRG formalism allows for all entries of the effective vertices to be tracked, we perform a low-dimensional parameterization in line with the parameters of the original Hamiltonian. Our cutoff scheme is the Litim regulator (Litim [2001]), without field-strength renormalization. For component z, this this takes the form

$$(V_z^T R_{z,k} V_z)_{ab} = \max\left\{k - \lambda_{z,a}, 0\right\} \delta_{ab} \,. \tag{3.7}$$

The ansatz for the flowing effective action contains two couplings $u^z_{2,k}$ and $u^z_{4,k}$ for each component z:

$$\begin{aligned}\Gamma^{(2)}_{z,k\,ab} &= K^z_{ab} + u^z_{2,k}\delta_{ab}\\ \Gamma^{(4)}_{z,k\,abcd} &= \frac{u^z_{4,k}}{3}(\delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc})\end{aligned}$$

The full RG analysis is given in Supplemental Material (SM) and is largely standard. For $\alpha_z \in (3/2, 2)$, (correspondingly $D_z \in (3, 4)$) the mass and interaction couplings $u_{z,2,k}$ and $u_{z,4,k}$ are relevant and describe the asymptotic properties. They obey the usual flow equations (Kopietz et al. [2010]) , up to differences in numerical factors due to neglecting subextensive diagrams. There are two fixed points, one at the origin and one at the WF fixed point. There is a $\mathbb{Z}_2$ symmetry-broken phase with two degenerate minima of the effective average action and a symmetric phase with a single minimum at s_z.

Because the ϕ^4 mixture model is multi-relevant, Γ can be obtained from the $k \to 0$ limits of $\Gamma_{A,k}$ and $\Gamma_{B,k}$. These individual analyses of A and B are simple, including only two relevant parameters each. When combined to compute the whole effective potential at the end, there are still only four relevant parameters which constrain measurements (averages). Yet, the functional dependence of these measurements on the four parameters is complicated due to the mixture construction. By contrast, it is clear that if we had not used the mixture construction and instead computed the vertex expansion flow for the whole Γ_k, our final answer would look a lot simpler, namely it would be some polynomial in the fields. This simpler representation is wrong in some regimes, since it cannot capture the essentially non-polynomial dependence of $\Gamma(\phi)$ on ϕ, due to the mixture construction. Despite this, there are regimes where the standard approach is justified.

Whether standard vertex expansion succeeds or fails is essentially determined by the parameters s_A and s_B. These act as displacements in state space, centering one ϕ^4 model at s_A and the other at s_B. Let $s = s_A - s_B$ be the separation vector. In Fig. 3.1, we show the separation dependence of the probability density of ϕ projected along the inter-basin axis $\hat{s}$. There is a transition as $|s|$ crosses some critical threshold s_c, which we calculate in the SM. To be concrete, we define this transition as the largest separation at which $P(\hat{s} \cdot \phi)$ has a local maximum at $\hat{s} \cdot \phi = 0$. This approximates the separation beyond which the A and B component densities can be separated by a hyperplane.

Consider first the well-separated phase, that is $|s| > s_c$. The total energy has a saddle point near s_A, and largely the local structure looks like $\mathcal{H}_A$. This is because the coefficient data due to the presence of $\mathcal{H}_B$ are suppressed exponentially with respect to a power of the separation $|s|$. For states ϕ near s_A,

$$\mathcal{H}(\phi) = \mathcal{H}_A(\phi) - e^{-\Delta\mathcal{H}_{BA}(\phi)} + O\left(e^{-2\Delta\mathcal{H}_{BA}(\phi)}\right),$$

where $\Delta\mathcal{H}_{AB} = \mathcal{H}_B - \mathcal{H}_A$. Then, using the regulator $R_{A,k}$ in (3.7), one can compute Γ_k in the neighborhood of s_A, using the above as the initial conditions of the flow. For sufficiently large separations $|s| > s_c$, the perturbations to $\mathcal{H}_A$ due to $\mathcal{H}_B$ do not change the fact that

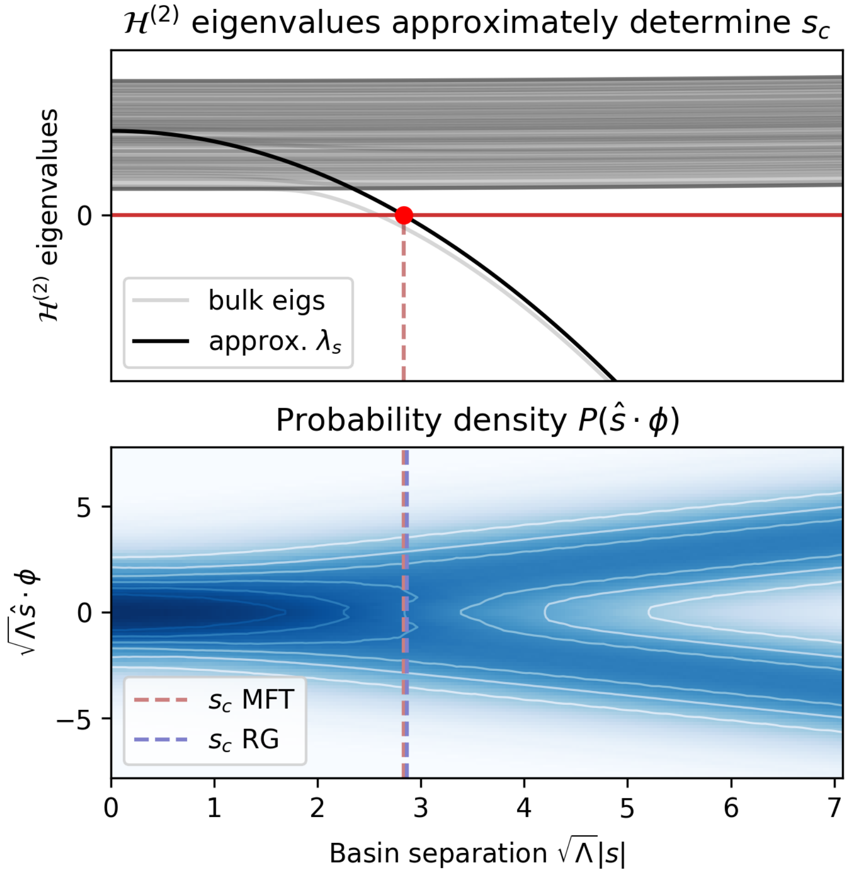

Figure 3.1: Demonstration of the coalescence transition. **Top**: spectrum of the energy Hessian at the saddle point separating basins A and B. Model parameters are $N = 50$, $\alpha_A = 3/2$, $\Lambda_A = N^{1/\alpha_A}$, $u_2^A = -\Lambda_A/10$, $u_4^A = 0.58\Lambda_A^{s_4}$. All A and B parameters are equal. The gray lines depict the true eigenvalues (neglecting $O(|s|^4)$ effects) while the black line represents $\lambda_s = \hat{s}\cdot\mathcal{H}^{(2)}\cdot\hat{s}$. The separation $|s|$ at which λ_s crosses 0 is the mean-field estimate. **Bottom**: probability density of the $\phi\cdot\hat{s}$ as a function of separation. Picking a value on the x-axis selects a single distribution with a specified separation. The red dashed line indicates the mean-field estimate of the critical separation, while the blue dashed line gives the RG estimate (see SM for expressions).

the calculation of Γ_k recovers a good approximation of $\Gamma_{A,k}$. This follows from the fact that when the basins are well-separated, a series expansion of $\Gamma(\phi)$ will approximate that of $\Gamma_z(\phi)$ for ϕ in a suitable neighborhood of s_z. In turn, this occurs because the Helmholtz free energy for one component, $W_z(J)$, dominates the mixture family for J near some J^*, making $\Gamma(\phi) \approx \Gamma_z(\phi)$ for ϕ in the neighborhood of $\phi(J^*)$. We discuss this further in the SM.

The vertex expansion approach can therefore, in the well-separated case, be sufficient to analyze a multi-relevant energy function $\mathcal{H}$. However, to get a full description of Γ, one needs to perform several RG flows, one at each s_z. In general, each requires a different regulator scheme, and will give different asymptotic properties than the others. In practice, this is equivalent to using the mixture construction. It is interesting to note that this analysis demonstrates how multi-relevant models can have differing low-energy physics near different points in state space. If there were, additionally, some notion of dynamics that led to effective ergodicity breaking, one could consider the system getting trapped in basin A or B for long periods of time, and in each case the other basin should have little influence.

In the coalesced phase, the lowest-order approximation to the vertex expansion approach fails. In this case, the separation $|s|$ is smaller than s_c. For simplicity, we consider the case $|s| = 0$. Expanding (3.6) about the point $s_A = s_B = 0$ yields, at the quadratic level,

$$\mathcal{H}^{(2)}_{ab} = K_{ab} + \frac{1}{2}(u_2^A + u_2^B)\delta_{ab}\,,$$

where $K = (K^A + K^B)/2$. Because K^A and K^B are diagonal in different, randomly chosen bases, their sum $2K$ defines a new notion of scale that does not completely agree with either system A or B. As before, we define scale using the eigenvalues of K, with the collective degrees of freedom (analogous to Fourier modes) defined by the eigenvectors. In our construction, K acquires a mass gap K_0, and after subtracting this off, the remaining eigenvalues will have some asymptotically scale-free behavior. Therefore let $\{\lambda_a + K_0\}_{a=1}^N$ denote the eigenvalues of K, and construct the density $\rho(\lambda)$ of these eigenvalues as we did in the finite ϕ^4 case. The units associated to these eigenvalues are defined through the scaling properties of this density. For example, since $\rho(\lambda) \sim \lambda^{\alpha-1}$, an extensive quantity has dimension α.

Expanding (3.6) to higher orders in ϕ, all of the terms come in the form of symmetrized outer products involving K^A, K^B, and δ, with δ the identity matrix. For example, the $\mathcal{H}^{(4)}$ has a term like $(K^A)_{(ab}(K^B)_{cd)}$, which we denote $(K^A \circ K^B)_{abcd}$. In the eigenbasis of K, the matrices K^A and K^B can be approximated as diagonal, but the diagonal elements pick up a mass gap just as K itself does. By defining $\tilde{K}^A = K^A - K_0^A\delta$ as the scale-free part of the K^A diagonal elements (and repeating for K^B), we can write the expansion in terms of approximately scale-free objects. After computing scaling dimensions, the relevant and marginal couplings are captured by the ansatz:

$$\Gamma^{(2)}_{k\,ab} = (v_k^A \tilde{K}^A + v_k^B \tilde{K}^B + v_k^\delta \delta)_{ab} \tag{3.8}$$

$$\Gamma^{(4)}_{k\,abcd} = v_k^{\delta\delta}(\delta \circ \delta)_{abcd} \tag{3.9}$$

$$\Gamma^{(6)}_{k\,a\ldots f} = v_k^{\delta\delta\delta}(\delta \circ \delta \circ \delta)_{a\ldots f} \tag{3.10}$$

The spectral density exponent α determined by this new quadratic coupling K is smaller than α_A and α_B when we take them to be equal. Because of this, the sixtic interaction becomes relevant and must be included. Yet, for $\alpha_A = \alpha_B$ in the range $(3/2, 2)$, no other terms appearing in the initial conditions generated by the expansion of $\mathcal{H}$ have positive scaling dimensions at the Gaussian fixed point. The couplings we include provide three unique constraints on the original model parameters, and the marginal coupling initial conditions provide no additional constraints.

To compute the flow equations, we employ the Litim regulator (without field-strength renormalization) as before. Keeping only extensive terms yields

$$
\begin{aligned}
\dot{g}_l^{\delta} &= g_l^{\delta} + \frac{1}{6}\frac{g_l^{\delta\delta}}{(1+g_l^{\delta})^2} \\
\dot{g}_l^{\delta\delta} &= (2-\alpha)g_l^{\delta\delta} - \frac{1}{3}\frac{(g_l^{\delta\delta})^2}{(1+g_l^{\delta})^3} + \frac{1}{10}\frac{g^{\delta\delta\delta}}{(1+g^{\delta})^2} \\
\dot{g}_l^{\delta\delta\delta} &= (3-2\alpha)g_l^{\delta\delta\delta} + \frac{5}{3}\frac{(g_l^{\delta\delta})^3}{(1+g_l^{\delta})^4} - \frac{g_l^{\delta\delta}\, g_l^{\delta\delta\delta}}{(1+g_l^{\delta})^3}
\end{aligned}
$$

where $l = -\log(k/\Lambda)$ and g are the non-dimensionalized couplings. These flow equations yield three finite fixed points: The Gaussian point at the origin, the WF point, and one other point that appears for $\alpha \leq 3/2$. In effect, a standard application of the vertex expansion approach predicts that the IR properties of the mixture model in the coalesced phase are the same as some kind of nonlocal ϕ^6 model near and below 3 dimensions (but still above $D = 8/3$ where ϕ^8 becomes relevant).

The crucial result that we come to now is that the vertex expansion approach on the full model $\mathcal{H}$ at $s_A = s_B$ led to oversimplification. A primary indication of this is there are only three relevant parameters, while according to the separate or mixture analysis there should be four. In systems with increasing N, the net effect of irrelevant couplings on the large-scale components of average quantities decreases, and the possibility of constraining their initial values using data disappears. By contrast, we know that all four parameters, $u_{z,2}$ and $u_{z,4}$, can affect predictions if the mixture construction is used, regardless of the basin separation $|s|$. This disagreement arises at the choice of truncation scheme (3.16), (3.17), and (3.18). Due to the non-polynomial structure of the microscopic model $\mathcal{H}$, an infinite number of irrelevant couplings with large initial values are thrown away. However, these have a significant effect on the flow of relevant and marginal couplings in the UV and, for accurate predictions, must be included. In essence, the single-flow approach oversimplifies because its notion of scale is poorly chosen and no good polynomial truncation scheme is available.

An important consequence of this calculation is that when principal component analysis (PCA) is applied to data drawn from a multi-relevant distribution, it can also lead to oversimplification. In data from our model, one finds approximately power-law distributed covariance eigenvalues and increasingly non-Gaussian behavior in the IR, suggesting a nontrivial RG flow. However, by performing a single RG flow starting at $s_A = s_B$ (or at the saddle point between the basins), we chose a collective basis and cutoff scheme which emulates PCA (Bradde and Bialek [2017], Lahoche et al. [2021]). Under coarse-graining with respect to this notion of scale, the RG analysis fails to effectively constrain the statistics of large-scale degrees of freedom using the original parameters. In contrast, the non-Gaussian statistics observed for large principal components is accounted for by the mixture model construction. Like the single-expansion point RG analysis, PCA can lead to oversimplification because it mixes up the collective bases defined by K^A and K^B, and after coarse-graining their effects cannot be disentangled.

In this letter, we introduced a way to formalize multi-relevance. This property endows a system with multiple coexisting and exactly independent RG flows, each with its own potentially distinct notion of scale. The existence of latent categorical variables is a simple mechanism which leads directly to multi-relevance, essentially by definition. We expect these results to be an important step towards implementing RG techniques in theoretical biology and complex systems more broadly. In these new domains, RG could offer paths to simplification in the space of models and thereby aid in the search for organizing principles. Our analysis reveals that multi-relevance does not simplify under the RG flow, and in this sense represents a degree of inherent complexity which must be accounted for when building and classifying theories. Whereas PCA fails to correctly coarse-grain multi-relevant systems, machine learning methods have demonstrated promise in discovering latent representations together with rich, nonlinear encoding and decoding schemes (Hinton and Salakhutdinov

[2006], Berman et al. [2016], Ding et al. [2019], Ahamed et al. [2021], Kalinin et al. [2021], Ziegler et al. [2023]). Through multi-relevance, these capabilities are brought closer to many-body formalism and new theories of collective computation.

3.1 NPRG on the finite ϕ^4 model

In this section, we work through the NPRG analysis of a finite ϕ^4 model as defined in the main text, Equation 3.5:

$$\mathcal{H}(\phi) = \frac{1}{2}\sum_{ab}(K_{ab} + u_2\delta_{ab})\phi_a\phi_b + \frac{u_4}{4!}\phi_a^2\phi_b^2$$

In Wetterich's formulation of the NPRG, the flow is defined in terms a slightly modified version of the 1PI generating functional, or Gibbs energy, known as the effective average action:

$$\Gamma_k(\varphi) = \max_J\left\{\varphi\cdot J - \log\int d\phi \exp\left[-\mathcal{H}(\phi) - \frac{1}{2}R_{k\,ab}\,\phi_a\phi_b + J\cdot\phi\right]\right\} - \frac{1}{2}R_{k\,ab}\,\varphi_a\varphi_b \quad (3.11)$$

The matrix R is known as the regulator. It ensures that the the largest eigenvalue of the propagator is no larger than $1/k$ (or $1/k^2$, traditionally, though we choose an alternative parameterization for convenience). The full propagator at scale k can be found using the effective action and the regulator:

$$\left[G_k^{-1}\right]_{ab} = \frac{\partial}{\partial\varphi_a}\frac{\partial}{\partial\varphi_b}\Gamma_k(\varphi) + R_{k\,ab}$$

It is important to note that unlike a hard cutoff scheme, the whole configurational integral over states ϕ in (3.11) is carried out, including fluctuations in IR modes. The regulator takes the role as the cutoff, and given that it satisfies the necessary boundary conditions, the resulting RG flow is well-defined for a finite number of degrees of freedom. This is not the case with a hard cutoff scheme, wherein one needs to rescale the total number of degrees of freedom in order to map the parameters of the coarse-grained Hamiltonian back in to the original space of couplings. This does not mean that there is not a rescaling step eventually, rather this new rescaling step happens after the flow has already been defined.

We use the vertex expansion, meaning the flow equations for vertices $\Gamma^{(n)}_{a_1,\ldots,a_n}$ are found by differentiating the Wetterich flow equation directly and keeping track of expansion coefficients:

$$\begin{aligned}
\dot{\Gamma}_k(\varphi) &= \frac{1}{2}\operatorname{Tr}\left[\dot{R}_k G_k(\varphi)\right] \\
\dot{\Gamma}^{(2)}_{ab} &= -\frac{1}{2}\operatorname{Tr}\left[\Gamma^{(4)}_{ab} G_k \dot{R}_k G_k\right] + \operatorname{Tr}\left[\Gamma^{(3)}_{a} G_k \Gamma^{(3)}_{b} G_k \dot{R}_k G_k\right] \\
\dot{\Gamma}^{(3)}_{abc} &= \ldots \\
\dot{\Gamma}^{(4)}_{abcd} &= -\frac{1}{2}\operatorname{Tr}\left[\Gamma^{(6)}_{abcd} G_k \dot{R}_k G_k\right] + 3\operatorname{Tr}\left[\Gamma^{(4)}_{ab} G_k \Gamma^{(4)}_{cd} G_k \dot{R}_k G_k\right]_{\text{(symmetrized)}} \\
&\qquad + \text{terms involving } \Gamma^{(3)}
\end{aligned}$$

From here on, we implicitly symmetrize where necessary. To evaluate these flow equations we need to truncate the hierarchy at some highest power n and parameterize the vertices with an ansatz. It is easiest to start with a good parameterization and verify that higher-order vertices will not be relevant. Working from the high-temperature phase, all odd vertices such as $\Gamma^{(3)}$ are set to zero. The initial condition on Γ_k is $\Gamma_{k\to\infty} = \mathcal{H}$. We approximate this by setting $\Gamma_\Lambda = \mathcal{H}$. Truncating at order $n = 4$, the effective action can be written in the same form as $\mathcal{H}$:

$$\begin{aligned}
\Gamma^{(2)}_{k\,ab} &= (\lambda_a + u_{2\,k})\delta_{ab} \\
\Gamma^{(4)}_{k\,abcd} &= \frac{u_{4\,k}}{3}\left(\delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}\right) \\
u_{2\,\Lambda} &= u_2 \\
u_{4\,\Lambda} &= u_4
\end{aligned}$$

From the truncation and $\phi \to -\phi$ symmetry we therefore have the following closed set of finitely many flow equations:

$$\begin{aligned}
\dot{\Gamma}^{(2)}_{ab} &= -\frac{1}{2}\operatorname{Tr}\left[\Gamma^{(4)}_{ab} G_k \dot{R}_k G_k\right] \\
\dot{\Gamma}^{(4)}_{abcd} &= 3\operatorname{Tr}\left[\Gamma^{(4)}_{ab} G_k \Gamma^{(4)}_{cd} G_k \dot{R}_k G_k\right]
\end{aligned}$$

To obtain a low-dimensional description of the flow, we re-express these flow equations in terms of the two parameters $u_{2\,k}$ and $u_{4\,k}$ of our ansatz. For example, we first insert $\Gamma^{(4)}$ into the RHS of the expression for $\dot{\Gamma}^{(2)}$ above.

$$\begin{aligned}
\partial_k \Gamma^{(2)}_{k\,ab} &= -\frac{1}{2}\frac{u_{4\,k}}{3}\sum_{cd}\left(\delta_{ab}\delta_{cd} + \delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}\right)\left(G_k \dot{R}_k G_k\right)_{cd} \\
&= -\frac{u_{4\,k}}{6}\left(\operatorname{Tr}\left(G_k \dot{R}_k G_k\right)\delta_{ab} + 2\left(G_k \dot{R}_k G_k\right)_{ab}\right) .
\end{aligned}$$

The first and second terms on the second line differ by a factor proportional to the number of degrees of freedom in the system, as we will show shortly. For this reason, we refer to the first term as extensive, while the second is sub-extensive. To actually calculate these terms, we need to introduce a regulator and get an expression for the propagator. In this context, the Litim regulator without field-strength renormalization can be written in the

$\Gamma^{(2)}$ eigenbasis as

$$R_{k\,ab} = \max\left\{k - \lambda_a, 0\right\} \delta_{ab}$$

This gives us

$$\begin{aligned}
\dot{R}_{k\,ab} &= 0\,; \quad a > n_k \\
&= \delta_{ab}\,; \quad a < n_k \\
G_{k\,ab} &= \left(\Gamma_k^{(2)}\right)^{-1}_{ab}\,; \quad a > n_k \\
&= \frac{1}{k + u_{2\,k}} \delta_{ab}\,; \quad a > n_k
\end{aligned}$$

and so

$$\dot{\Gamma}^{(2)}_{k\,ab} = -\frac{1}{2}\frac{u_{4\,k}}{3}\left(\frac{n_k}{(k+u_{2\,k})^2}\delta_{ab} + \frac{2}{(k+u_{2\,k})^2}\dot{R}_{k\,ab}\right)$$

By n_k we mean the number of eigenvalues λ_a below the cutoff k. We will mainly be interested in the intermediate part of the flow, for which n_k is a number much larger than unity, even though it may be much less than N. This is the regime in which the scale-free parts of these flow equations are accurate and look similar to traditional results in infinite systems. Dropping the sub-extensive term and inserting our ansatz for $\Gamma^{(2)}$ into the RHS of the above equation, we find:

$$\dot{u}_{2\,k} = -\frac{1}{6}\frac{n_k u_{4\,k}}{(k+u_{2\,k})^2}$$

The computation for $\partial_k u_{4\,k}$ proceeds similarly. Inserting the effective action ansatz into the

RHS of the truncated flow equations yields

$$\begin{aligned}
\dot{\Gamma}^{(4)}_{k\,abcd} &= 3\,\mathrm{Tr}\left[\Gamma^{(4)}_{ab} G_k \Gamma^{(4)}_{cd} G_k \dot{R}_k G_k\right] \\
&= 3u_{4\,k}^2 \sum_{a'b'c'd'} \frac{1}{3}(\delta_{ab}\delta_{a'b'} + \delta_{aa'}\delta_{bb'} + \delta_{ab'}\delta_{ba'})\times \\
&\qquad\qquad \frac{1}{3}(\delta_{cd}\delta_{c'd'} + \delta_{cc'}\delta_{dd'} + \delta_{cb'}\delta_{dc'})(G_k \dot{R}_k G_k)_{a'c'}\, G_{k\,b'd'} \\
&= \frac{1}{3} u_{4\,k}^2 \left(\delta_{ab}\delta_{cd}\,\mathrm{Tr}\left(G_k^3 \dot{R}_k\right) + \text{sub-extensive terms}\right) \\
&\to \frac{1}{3}\frac{n_k u_{4\,k}^2}{(k+u_{2\,k})^3}\delta_{ab}\delta_{cd}
\end{aligned}$$

where the RHS is symmetrized across $abcd$. Substituting on the LHS, we have the scale-free part

$$\dot{u}_{4\,k} = \frac{1}{3}\frac{n_k}{N}\frac{u_{4\,k}^2}{(k+u_{2\,k})^3}$$

In this case, we can assign dimensions to the couplings simply by inspecting the flow equations. The key is to note that, for $[k] = 1$ by definition,

$$\alpha = \partial_{\log\lambda} \log \rho(\lambda) + 1 \quad \Rightarrow \quad [n_k] = \alpha\,.$$

A heuristic way to see this is to give λ_n a simple expression consistent with $\rho(\lambda) \sim \lambda^{\alpha-1}$. This is satisfied by

$$\lambda_n \sim \Lambda\left(\frac{n}{N}\right)^{1/\alpha}\,.$$

Proceeding, we can straightforwardly identify:

$$s_2 = [u_2] = 1 \quad s_4 = [u_4] = 2-\alpha$$

Finally, we define rescaled dimensionless couplings $g_2 = k^{-s_2}u_2$ and $g_4 = k^{-s_4}u_4$, and let

$l = -\log(k/\Lambda)$. The dimensionless flow equations are given by:

$$\partial_l g_{2\,l} = g_{2\,l} + \frac{1}{6}\frac{g_{4\,l}}{(1+g_{2\,l})^2}$$
$$\partial_l g_{4\,l} = (2-\alpha) g_{4\,l} - \frac{1}{3}\frac{g_{4\,l}^2}{(1+g_{2\,l})^3}$$

These have the standard properties describing the ϕ^4 system, up to some differences in coefficients, and with anomalous scaling neglected. Save for the infinite fixed points, when $\alpha > 2$ the only physical ($g_4 \geq 0$) fixed point is at $(0, 0)$. For $\alpha < 2$, the Wilson-Fisher fixed point appears at

$$g_2^* = \frac{2-\alpha}{4-\alpha} \qquad g_4^* = \frac{24(2-\alpha)}{(4-\alpha)^2}\,.$$

3.2 Approximate behavior of Γ in well-separated phase

In the main text, we discussed that for a multi-relevant Hamiltonian $\mathcal{H}$ with component Hamiltonians $\mathcal{H}_z$, the full effective potential Γ satisfies the exact relation

$$\exp \mathcal{L}[\Gamma](J) = \exp W(J) = \sum_z \exp W_z(J) = \sum_z \exp \mathcal{L}[\Gamma_z](J)$$

In this section, we argue that in the well-separated phase, Γ behaves like Γ_z for ϕ in the neighborhood of s_z. Further, we argue that although Γ is convex in principle, it may still be useful to consider it as the convex hull of a non-convex combination of the potentials Γ_z.

To begin, we work out a toy model. Consider a state variable $\phi \in \mathbb{R}$ which is distributed according to a mixture of Gaussians with unit variance and with means at $\pm a$. That is,

$$\mathcal{H}(x) = -\log\left[\exp(-\mathcal{H}_+(\phi)) + \exp(-\mathcal{H}_-(\phi))\right] = -\log\left[\exp\left(-\frac{1}{2}(\phi-a)^2\right) + \exp\left(-\frac{1}{2}(\phi+a)^2\right)\right]$$

The Helmholtz energy of $\mathcal{H}_{\pm}(\phi)$ is

$$W_{\pm}(J) = \log \int d\phi \exp(-\mathcal{H}_{\pm}(\phi) + J\phi) = \frac{1}{2}J^2 \pm Ja$$

By applying a Legendre transform to these functions we get the effective potentials

$$\Gamma_z(\phi) = \frac{1}{2}(\phi - s_z)^2 = \mathcal{H}_z(\phi)$$

where $s_{\pm} = \pm a$. Define $\varphi = \partial_J W(J)$, the average of ϕ given source J. Then

$$\varphi = J + a \tanh Ja$$

So for $J > 1/a$, $\varphi \to J + a$. As the separation a is made larger, $1/a$ decreases, and this asymptotic behavior occurs for smaller values of J. To get a better intuition for this result, consider the full effective potential in this asymptotic case, where a is large (relative to unity, the standard deviation of a single mixture component). When $J > 1/a$, the sum over z is dominated by the term $\exp W_+(J)$. This yields:

$$\Gamma(\varphi) = \mathcal{L}[W](\varphi) \approx \mathcal{L}[W_+](\varphi) = \Gamma_+(\varphi)$$

In words, the full Helmholtz energy $W(J)$ is well-approximated by $W_+(J)$ for large positive sources J, i.e. $J > 1/a$ (and $W_-(J)$ for $J < -1/a$). Therefore, when we calculate Γ from W via Legendre transform, the fact that W is dominated individual components W_z in various regions causes Γ to approximate Γ_z in those regions.

How does this relate to our model and the definition of the critical separation s_c? First note that the toy model can be extended to a larger state space and non-Gaussian component energy functions. What is necessary is simply that the component densities are well-separated, so that $\langle \phi \rangle_J$ can take on large values at relatively small sources J. Next, our choice to define s_c in terms of the maximum of the probability density of $\phi \cdot \hat{s}$ was not crucial. The important point is that the separation between the basins is greater than the variance (in the inter-basin direction) of each basin individually. In fact, we compute s_c according to this condition later in the SM. At $J = 0$, the second derivative of $W(J)$ with respect to J gives the total covariance $\langle \phi^2 \rangle - \langle \phi \rangle^2$. For large separations, this covariance value drops off quickly as J is moved away from zero, since a slight source "pushes" the system into one basin or another. By assumption, these basins individually have much less variance than the total distribution at $J = 0$.

Typically, this structure is associated with symmetry breaking and ergodicity breaking, wherein a small external source causes the system to find a state in an ergodically broken region. In this paper, we have only been concerned with distributions over finitely many variables, and adding dynamics would require additional assumptions and structure. However, given the close analogies to standard analysis in many-body systems, we believe that an effective description of the structure of $\Gamma(\phi)$ in the well-separated phase is provided by $\Gamma_z(\phi)$ when ϕ is near s_z, i.e., the minima of $\Gamma_z(\phi)$ can sometimes be treated as metastable states. While this manifestly breaks convexity of Γ, the convex hull can always be taken if needed, and otherwise Γ can be interpreted as a constrained free energy.

3.3 RG flow in the coalesced phase

3.3.1 Units and scaling dimensions

In the previous section, the units of couplings u_2 and u_4 were not determined until after the flow equations were found. In general, this is the correct approach, namely that the units of different couplings should be verified by showing that they remove explicit scale-dependence in the flow equations. In the following discussion, the full (before irrelevant terms are removed) flow equation hierarchy is very complicated and a similar, 'by inspection' approach is difficult. In this section, we discuss heuristic reasoning that provides scaling dimensions in terms of the original energy function. However, we wish to emphasize that the true scaling dimensions may not always appear when using this heuristic, since they are really defined by a cutoff-scheme and are revealed in the flow equations.

The basic strategy follows (Bradde and Bialek [2017]) and (Lahoche et al. [2021]), though we do not take into account non-power law dependence of ρ on λ. As the cutoff scale k is changed, the number n_k of remaining modes below this cutoff scales like $n_k \sim k^\alpha$, which we denote by $[n_k] = \alpha$. What we would like to compute are the so-called 'engineering dimensions' of our couplings, which can be defined as the critical exponents at the non-interacting (Gaussian) fixed point. First, consider the trace of the propagator under a change of k, where k is the UV cutoff:

$$\mathrm{Tr}_k[G] = \sum_{a<n_k} \langle \phi_a^2 \rangle = \int_0^k d\lambda\, \rho(\lambda) \langle \phi(\lambda)^2 \rangle = \int_0^k d\lambda\, \rho(\lambda) \lambda^{-1} = \frac{\alpha N}{\Lambda^\alpha} \int_0^k d\lambda\, \lambda^{\alpha-2} \sim k^{\alpha-1}$$

We conclude that

$$[\phi] = \frac{1}{2}(\alpha - 1)$$

Note that under the identification $\lambda \sim q^2$ and $\alpha \sim D/2$ as pointed out in the Main Text, this agrees with the standard result $[\phi] \sim D/2 - 2$. Now consider a positive-definite matrix C which is diagonal in the same basis as $\langle \phi_a \phi_b \rangle$, and with eigenvalues $\tilde{\lambda}_a$, where a is ordered so that λ_a are monotonically increasing. Let the density of $\tilde{\lambda}_a$ be denoted $\tilde{\rho}$, which integrates to N. Then define $\tilde{\alpha}(k)$ as

$$\tilde{\alpha} = \partial_{\log k} \log \tilde{n}_k\,; \quad \tilde{n}_k = \int_0^k d\lambda\, \tilde{\rho}(\lambda)$$

This gives a mapping between the eigenvalues $\tilde{\lambda}_a$ of C and the eigenvalues λ_a of K through

$n_\lambda = \tilde{n}_{\tilde{\lambda}(\lambda)}$. When $\tilde{\rho}$ has power-law dependence on $\tilde{\lambda}$, $\tilde{n}_k \sim k^{\tilde{\alpha}}$, which yields:

$$\tilde{\lambda}(\lambda) \sim \lambda^{\alpha/\tilde{\alpha}} \equiv \lambda^{\beta}$$

Now the scaling dimension of C can be computed in the same way:

$$\mathrm{Tr}_k[CG] = \sum_{a<n_k} \langle C_{aa}\phi_a^2\rangle = \sum_a \lambda_a^{-1}\tilde{\lambda}_a \sim \int_0^k d\lambda\, \rho(\lambda)\lambda^{\beta-1} \sim k^{\alpha+\beta-1}$$

Hence

$$[C] = \frac{\alpha}{\tilde{\alpha}} = \beta$$

In our toy model, we use this to calculate the effective scaling dimensions of $\tilde{K}^A$ and $\tilde{K}^B$ in the mixed basis. We find that $\beta \approx 1$, meaning that, as far as power-counting rules are concerned, these matrices can be treated like derivative operators. Terms with higher powers of $\tilde{K}^z$ are increasingly irrelevant near the Gaussian fixed point.

Finally, let us consider a quartic coupling comprised of the symmetrized outer product of two matrices C_1 and C_2 which satisfy the same properties as C in the previous calculation and have dimensions β_1 and β_2 respectively. Suppose only the n_k modes below the cutoff are involved. Further, recall that these dimensions are defined as the exponents of the non-interacting theory, so the four-point function can be broken into two-point functions.

$$\sum_{a,b<n_k} (C_1 \circ C_2)_{aabb}\langle \phi_a^2\phi_b^2\rangle \sim \left(\int_0^k d\lambda\, \rho(\lambda)\lambda^{1-\beta_1}\right)\left(\int_0^k d\lambda\, \rho(\lambda)\lambda^{1-\beta_2}\right) + \text{sub-leading terms}$$
$$\sim k^{2\alpha+\beta_1+\beta_2-2}$$

And so the engineering dimensions of couplings like $C_1 \circ C_2$ are just the sum of the dimensions of C_1 and C_2, that is:

$$[C_1 \circ C_2] = \beta_1 + \beta_2\,.$$

The dimensions of couplings in the energy can be found using these rules. As an example, consider the energy of the finite ϕ^4 model, analyzed in the first section:

$$\mathcal{H}(\phi) = \frac{1}{2}\sum_{ab}(K_{ab} + u_2\delta_{ab})\phi_a\phi_b + \frac{u_4}{4!}\phi_a^2\phi_b^2$$

The dimension of $\mathcal{H}$ must be α:

$$[\mathcal{H}] = [K \cdot \phi^2] = 1 + 2[\phi] = \alpha$$

Following the rules above, we further find:

$$[u_2] = 1, \quad [u_4] = 2 - \alpha$$

We reiterate that this method is heuristic, and rescaling by these dimensions will not always give scale-free flows. The primary reasons why this calculation may fail are anomalous scaling, the presence of operators like $\tilde{K}^z$ (which will be defined in the next section) which do not commute with the propagator G. When this latter situation is the case, $\tilde{K}^z$ is not diagonal in the collective variable basis, and so the k-scaling dimensions of loop sums like $\mathrm{Tr}[\dot{R}_k G_k \tilde{K}^A G_k \tilde{K}^A G_k]$ are not simply given by the sum of dimensions of operators under the trace. We are fortunate in this work that although this non-commutativity is present, the $\tilde{K}^z$ matrices appear to be approximately or effectively diagonal in the collective basis, and the loop sums have scaling dimensions which can be naïvely approximated.

A third mechanism which can break this heuristic is the presence of operators which are not scale-free. One example could be if K were not scale free, causing the eigenvalue density $\rho(\lambda)$ to not have simple power-law dependence on λ. This can cause the scaling dimensions to change along the flow (Lahoche et al. [2021]) and demonstrates why it is necessary to define the true scaling dimensions in terms of the flow equations and not *a priori.* This contingency also demonstrates the power of NPRG formalism, since the flow equations are defined whether or not a scale-free description is available.

3.3.2 RG calculation

In the coalesced phase, the separation s between s_A and s_B is smaller than a critical value s_c, which is calculated in the next section. Here we set $|s| = 0$ which significantly simplifies the analysis. Expanding (3.6) with $s_A = s_B$ about the point $\phi = 0$ yields at the quadratic level,

$$\begin{aligned}\mathcal{H}^{(2)} &= \frac{1}{2}\left(K^A + K^B + (u_2^A + u_2^B)\delta\right) \\ &= K + u_2\delta\,.\end{aligned}$$

In the above expressions, we use δ to denote the identity matrix. Let us first consider the quadratic part, which we use to define scale. The density of eigenvalues of K is given by

$$\rho(\lambda) = \sum_{a=1}^{N} \delta(\lambda - \lambda_a)\,,$$

which for large N can be treated as an effectively smooth distribution, with proper care. Approximating these eigenvalues to be power-law distributed, define the scaling dimension α as

$$\alpha(k) = \partial_{\log k} \log \int_0^k d\lambda\, \rho(\lambda) \approx \alpha.$$

Next, We can calculate the scaling dimensions of K^A and K^B as discussed in the previous section. When written in the K-eigenbasis, K^A and K^B appear as approximately diagonal. The precise reasons for this are beyond the scope of this work, so we merely take it as an experimental fact. We must be careful to note that these matrices are not truly diagonal in the K-eigenbasis. Additionally, K^A and K^B display mass gaps in their diagonal elements just as K does. We define their scale-free parts by subtracting off these mass gaps:

$$\tilde{K}^A = K^A - K_0^A \delta$$
$$\tilde{K}^B = K^B - K_0^B \delta$$

This allows us to compute the k-scaling dimension of $\tilde{K}^A$, which we also approximate as constant with respect to k

$$[\tilde{K}^A] = \beta_A(k) = \partial_{\log k} \log \int_0^k d\lambda \sum_a \delta\left(\lambda - \mathrm{Diag}_a\{\tilde{K}^A\}\right) \approx \beta_A,$$

and similarly for $\tilde{K}^B$, where by $\mathrm{Diag}\{\tilde{K}^A\}$ we mean the diagonal element in the K-eigenbasis corresponding to the $\lambda + K_0$ eigenvalue of K.

In the topmost plot of Fig. 3.2, we show that by restricting to $\alpha_A = \alpha_B$ and sweeping, β_A and β_A are essentially unity, and that α has a predictable functional dependence. While small deviations from power-law behavior may be present, we approximate $\beta_A = \beta_B = 1$ for the rest of the analysis, unless otherwise specified.

We now turn our attention to the quartic couplings, which will contain most of the rest of the building blocks of our ansatz.

$$\begin{aligned}\mathcal{H}^{(4)} = &-\frac{3}{4}K^A \circ K^A - \frac{3}{4}K^B \circ K^B + \frac{3}{2}K^A \circ K^B \\ &- \frac{3(u_2^A - u_2^B)}{2}K^A \circ \delta - \frac{3(u_2^B - u_2^A)}{2}K^B \circ \delta \\ &+ \left(\frac{u_4^A}{2} + \frac{u_4^B}{2} - \frac{3((u_2^A)^2 + (u_2^B)^2)}{4}\right)\delta \circ \delta\,.\end{aligned}$$

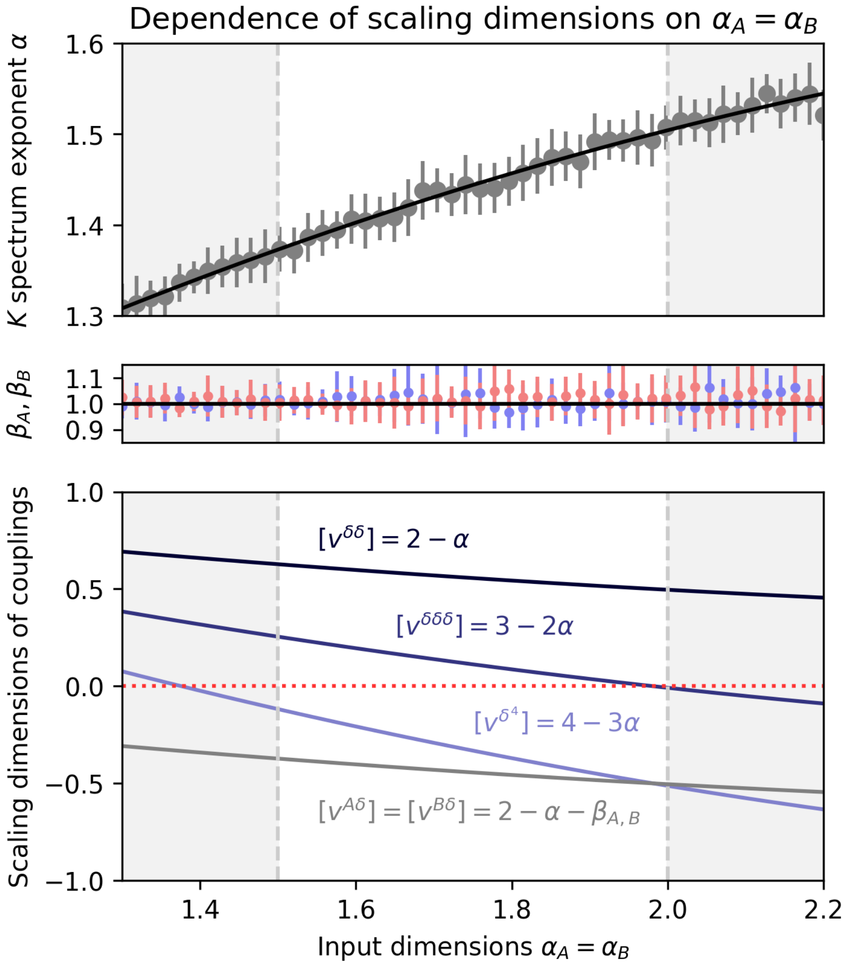

Figure 3.2: Numerical examination of the exponent α describing the scaling properties of eigenvalues of K. Areas without grey shading denote $\alpha_A = \alpha_B \in (1.5, 2)$, which corresponds to $D \in (3, 4)$. **Top**: Estimates of the exponent α were obtained by explicitly constructing K and measuring power laws in the scale-free part of their spectra. Error bars represent the root mean square error over 1 000 trials at each value of α_A. Matrix size is $N = 75$. Black line is a polynomial fit which we use to find scaling dimensions. **Middle**: Estimates of β_A and β_B show no significant deviation from unity. **Bottom**: scaling dimensions of various couplings. The only couplings which are relevant in the range $\alpha_A = \alpha_B \in (1.5, 2)$ are $v^{\delta\delta}$ and $v^{\delta\delta\delta}$

The use of the symbol $\circ$ here denotes a symmetrized outer product:

$$\begin{aligned}(K^A \circ K^A)_{abcd} &= K^A_{(ab}K^A_{cd)} \\ &= \frac{1}{4!}\sum_{\text{perms}\,p} K^A_{p_1p_2}K^A_{p_3p_4}\,.\end{aligned}$$

Again, it is useful to separate out the couplings into scale-free parts. This yields:

$$\begin{aligned}\mathcal{H}^{(4)} &= -\frac{3}{4}\tilde{K}^A \circ \tilde{K}^A - \frac{3}{4}\tilde{K}^B \circ \tilde{K}^B + \frac{3}{2}\tilde{K}^A \circ \tilde{K}^B \\ &+ \frac{3}{2}(u_2^B + K_0^B - u_2^A - K_0^A)\tilde{K}^A \circ \delta \\ &+ \frac{3}{2}(u_2^A + K_0^A - u_2^B - K_0^B)\tilde{K}^B \circ \delta \\ &+ \left(\frac{u_4^A}{2} + \frac{u_4^B}{2} - \frac{3}{4}\left(u_2^A + K_0^A - u_2^B - K_0^B\right)^2\right)\delta \circ \delta\,.\end{aligned}$$

Finally, for the sixth-order part, we discard all but the $\delta \circ \delta \circ \delta$ term, which we shall justify shortly.

$$\begin{aligned}\mathcal{H}^{(6)}_{a\ldots f} = \frac{15}{4}&(u_4^A - u_4^B)\times \\ &(u_2^B - K_0^B - u_2^A + K_0^A)(\delta \circ \delta \circ \delta)_{a\ldots f}\end{aligned}$$

For the moment, we explicitly include all of these operators in our ansatz for the effective action Γ_k. This gives at least ten couplings in total; three quadratic operators, six quartic, and one sixtic:

$$\Gamma^{(2)}_{k\,ab} = (v^A_k \tilde{K}^A + v^B_k \tilde{K}^B + v^\delta_k \delta)_{ab} \tag{3.12}$$

$$\begin{aligned}\Gamma^{(4)}_{k\,abcd} = (&v^{AA}_k \tilde{K}^A \circ \tilde{K}^A + v^{AB}_k \tilde{K}^A \circ \tilde{K}^B + \\ &v^{BB}_k \tilde{K}^B \circ \tilde{K}^B + v^{A\delta}_k \tilde{K}^A \circ \delta + \\ &v^{B\delta}_k \tilde{K}^B \circ \delta + v^{\delta\delta}_k \delta \circ \delta)_{abcd}\end{aligned} \tag{3.13}$$

$$\Gamma^{(6)}_{k\,a\ldots f} = (\cdots + v^{\delta\delta\delta}_k \delta \circ \delta \circ \delta)_{a\ldots f} \tag{3.14}$$

Many of these couplings are irrelevant for $\alpha_A = \alpha_B \in (1.5, 2)$, including all sixtic couplings except $v^{\delta\delta\delta}$. This follows from the dimensional analysis rules we discussed, together with our definition of α.

$$\begin{aligned}[]
[\phi] &= \frac{1}{2}(\alpha - 1), \quad [\mathcal{H}^{(n)}] = \alpha\left(1 - \frac{n}{2}\right) + \frac{n}{2} \\
[v^A] &= 1 - \beta_A = 0, \quad [v^B] = 1 - \beta_B = 0, \quad [v^\delta] = 1 \\
[v^{z_1 z_2}] &= 2 - \alpha - \beta_{z_1} - \beta_{z_2}\,; \quad z_i \in \{A, B, \delta\}, \quad \beta_\delta = 0 \\
[v^{\delta\delta\delta}] &= 3 - 2\alpha
\end{aligned}$$

A few of these scaling dimensions are represented as a function of $\alpha_A = \alpha_B$ in Fig. 3.2. Terms like $\tilde{K}^A \circ \delta$ and $\tilde{K}^A \circ \tilde{K}^B$ have negative dimensions for the range of spectral exponents we are considering, so these are termed irrelevant. Such terms can be important for obtaining accurate estimates of non-universal properties in terms of the original couplings, but deep in the IR, they only have the effect of shifting values of relevant couplings. As discussed, these dimensions are derived heuristically, and must be verified.

The ansatz (3.12-3.14) generates an approximately closed set of flow equations given the truncation, since the new couplings generated in the flow are driven only by sub-extensive contributions. As an example, we verify the predicted value of $s_{z_1 z_2} = [v_k^{z_1 z_2}]$ above. Looking only at the diagram with two 4-point vertices and keeping only extensive terms (see first section for explanation),

$$\partial_k v_k^{z_1 z_2} = \frac{1}{3} \sum_{\substack{x\in\{A,B,\delta\} \\ y\in\{A,B,\delta\}}} v_k^{z_1 x} v_k^{z_2 y} \,\mathrm{Tr}\left[\dot{R}_k G_k C_x G_k C_y G_k\right] \; ; \quad C_A = \tilde{K}^A,\, C_B = \tilde{K}^B,\, C_\delta = \delta$$

The LHS must have dimension $s_{z_1 z_2} - 1$, while the RHS is a little trickier. Consider a single term:

$$s_{z_1 z_2} = 1 + s_{z_1 x} + s_{z_2 y} + [\Phi_k] \, ; \quad \Phi_k^{xy} = \mathrm{Tr}\left[\dot{R}_k G_k C_x G_k C_y G_k\right] \tag{3.15}$$

In reality, these C matrices do not generally commute with G_k. Yet, given the setup of our problem, the $\tilde{K}^z$ matrices effectively act as if they are diagonal when in the K eigenbasis. The catch is that their diagonals are not eigenvalues and the scale free parts are no longer described by the exponents α_A and α_B, but instead with $\alpha(\alpha_A, \alpha_B)$, as can be easily numerically verified. Therefore,

$$[\Phi_k^{xy}] = -3 + \alpha + \beta_x + \beta_y$$

Which shows, after direct substitution into (3.15) that the free-field guess at $s_{z_1 z_2}$ is correct.

The ansatz (3.12-3.14) can be used to find flow equations for all couplings, and these can be integrated numerically. At present we are only interested in asymptotic properties, so we can thin out the effective action ansatz by removing irrelevant terms. This takes us from ten down to five couplings:

$$\Gamma^{(2)}_{k\,ab} = (v_k^A \tilde{K}^A + v_k^B \tilde{K}^B + v_k^\delta \delta)_{ab} \tag{3.16}$$

$$\Gamma^{(4)}_{k\,abcd} = v_k^{\delta\delta} (\delta \circ \delta)_{abcd} \tag{3.17}$$

$$\Gamma^{(6)}_{k\,a\ldots f} = v_k^{\delta\delta\delta} (\delta \circ \delta \circ \delta)_{a\ldots f} \tag{3.18}$$

Because $v^{A\delta}$ and $v^{B\delta}$ are irrelevant and have been dropped, v^A and v^B are not significantly renormalized by any couplings in this ansatz. The dimensionless flow equations for relevant parameters, computed using the Litim regulator and not accounting for anomalous scaling, are given by:

$$\begin{aligned}
\dot{g}_l^{\delta} &= g_l^{\delta} + \frac{1}{6}\frac{g_l^{\delta\delta}}{(1+g_l^{\delta})^2} \\
\dot{g}_l^{\delta\delta} &= (2-\alpha)g_l^{\delta\delta} - \frac{1}{3}\frac{(g_l^{\delta\delta})^2}{(1+g_l^{\delta})^3} + \frac{1}{10}\frac{g^{\delta\delta\delta}}{(1+g^{\delta})^2} \\
\dot{g}_l^{\delta\delta\delta} &= (3-2\alpha)g_l^{\delta\delta\delta} + \frac{5}{3}\frac{(g_l^{\delta\delta})^3}{(1+g_l^{\delta})^4} - \frac{g_l^{\delta\delta}\, g_l^{\delta\delta\delta}}{(1+g_l^{\delta})^3}
\end{aligned}$$

3.4 Calculation of multi-relevance breakdown separation s_c

3.4.1 Mean-field saddle point method

To begin, assume the A and B models have all the same parameters, e.g. $u_2^A = u_2^B$, etc., but the eigenbases of K^A and K^B are random with respect to each other. (Each drawn from the Gaussian Orthogonal Ensemble.) The saddle point state is given by:

$$\phi_{\text{sp}} = \frac{1}{2}M^{-1}\left(M^B - M^A\right)\cdot s + O(|s|^3)$$

Where $M^A = K^A + u_2^A$ and $M = M^A + M^B$. The truncation in powers of $|s|$ is due to the fact that s_c should go as $\Lambda^{-1/2}$, which we take to be small. Set $s_A = |s|/2$ and $s_B = -s_A$ without loss of generality. The saddle point condition is approximately satisfied at the origin:

$$\frac{1}{2}M^{-1}\left(M^B - M^A\right)\cdot s \approx \frac{1}{2}M^{-1}\left(\frac{1}{N}\operatorname{Tr} M^B - \frac{1}{N}\operatorname{Tr} M^A\right)|s| \approx 0$$

An important approximation we made here is that since s is randomly oriented with respect to the eigenbases of K^A and K^B, it is approximately an eigenvector for the ranges

α we are dealing with, as can be verified numerically. Explicitly,

$$M^A \cdot s \approx \frac{1}{N}\left(\operatorname{Tr} M^A\right)|s| \,.$$

At the origin, the energy Hessian up to $O(|s|^4)$ is

$$\mathcal{H}^{(2)}_{ab} = M_{ab} - \frac{1}{4}(M \cdot s)_a (M \cdot s)_b + \frac{1}{2^4 3}(u_4^A + u_4^B)(s^2 \delta_{ab} + 2 s_A s_B) + O(|s|^4) \tag{3.19}$$

$$= M_{ab} + \frac{1}{2^3 3} u_4 s^2 \delta_{ab} - \frac{1}{4}\left(\left(\frac{1}{N}\operatorname{Tr} M\right)^2 - \frac{1}{3}u_4\right) s_A s_B \tag{3.20}$$

Though s is not an exact eigenvector of this Hessian, we approximate it as one. Its ‘eigenvalue’ can be found from the above:

$$(\mathcal{H}^{(2)} \cdot s)_a = \left(\frac{1}{N}\operatorname{Tr} M + \frac{1}{2^3}u_4 s^2 - \frac{1}{2^2}\left(\frac{1}{N}\operatorname{Tr} M\right)^2 s^2\right) s_A$$

Setting this to zero, we can solve for the point at which curvature along the separation vector between the basins goes from positive to negative:

$$s_c^2 = 4\left(\left(\frac{1}{N}\operatorname{Tr} M\right)^2 - \frac{1}{2}u_4\right)^{-1} \tag{3.21}$$

3.4.2 Mixture component variance calculation

The alternative method for calculating the critical separation is to leverage the fact that our model is a mixture of densities:

$$P_z(\phi) \propto \exp(-\mathcal{H}_z(\phi))\,, \quad P(\phi) \propto \exp(-\mathcal{H}_A(\phi)) + \exp(-\mathcal{H}_B(s))$$

Because we have assumed all of the same parameters for the A and B states (except basis),

$$P(\phi) = \frac{1}{2}P_A(\phi) + \frac{1}{2}P_B(\phi)$$

To calculate the probability density of ϕ along $\hat{s}$, the unit vector along the basin separation, we evaluate

$$\begin{aligned} P(\hat{s}\cdot\phi = s) &= \int d\phi\, \delta(s - \hat{s}\cdot\phi)P(\phi) \\ &= \frac{1}{2}P_A(\hat{s}\cdot\phi = s) + \frac{1}{2}P_B(\hat{s}\cdot\phi = s) \end{aligned}$$

In the independent phase, $P_A(\hat{s}\cdot\phi)$ does not significantly overlap $P_B(\hat{s}\cdot\phi)$. If A and B are in their disordered phases, these distributions will be Gaussian, while in their ordered phases they will each look like a mixture two Gaussians placed at $\pm\hat{s}_0|\langle\phi_0\rangle_z|$, where by $|\langle\phi_0\rangle_z|$ we simply mean the magnetization in the standard sense. In both cases the coalescence transition occurs roughly when the distance between the centers of these distributions is twice the standard deviation of one of them, causing overlap:

$$\left(\frac{s_c}{2}\right)^2 = \langle(\hat{s}\cdot\phi)^2\rangle_z = \sum_{ab} \hat{s}_a\hat{s}_b \left(\Gamma^{(2)}_{z,k\to 0}\right)^{-1}_{ab} \tag{3.22}$$

From the ansatz for Γ_k,

$$\left(\Gamma^{(2)}_{z,0}\right)_{ab} = (\lambda^z_a + u^z_{2,0})\delta_{ab}$$

In the K^z eigenbasis. Using the fact that s has a random orientation in this basis,

$$\sum_{ab} \hat{s}_a\hat{s}_b \left(\Gamma^{(2)}_{z,0}\right)^{-1}_{ab} \approx \frac{1}{N}\,\mathrm{Tr}\left(\Gamma^{(2)}_{z,0}\right)^{-1}$$

In the disordered phase, $u^z_{2,0} > 0$, so

$$\frac{1}{N}\,\mathrm{Tr}\left(\Gamma^{(2)}_{z,0}\right)^{-1} = \frac{1}{N}\int_0^{\Lambda_z} d\lambda\,\rho_z(\lambda)\frac{1}{\lambda + u^z_{2,0}}$$

In the disordered phase, $u^z_{2,0} < 0$, so expand the potential instead around one of the minima at $\langle\phi_0\rangle = \pm(-6u^z_{2,0}/u^z_{4,0})^{1/2}$. The total variance is $\langle\phi_0\rangle^2$ plus trace the inverse of $\Gamma^{(2)}$ evaluated at the symmetry-broken expansion point:

$$\sum_{ab}\hat{s}_a\hat{s}_b\left(\Gamma^{(2)}_{z,0}\right)^{-1}_{ab} \approx \frac{1}{N}\left(-\frac{6u^z_{2,0}}{u^z_{4,0}} + \int_0^{\Lambda_z} d\lambda\,\rho_z(\lambda)\frac{1}{\lambda - 2u^z_{2,0}}\right)$$

There is a normalization factor on ρ_z so that it integrates to N. For large N, our assumption is that

$$\rho_z(\lambda) \to \frac{N\alpha_z}{\Lambda_z^{\alpha_z}}\lambda^{\alpha_z - 1}$$

Hence the disordered phase gives

$$\left(\frac{s_c}{2}\right)^2 \approx \frac{\alpha_z}{\Lambda_z^{\alpha_z}}\int_0^{\Lambda_z} d\lambda\,\frac{\lambda^{\alpha_z-1}}{\lambda + u^z_{2,0}}$$

While the ordered phase gives

$$\left(\frac{s_c}{2}\right)^2 \approx -\frac{6}{N}\frac{u^z_{2,0}}{u^z_{4,0}} + \frac{\alpha_z}{\Lambda_z^{\alpha_z}}\int_0^{\Lambda_z} d\lambda\,\frac{\lambda^{\alpha_z-1}}{\lambda + u^z_{2,0}}$$

CHAPTER 4

COARSE-GRAINING TO REVEAL DIFFERENT COLLECTIVE COMPUTATIONS IN A SENSORY POPULATION

AGK[1], Aleksandra M. Walczak[2], Thierry Mora[2], Maciej Koch-Janusz[1,3], and Stephanie E. Palmer[1,4]

1. *Department of Physics, The University of Chicago, Chicago IL 60637*

2. *Laboratoire de physique de l' École normale supérieure, CNRS, PSL University, Sorbonne Université and Université de Paris, 75005 Paris, France*

3. *Haiqu, Inc., 95 Third Street, San Francisco, CA 94103, USA*

4. *Department of Organismal Biology and Anatomy, The University of Chicago, Chicago IL 60637*

4.1 Abstract

The vertebrate retina performs prediction on incoming visual signals, which can compensate for lags in neural processing. It has been hypothesized that prediction occurs at each successive layer of the visual stream, but downstream prediction of the retina is not yet understood. One challenge is that retinal responses are collective, and full recovery of predictive information must take into account correlations in the joint activity of large populations; this incurs the curse of dimensionality. Another challenge arises when the stimulus is naturalistic, since relevant features in complex scenes are typically unknown. Furthermore, estimates of available predictive information in complex scenes and the responses they elicit are difficult. In this work, we address these challenges simultaneously by searching for maximally-predictive collective variables in a large population of 93 salamander retinal ganglion cells under naturalistic stimulus. To achieve this, we apply a tractable, approximate implementation of the information bottleneck method to our neural data and infer a lower-dimensional representation that is maximally informative about the future neural activity. We observe that across stimuli and intervals, all predictive information in the retinal outputs is captured by a few linear collective variables. We further show that predictive signals are collectively encoded. At short timescales, this coding is less collective and noise correlations contribute significantly, while at later timescales predictive features are highly collective and stimulus-induced correlations dominate. Our analysis demonstrates the feasibility of finding biologically relevant coarse-graining schemes in high-dimensional data using variational inference and basic machine learning tools.

4.2 Introduction

A feasible account for the ubiquity of prediction in biology is that it allows agents to exploit cause-and-effect relationships in their environment. The success of a species depends on its ability to survive and reproduce, and prediction is crucial to reacting quickly in life-or-death situations. Prediction has therefore been explored as a normative principle for understanding how organisms assign utility to various components of incoming signals in early stages of sensory processing (Creutzig and Sprekeler [2008], Creutzig et al. [2009], Rust and Palmer [2021]). In the vertebrate visual processing stream, visual signals are propagated through several layers of neurons, each with significant sensory delays, and it has been proposed that prediction may help compensate for these delays. This begins in the eye, which can optimally predict visual signals in some environments (Berry et al. [1999], Gollisch and Meister [2010], Palmer et al. [2015], Sederberg et al. [2018], Liu et al. [2021]). While optimal prediction has been observed for small groups of cells in the vertebrate retina under simple stimuli (Palmer et al. [2015]), we expect that this computation should not only extend to richer more complex stimuli, but should also continue deeper along the visual stream, layer upon layer, in order to produce the fastest possible reaction time in natural scenarios.

Visual signals are first processed in the retina, then sent downstream to ultimately drive behavior. This population of neurons contains all that the brain sees, and all predictive computations rely on the collective correlation structure of this code. If neurons downstream from the retina are predicting its activity, they must do so without recourse to the visual inputs. We therefore aim examine the statistics of retinal responses to complex, naturalistic stimuli, in an effort to understand the downstream prediction problem. In particular, are there a few coarse-grained features in the output code which capture all of the predictive information? Do these features depend on the prediction interval? If we are able to find them, is it conceivable that they could be read out by real downstream neurons? And finally, how collective is predictive information? That is, does effective prediction of retinal activity require pooling the outputs of many neurons?

To find and study predictive features in the retinal code, we leverage recent machine learning methods with foundations in information theory. At its core, our method solves a variational version of the predictive information bottleneck problem (Tishby et al. [2000]) by simultaneously searching over predictive features and performing inference. Letting X represent the states of retinal outputs at the present and Y the outputs in the future, we seek deterministic, coarse-grained representations $Z = \Lambda(X)$ of a fixed dimensionality that carry maximal information about Y. Both the features Λ and a trial distribution $q(y|z)$ are expressed as neural networks and optimized simultaneously. This method is adapted from approaches originally developed to find coarse-graining schemes in physical systems by maximizing long-range information (Koch-Janusz and Ringel [2018], Gökmen et al. [2021]), and aligns with recent efforts to connect predictive information bottleneck to a generalized approach to model reduction (Schmitt et al. [2023]). We wish to highlight that by using the expressive power of neural networks to solve an inference problem, this method is able to estimate information quantities on a state space of unprecedented size, that is, the state space of joint activity among hundreds of neurons. The performance of this method is perhaps striking given the simplicity of its theoretical motivation.

To study how the collective properties of retinal responses might allow for downstream prediction, we must be able to observe correlations in the simultaneous activity of a large number of neurons. Moreover, given the wide range of timescales present in natural visual environments (Salisbury and Palmer [2016]) and the retina's ability to resolve very fast features, we need access to long recordings with high resolution in time. The dataset we examine contains electrode recordings of 93 ganglion cells in the salamander retina under five different naturalistic stimuli, giving us access to large scale features (population-scale correlations) and a wide range of timescales, down to 60 Hz, or the frame rate of the stimuli. Moreover, the dataset offers responses to repeated stimuli, allowing us to tease apart the contributions to predictive information which are directly due to the stimulus from those which are due to the underlying physiology of the retina.

4.3 Methods

4.3.1 Data collection

In order to investigate the statistics of retinal outputs, we examine response data taken from 93 retinal ganglion cells (RGCs) in salamander retina under naturalistic stimuli. The data collection process is outlined in Fig. 4.1. A retina taken from a larval tiger salamander was placed on an electrode array with a density of roughly one electrode per RGC. These data were then spike-sorted, yielding the times at which each neuron fired an action potential. For our purposes, these discrete events are effectively identical, so all of the information they convey is in their timing (Strong et al. [1998]). Next, a binary representation of the neural code was generated by discretizing time into time bins of size $\Delta t = 1/60$ s, and assigning

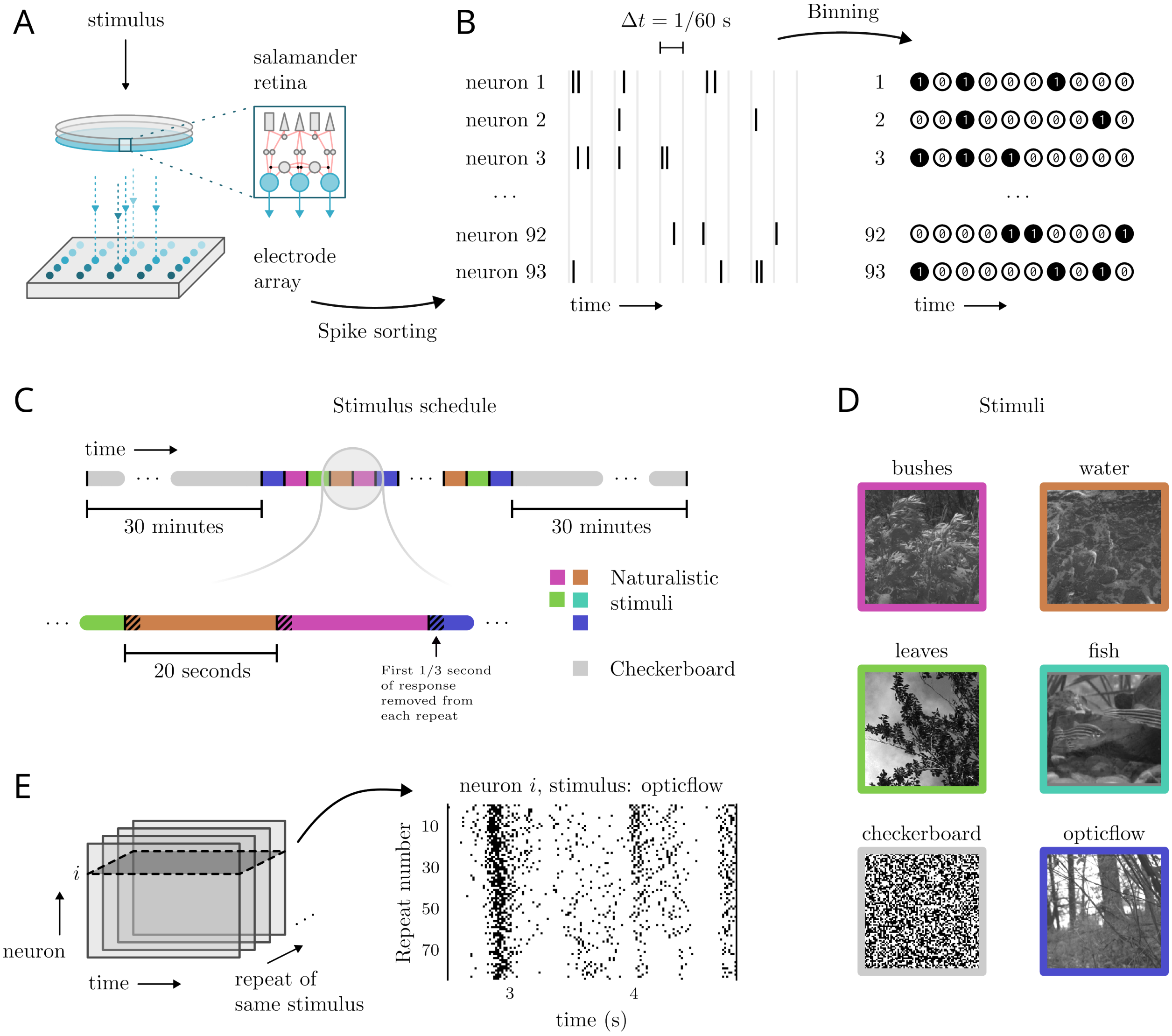

Figure 4.1: Experimental protocol enables simultaneous recording of responses across a large population of salamander retinal ganglion cells under repeated, naturalistic stimuli. A. Stimuli are shown to a salamander retina while ganglion cell responses are recorded by an electrode array. B. Following spike sorting from electrode array, neuron spike times are binned into $\Delta t = 1/60$ second time bins (aligned with stimulus movie frames), which yields a binary representation of the neural code. C. Naturalistic stimuli (60 fps) are shown to the retina in random order, in 20 second intervals. In our analysis, the responses from the first 20 frames (0.3 seconds) of each interval are removed, as they contain transients due to switching stimulus. At the beginning and end of the experiment are 30 minute intervals of checkerboard stimulus. D. Snapshots taken from each stimulus. E. Each stimulus is shown to the retina ~80 times, allowing us to observe noise correlation effects.

the state “1” to a neuron in any time bin in which it fired a spike and a “0” otherwise. In our data, the marginal probability of the “1” state across all neurons and time bins can be as high as 0.016 and as low as 0.003, depending on stimulus.

Five different naturalistic stimulus clips were used, each 20 seconds long. These were shown to the retina in random order for 141 minutes, yielding roughly 80-90 repetitions of each stimulus over the course of the experiment. This repeated structure allows us to compare neural responses to a given stimulus across trials, revealing both reliable response features and stochastic effects inherent to the physiology of the retina. Correlations in these latter effects are referred to as “noise correlations” and can be studied by considering the ensemble of repeated trials of the same stimulus. The stimuli themselves are greyscale videos at 60Hz, depicting a range of visual scenes that could conceivably be present in the environment of the specimen. Such stimuli are characterized by heavy-tailed distributions of contrast, velocity, time-, and length-scales, in contrast to simpler, engineered stimuli such as moving gratings (Salisbury and Palmer [2015, 2016]). Because the retina encodes its stimuli nonlinearly, an ethologically relevant understanding of these encodings may require naturalistic stimuli.

4.3.2 Estimating mutual information

A central focus of this work is quantifying the amount of information encoded in the outputs of the retina about its future outputs. Generally, given a joint distribution $p(x, y)$ for two random variables X and Y, the mutual information between them is given by

$$I(X;Y) = \left\langle \log \frac{p(x,y)}{p(x)p(y)} \right\rangle_{p(x,y)} . \tag{4.1}$$

In general, mutual information quantities can be very challenging to compute. A common difficulty is if X or Y are variables with many dimensions, say $X \in \{0,1\}^N$ with N large.

In such cases, even if given an analytical expression for $p(x, y)$, the integral in $\langle \cdot \rangle_{x,y}$ will typically be intractable. Moreover, when working with data, one has the issue of determining $p(x, y)$. This problem is known as inference, and also becomes increasingly difficult as the dimensionality of random variables increases.

In this work, we leverage recently-developed tools to compute a tractable lower bound to (4.1). This lower bound depends on an ansatz for $p(x, y)$, represented using neural networks, and maximizing this lower bound implies that the ansatz approaches the true distribution. In other words, this method both solves an inference problem and in the process computes an approximation of $I(X; Y)$.

More explicitly, we consider a trial distribution $q(x, y|\theta)$ which represents our best guess at $p(x, y)$, where θ are free parameters. The Barber-Agakov (BA) bound on mutual information is given by

$$\begin{aligned} I_{\text{BA}}(X;Y)(\theta) &= \left\langle \log \frac{q(x, y|\theta)}{p(x)p(y)} \right\rangle_{p(x,y)} \\ &= I(X;Y) - D_{\text{KL}}[p(x, y)||q(x, y|\theta)] \, . \end{aligned}$$

Because $D_{\text{KL}}[p||q] \geq 0$, we have that $I_{\text{BA}}(X;Y) \leq I(X;Y)$ with equality only when $q(x, y|\theta) = p(x, y)$. Therefore the act of maximizing this bound is precisely inference. We parameterize this trial distribution q with the ansatz

$$q(x, y|\theta) = \frac{p(x)p(y)}{Z(y, \theta)} e^{f(x,y|\theta)} \, , \tag{4.2}$$

Where $p(x)$ is the true marginal of x and $f(x, y|\theta)$ is computed by a neural network with parameters θ. The partition function $Z(y, \theta)$ is then defined by $f(x, y|\theta)$ and $p(x)$ through the requirement that $q(x|y, \theta)$ is normalized for all y. Inserting this ansatz into the BA bound yields the unnormalized Barber-Agakov Bound

$$I_{\text{UBA}}(X;Y)(\theta) = \langle f(x,y|\theta) - \log Z(y,\theta) \rangle_{p(x,y)} \ .$$

This bound is intractable due to the log partition function. The Nguyen, Wainwright, Jordan (NWJ) lower bound effectively deals with this problem but introduces a large amount of variance in the estimate of $\log Z(y)$. This is an issue given that we are interested in treating this bound as an estimate of mutual information. In the noise-contrastive estimate (NCE) bound, a Monte-Carlo estimate of the partition function is computed using minibatches containing B independent samples $\{(x_i, y_i)\}_{i=1}^{B}$, and this effectively reduces variance of the final estimate. Explicitly, the NCE bound is

$$I_{\text{NCE}}(X;Y)(\theta) = \frac{1}{B}\left\langle \sum_{j=1}^{B} f(x_j, y_j|\theta) - \log\left(B^{-1}\sum_{i=1}^{B} e^{f(x_i,y_j|\theta)} \right) \right\rangle_{\prod_{i=1}^{B} p(x_i,y_i)} \tag{4.3}$$

In summary, we aim to compute $I_{\text{NCE}}(X;Y)$, which is bounded in the following way

$$I_{\text{NCE}}(X;Y)(\theta) \leq I_{\text{BA}}(X;Y)(\theta) \leq I(X;Y)\,,$$

With each bound becoming tight in the case that $q(x,y|\theta) = p(x,y)$.

Several discussions of this method and its applications are available (Gökmen et al. [2021], Poole et al. [2019]). We implement the critic function $f(x,y)$ using fully connected neural networks $u_a(x)$ and $v_a(y)$ with $a = 1, \ldots, N_{\text{embed}}$ as:

$$f(x,y|\theta) = \sum_{a=1}^{N_{\text{embed}}} u_a(x|\theta) v_a(y|\theta)$$

4.4 Predictive information persists at long timescales

Ultimately, the retina needs to reliably encode and transmit visual information to further processing stages downstream, and a full understanding of its dynamics requires some knowledge about stimulus. However, it is clearly true that these downstream populations do not have direct access to the stimulus in addition to retinal outputs. If we look at the code by itself, as if it was produced by an autonomous dynamical system, what can we say about its dynamics?

As a first step, we examine predictive information and find that it persists out to very long timescales for a range of stimuli. Specifically, we estimate the mutual information between all 93 neurons in a single time bin at any given time t and the same set of degrees of freedom at τ into the future (Fig 4.2A). These two random variables, each a word of 93 binary numbers, are denoted X and Y respectively. Our estimate of $I(X;Y)$ is given by $I_{\text{NCE}}(X;Y)$ (in Eq (4.3)), evaluated on data which were held out during training of the critic function.

In Fig 4.2B, estimates of $I_{\text{NCE}}(X;Y)$ are given as a function of the prediction interval τ, for each stimulus. Since I_{NCE} is a lower bound of mutual information, we can conclude that for some stimuli, such as "fish" and "opticflow", the true mutual information $I(X;Y)$ achieves large values at prediction intervals as large as ∼500 ms. Additionally, there is quite a bit of variation in the overall scale of information across stimuli, and in the shape of dependence on τ.

To better observe this variation in dependence, we examine in Fig 4.2C the decay of predictive information on a log scale, after normalizing by the maximum value of $I_{\text{NCE}}(X,Y)$ at $\tau = \Delta t$. These trends show approximately exponential decay for most stimuli over some initial transient period, followed by a plateau. In the "opticflow" stimulus, this profile is quite exaggerated, with the plateau onset occurring after only about 150 ms, followed by apparently a slight *increase* in information. To be sure that these are real effects, we

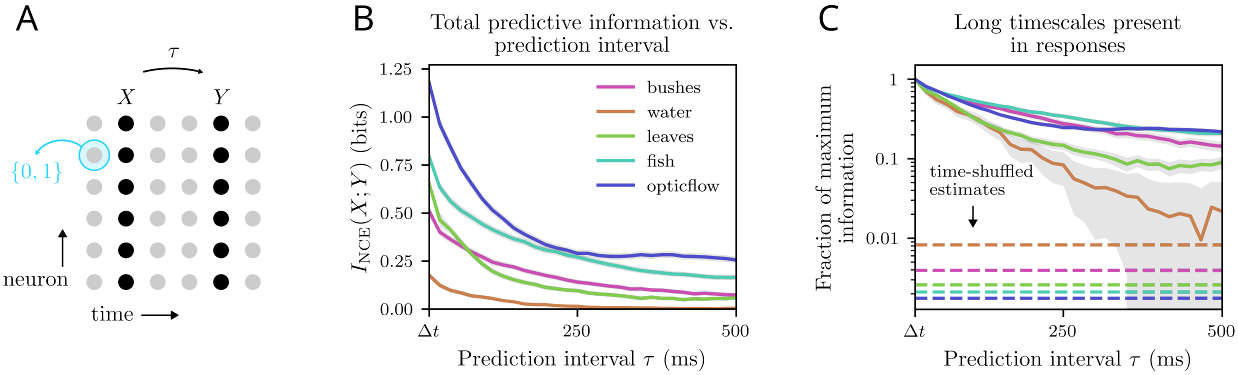

Figure 4.2: Naturalistic stimuli induce very long timescales in retinal responses, which we resolve using our mutual information estimate. A. We use the NCE bound (4.3) to estimate mutual information $I(X;Y)$ between X and Y. X is the binary code word of length $N = 93$ representing ganglion cell responses in a single time bin, while Y is the code word of the same size at τ in the future. B. Dependence of total predictive information on prediction interval for each stimulus. C. Trends from B are presented on a log scale, normalized by their maximum values, with the same coloring scheme. Each stimulus response displays multiple timescales and correlations at very long times. Dotted lines are values obtained by measuring data with time bins randomly shuffled and represent the noise floor.

have included estimates using the same method on shuffled data. This shuffling randomly permutes the time index in a way that is consistent across repeated trials of the same stimulus, which allows for the shuffled estimate to include the over-fitting contribution due to "memorizing" frame number while destroying all other correlations in time.

4.5 Predictive information is compressible

The presence of mutual information between points separated by large time intervals τ reveals that there are features of the neural code that are predictable in principle. However, is it reasonable to expect that neurons downstream of the retina can do this prediction task? For example, given some stimulus, how many distinct features of the random variable $X \in \{0,1\}^{93}$ need to be measured to learn all there is to know about the future state Y, and what are these features? Given the importance of prediction for survival, it would not make much sense if the outputs of every single neuron had to be known in full fidelity to predict what will happen next. For one, we know that the neural code is noisy, with repeated exposure to the same stimulus giving variable responses. Secondly, we know that the optic tectum, the next stage downstream of the retina in the visual stream, has fewer neurons and acts as a bottleneck for the signal. Furthermore, signals which drive motor responses also converge to a small set of motor neurons, indicating a need for compressibility. Such considerations also suggest that an important aspect of retinal computation is its ability to make signals accessible through linear readout downstream Gollisch and Meister [2010]

Here, we directly quantify compressibility of predictive information in the retinal code by solving for maximally predictive coarse-grained variables. Our method follows Gökmen et al. [2021] and essentially consists of a variational solution to a type of information bottleneck (IB) problem. Suppose $X \in \mathcal{X}$ is the current state of the retinal code and Y is the future, as in Fig. 4.3A. We seek coarse-graining maps $\Lambda : \mathcal{X} \to \mathbb{R}^K$ that preserve a maximal amount of mutual information between $Z = \Lambda(X)$ and Y. The critic function $f(z, y)$ which parameterizes the variational estimate $q(z, y)$ of the true joint distribution $p(z, y)$ is optimized simultaneously. We therefore have

$$\Lambda, f = \underset{\Lambda', f'}{\arg\max}\, I_{\text{NCE}}(\Lambda'(X); Y)[f'] \tag{4.4}$$

where K is fixed during optimization. In this work, we restrict our search to linear maps Λ and denote the compressed representation $Z = \Lambda \cdot X$. We refer to these coarse-grained degrees of freedom as "meta-neurons".

Some generic statements can be made about the behavior of $I_{\text{NCE}}(Z; Y)$ in solutions to (4.4), and these provide reasonable guidelines to interpreting the solutions of (4.4). First, suppose that $I_{\text{NCE}}(Z; Y)$ has been properly maximized over f is a good estimate of $I(Z; Y)$ for a given Λ. Because of the data processing inequality, $I(Z; Y) \leq I(X; Y)$, which we refer to as the total predictive information. When K is increased, $I(Z; Y)$ cannot decrease, and when $K = \dim\mathcal{X}$, the coarse-graining map Λ becomes invertible, leading to $I(Z; Y) = I(X; Y)$.

Together, these constraints tell us that $I(Z;Y)$ is a monotonically increasing function of K which saturates at $I(X;Y)$. The shape of this dependence beyond these basic requirements is the object of study here. In particular, to quantify compressibility of the retinal code, we examine how quickly $I(Z;Y)$ approaches its upper bound as K is increased from the fully-compressed limit.

Across stimuli, we find that for X representing all neurons in a single time bin and Y representing all neurons at $\tau = \Delta t$ in the future, only about $K = 8$ meta-neurons are required to capture all of the available predictive information. This is depicted in Fig 4.3B and 4.3C, where we additionally show $I_{\text{NCE}}(Z;Y)/I_{NCE}(X;Y)$ for $K = 1, 2, 4$, and 8 meta-neurons. We also perform the same measurements across prediction intervals τ and find that this result persists out to τ of at least 500 ms. Recall that $\dim \mathcal{X} = 93$, meaning only about a tenth of these dimensions actually contain information about future responses at any given time interval. One should note, however, that since each prediction interval τ actually represents a distinctly defined pair (X,Y) of input and relevance variables, these 8 predictive directions in state space could actually vary with respect to τ.

We directly investigate this dependence of relevant features on prediction interval, as depicted in Fig 4.3D and 4.3E. First we pick a "training" time interval, say for example $\tau_{\text{train}} = \tau_1 = 83$ ms, and solve (4.4), yielding (Λ_1, θ_1). The coarse-graining map Λ_1 represents the K linear features of X which maximize the mutual information between $Z_1 = \Lambda_1 \cdot X$ and $Y(\tau_1) = X_{t+\tau_1}$. Meanwhile, θ_1 represents part of an estimate of $p(z_1, y)$. With the features Λ_1 fixed, we then optimize $I_{\text{NCE}}(Z_1; Y(\tau))(\theta)$ over θ for a range of prediction intervals τ. This gives us an estimate of $I(Z_1; Y(\tau))$, or the information between meta-neurons trained to predict at interval τ_1 and the state at all other intervals τ. In Fig 4.3E, we show the dependence of $I_{\text{NCE}}(Z; Y(\tau))/I_{\text{NCE}}(X; Y(\tau))$ on τ for two different training intervals, $\tau_1 = 83$ ms and $\tau_2 = 350$ ms, and for $K = 1, 2, 4$, and 8 meta-neurons.

For both training intervals τ_1 and τ_2, the predictive features generalize well, but not

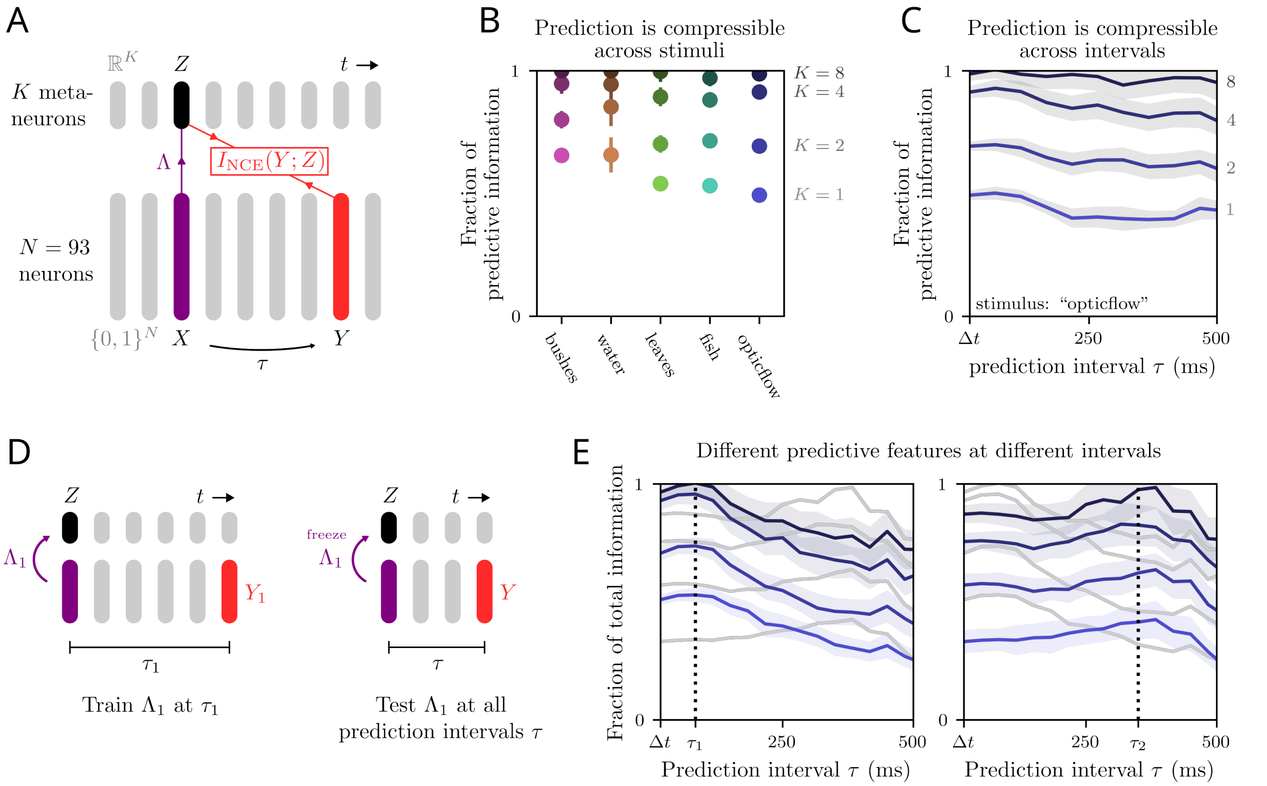

Figure 4.3: Predictive information in the retinal code is compressible, regardless of prediction interval. However, linear predictive features depend on prediction interval. A. Schematic of computation. Input domain X is coarse-grained linearly into a compressed, K-dimensional representation $Z = \Lambda \cdot X$. The optimal coarse-graining Λ is found by maximizing the InfoNCE lower bound (4.3). A neural network ansatz for the conditional distribution $q(z|y)$ is also optimized over simultaneously. B. Fraction of total predictive information $I_{\text{NCE}}(Z;Y)/I_{\text{NCE}}(X;Y)$ for a single time bin prediction interval $\tau = \Delta t$. ($Y_t = X_{t+\tau}$ here). This is provided for each stimulus at four values of compressed representation dimensionality $K = 1, 2, 4,$ and 8. C. Fraction of total predictive information $I_{\text{NCE}}(Z;Y)/I_{\text{NCE}}(X;Y)$ as a function of prediction interval τ for the "opticflow" stimulus. D. Example test of predictive feature generalization. We first find optimal features Λ_1 at prediction interval τ_1, then measure $I_{\text{NCE}}(\Lambda_1(X);Y(\tau))$ for τ across the whole range of timescales. E. Predictive feature generalization for compressed dimensionalities $K = 1, 2, 4, 8$, at two different training timescales τ_1 and τ_2. All values are the fraction of total available predictive information as estimated by I_{NCE}.

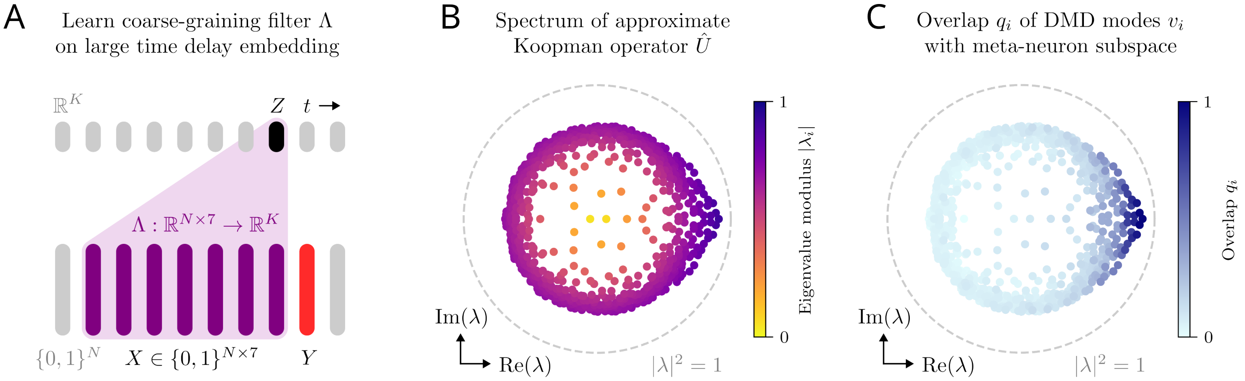

Figure 4.4: Meta-neuron features are spanned by leading modes from dynamic mode decomposition. A. Schematic depicting compression of time-delay-embedded input variable X. We take a time delay embedding with 7 time steps, leading to a state X that is a $93 \times 7 = 651$-entry binary array. This is transformed linearly into $K = 20$ meta-neurons Z_a by filters Λ_a, with $a = 1, \ldots, 20$. That is, $Z = \Lambda \cdot X$. B. Using the same time delay embedding, we construct an approximate Koopman operator $\hat{U}$ which is the least-squares solution to the linear regression problem $X_{t+\Delta t} = \hat{U} X_t + \varepsilon_t$. B depicts eigenvalues λ_i of $\hat{U}$. Coloring represents $|\lambda_i|$. Dotted line is unit circle. C. Overlap q_i of DMD modes v_i with subspace spanned by meta-neuron features.

perfectly, across testing intervals. When the testing interval is equal to the training interval, we see a peak in predictive information carried by these specialized meta-neurons. While it appears to be the case that only 8 meta-neurons trained at any timescale successfully encode a the majority of predictive information at all intervals, the two examples we give reveal that these compressions are highly sub-optimal. For example, 8 meta-neurons trained to predict $Y(\tau_2)$ do worse at predicting $Y(\tau_1)$ than only 4 meta-neurons trained to predict at that interval. Together, these results suggest that retinal responses for complex naturalistic stimuli encode both general and timescale-specific predictive information, and that these features could be disentangled by downstream neurons solving different prediction problems.

4.6 Meta-neurons carry long-timescale predictive information

We have so far shown that retinal responses to naturalistic stimuli have predictable long-timescale features, and that these features are compressible by linear coarse-graining. Moreover, by predicting future outputs at one timescale, neurons downstream of the retina can also recover information about the future at other timescales, indicating that meta-neurons encode generalized predictive features. On the other hand, at least in our linear encoding scheme, this occurs with slightly sub-optimal fidelity and efficiency due to the presence of interval-specific information. Now we ask, what are these features that the meta-neurons encode, and can they be interpreted without recourse to the specifics of the stimuli we are investigating?

Recent work has demonstrated connections between predictive IB and dynamic mode decomposition (DMD) (Schmitt et al. [2023]). In the latter, one is interested in finding a linear time evolution operator for some set of observables $\{f_a(x_t)\}$ for a system with underlying state x_t. For a deterministic system, some such set will admit linear dynamics of the form $f_b(x_{t+1}) = \sum_a U_{ba} f_a(x_t)$, and so the object of study becomes this time evolution operator U, also known as the Koopman operator. In practice, one rarely knows *a priori* which set of observables will satisfy this construction, and the system may be nondeterministic, so some set $\{\hat{f}_a\}$ is chosen and an estimate $\hat{U}$ of the Koopman operator is found by solving the linear regression problem $\hat{f}_b(x_{t+1}) = \hat{U}_{ba}\hat{f}_a(x_t) + \varepsilon_t$.

Here, inspired by the fact that predictive information in retinal responses can be captured by linear coarse-graining, we study the data using DMD with a set of observables consisting only of a time-delay-embedding (TDE) of the neural state. This allows us to examine the meta-neurons in terms of the collective modes identified by DMD. To this end, we choose as our observable $\hat{f}(X_t) = X_t$, which consists of all binarized neural responses in the 7 time bins leading up to time t. Explicitly, $X_t \in \{0,1\}^{7\times 93}$ since there are 93 neurons. The least-squares estimate of the Koopman operator in this observable basis is then given by $\hat{U} = \mathbf{X_2}\mathbf{X_1}^T(\mathbf{X_1}\mathbf{X_1}^T)^{-1}$ with $\mathbf{X}_1 = [X_1, \ldots, X_{T-1}]$ and $\mathbf{X}_2 = [X_2, \ldots, X_T]$.

The spectrum of $\hat{U}$ is shown in the complex plane in Fig 4.4B. For reference, the unit circle $|\lambda| = 1$ is provided by the dotted grey line. Because U is a time evolution operator, U^t

represents evolution by a time t, and eigenvalues λ_i with larger modulus therefore contribute more at larger times. In the spectrum of $\hat{U}$, there is a bulk part which is mostly distributed around a characteristic radius, as well as a set of eigenvalues bunched up past the right edge of this bulk. These latter eigenvalues correspond to the long-timescale modes of X_t, and should encode the predictive information, in principle.

How do meta-neuron features compare to the long-timescale modes as discovered by DMD? Our basic approach is to find optimally predictive features in the time-delay-embedded state, compare these features to the eigenmodes of the approximate Koopman operator, and finally connect these features in this TDE state space back to optimal features of individual time steps, such as those we have discussed so far. As depicted in 4.4A, we solve for a linear coarse-graining of the TDE state $X_t \in \{0,1\}^{7\times 93}$ which optimally predicts the state Y in the single following time step at a fixed number of K features. To see how this choice of relevance variable relates to the construction of the DMD problem, note that in DMD we essentially perform linear regression between the TDE state X_t and the same state propagated forward by one time step, X_{t+1}. In this propagated state, all but the leading time step are deterministically given by entries in X_t, and so all of the nontrivial coefficients in $\hat{U}$ are couplings from X_t to Y.

First, we examine the optimal meta-neuron features in the TDE state in terms of DMD modes. Let Λ denote the optimal linear coarse-graining map, represented as a $K \times 651$ matrix, where we take $K = 20$ (this captures $> 95\%$ of predictive information). Because any invertible transformation of Z leaves the mutual information invariant, all invertible linear operations $\Lambda \to L\Lambda$ yield equally good coarse-graining performance. What is important, then, is the rowspace of Λ and how it relates to the eigenvectors v_i of $\hat{U}$. Here we study this by constructing an orthogonal projection operator out of Λ and looking at the magnitude of each DMD eigenvector in the subspace it projects to. Explicitly, the projection is given by

$P_\Lambda = \Lambda^T(\Lambda\Lambda^T)^{-1}\Lambda$. Then, for $\hat{U}v_i = \lambda_i v_i$ and $v_i^\dagger v_i = 1$,

$$q_i = v_i^\dagger P_\Lambda v_i \in [0,1]\,.$$

This overlap quantifies how much of the i^{th} DMD mode is spanned by the meta-neuron features. For example, if a mode has an overlap $q = 1$, then it can be read out of the meta-neuron state Z with perfect fidelity, whereas an overlap of zero indicates that the corresponding mode is totally in the nullspace of Λ and does not influence Z. In Fig 4.4C, we depict q_i by the darkness of a point on the complex plane located at λ_i, and contours of the approximate density of λ_i are given for visual reference. We observe that the greatest overlap with meta-neuron features occurs in the longest timescale modes. This is consistent with the findings of (Schmitt et al. [2023]), in which it is shown that in systems with long-timescale eigenvalues that are well-separated from a bulk essential spectrum, the leading DMD eigenvalues correspond to the leading features learned by predictive IB. Here, although this separation is evidently absent, we still see that the IB features learned are spanned by these leading modes.

Next, we discuss how these meta-neuron features of the time-delay embedded state relate to those found by coarse-graining from individual time steps (non-TDE). In brief, TDE features are composed of single-time bin features, in the sense that the latter can be obtained by looking at entries in the former. More precisely, the optimal features at a given time step within the TDE state very nearly match the optimal features for the corresponding single-time step IB problem at the corresponding interval. For example, consider the single-time step problem with $\tau = 3\Delta t$. If we coarse-grain down to $K = 1$ meta-neuron, we will obtain some optimal vector $\Lambda_3 \in \mathbb{R}^{1\times 93}$ representing the corresponding feature. Now let $\Lambda = [\tilde{\Lambda}_1, \tilde{\Lambda}_2, \ldots, \tilde{\Lambda}_7] \in \mathbb{R}^{1\times 651}$ be the optimal feature obtained in the IB problem depicted in 4.4A with $K = 1$. The crucial connection is that $\Lambda_3 \approx a_3\tilde{\Lambda}_3$ for a some constant. Beyond $\tau = 3\Delta t$, we find generally that $\tilde{\Lambda}_\tau \approx a_\tau \Lambda_\tau$. For more general K, direct comparison is a bit more subtle, as reparameterization invariance in the objective function allows meta-neurons to mix. Moreover, variations in features only matter to the extent that they change predictive information. To address this, we built solutions to the time delay embedded IB problem out of single time step features and confirmed that they nearly maximize $I_{NCE}(Z;Y)$.

The overlap measurements may provide some explanation as to how meta-neurons trained to predict short time intervals generalize well to prediction at longer timescales. As discussed, the eigenvalues λ_i with the largest modulus are those that persist under time evolution, since $\lambda_i \to \lambda_i^t$. Because the meta-neuron features are spanned by these leading modes, a similar statement in the reverse direction is also true, namely that information about the leading modes can be read out of the meta-neurons. On the other hand, we know that with our data and method, different timescales require different features, and so these two pictures of DMD and predictive IB should not agree exactly here. Indeed, they do not; many of the bulk DMD modes have some nonzero but small overlap with the meta-neuron features, and when restricting to smaller K, the first features learned are not always the longest timescale modes. A number of changes in the analysis might bring about better agreement, for example by taking even larger time delay embeddings, expanding the set of DMD observables to contain nonlinearities, and allowing for nonlinearity in the meta-neuron coarse-graining map.

4.7 Predictive information is collectively encoded

From the perspective of a downstream predictor, how important are collective effects in predicting retinal activity? We address this by examining how predictive information in a population depends on how many neurons are included in it. If neurons are independent or inter-neuron correlations are irrelevant to prediction, the total predictability of a population is simply the sum of individual neuron contributions. On the other hand, if correlations between neurons exist in a way that can be leveraged by downstream predictors, the total predictive information in a population will differ from this independent component-wise sum.

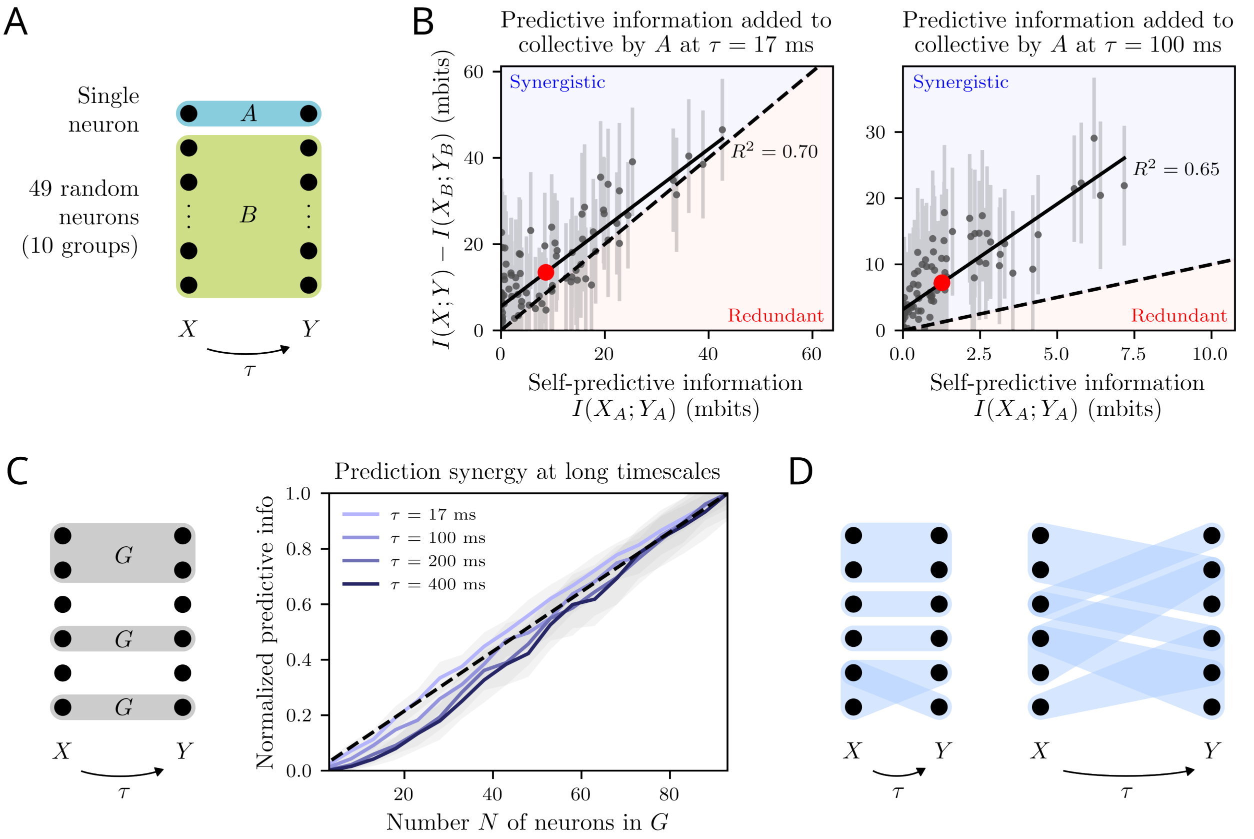

Figure 4.5: The dependence of predictive information on neuron subset suggests that predictive information is collectively encoded. A. Depiction of single neuron contribution test. For each neuron A we randomly sample 10 groups B of 49 neurons each (A and B are disjoint) then estimate $I(X_A;Y_A)$, $I_{\mathrm{NCE}}(X_B;Y_B)$, and $I_{\mathrm{NCE}}(X;Y)$, with $X = X_{A \cup B}$. We perform this test at $\tau = \Delta t$ and at $\tau = 100$ ms. B. Contribution of neuron A to predictability of a collective $A \cup B$, averaged over B, as a function of A self-information. Deviation from the dashed unity line indicates the presence of collective effects. The red dot is placed at the average over all points and the difference of its components indicates average prediction synergy for the division into groups sized 1 and 49 neurons. Collective effects appear to become much more important for prediction at the longer interval. C. Scaling of predictive information in 20 random groups at each size N for four different prediction intervals. Each curve is normalized by its largest value to compare shapes. A geometric argument shows that a concave-up curve implies positive prediction synergy after averaging over group choices. D. Schematic depiction of hypothesized correlation structure. At short timescales, prediction synergy is positive for small groups and neutral or redundant in large groups, suggesting greater autocorrelation content, less widespread cross-correlations, and effective independence between groups past some scale. Meanwhile, neurons which are highly self-predictive at long timescales also exhibit strong prediction synergy in random collectives, which might be explained by a proliferation of strong cross-correlations.

In our data, we should expect to see something more like this latter case. One reason is that naturalistic stimuli induce correlations across timescales, and these long-timescale features are associated with longer length-scales (Salisbury and Palmer [2015]). That is, because we expect the spatiotemporal correlations present in the stimulus to couple cell responses across the retina, and because these features carry predictive information in the stimulus, we expect predictive information to be collectively encoded.

As an initial probe of collective effects, we examine the contributions of individual neurons to the predictability of random groups. Specifically, we measure the relationship between a neuron's information about itself (self-information) and the change in information it provides to a collective upon its inclusion (information contribution). For each neuron A in the 93 cell population, we compute its self-information directly from an empirical histogram and sample ten random groups of 49 neurons B (Fig 4.5A). We then use InfoNCE to estimate $I(X_B; Y_B)$, the predictive information in these groups without A, as well as $I(X, Y)$, the predictive information of all neurons in $A \cup B$ at the given interval. The difference $\delta_{A|B} = I(X;Y) - I(X_B; Y_B)$, or the information contribution due to A, represents the actual increase in total predictability to a downstream predictor upon including A in the collective.

A particularly striking feature of figure 4.5B is the existence of neurons with small or vanishing self-information relative to significant, nonzero information contribution in random collectives. Such neurons are individually useless at the given timescale for downstream predictors; their utility only arises when read out within some group. While these neurons can be found at both prediction intervals shown, their effect on the mean values $(\langle I(X_A; Y_A)\rangle_A, \left\langle \delta_{A|B} \right\rangle_{B,A})$ (large red circles) seems to be larger in the short timescale case. These neurons may account for the fact for the fact that on the whole, the 93 cell population has about 1.2 bits of predictive information at 17 ms, or about 13 mbits/neuron even though an individual neuron has 8 mbits of self-predictive information on average. In other words, it seems that correlations between weakly-firing, low self-information neurons and their environments cause the scaling of total predictive information to exceed the independent-neuron estimate.

Neurons which have more information about their own future tend to contribute more information to the future of collectives, as indicated by positive linear correlations in figure 4.5B. This is perhaps unsurprising, since if neurons were totally independent, one would observe perfect correlation with a slope of unity (dashed black line). At a short prediction interval, we observe a correlation which is close to unity but with a slight positive bias in $\delta_{A|B}$, indicating a kind of synergy. This idea will be formalized shortly. While low-self information neurons contribute synergistically at both timescales shown, the overall synergistic deviation from unity is much stronger at the prediction timescale of 100 ms. From the downstream perspective, the utility of a highly active neuron at 17 ms seems to be limited to predicting its own future, while for prediction at 100 ms, the most self-informative neurons seem to be much more useful in random collectives. These features suggest a shift towards collective encoding of predictive information as prediction interval is increased.

In this analysis, the choice of 50-neuron collectives was arbitrary, as was the decision to look at differences due to single-neuron changes. To extend our reasoning, we introduce prediction synergy, which generalizes $\delta_{A|B} - I(X_A; Y_A)$ to groups A of arbitrary size. This quantifies the increase (or decrease) in predictive information obtained by a downstream predictor when pooling inputs from subsets A and B, as compared to considering them independently. For two disjoint sets A and B of neurons, and with X denoting present states and Y denoting future states, the prediction synergy is given by

$$S(A:B) = I(X;Y) - I(X_A;Y_A) - I(X_B;Y_B)\,. \tag{4.5}$$

What is the biological or functional interpretation of prediction synergy? Supposing that a group A has low self-information but high contribution to various collectives $A \cup B$, that group A is only useful to downstream predictors within some greater context. According to our data, benefits to predictability seem to be robust to the choice of context. In the single-neuron case, strong prediction synergy arises at long prediction intervals when we consider random environments B, which suggests that the benefit of pooling inputs is robust to this choice. This does not preclude a scenario in which the maximal value $\max_B S(A : B)$ far exceeds $\langle S(A : B)\rangle_B$, and it could be interesting to search for such maximal pairings.

Now we turn our attention to prediction synergy with larger subset sizes at various prediction intervals. As in the single-neuron case, we are less interested with optimizing information over the groups of neurons which are selected, and more in understanding how important collective effects are when predicting random groups at different scales. To probe this, we make use of a geometric interpretation of prediction synergy, which relates it to the convexity of predictive information scaling with group size N. Indeed, part of our motivation with studying prediction synergy as opposed to the standard definition of synergy is that it can extracted from scaling measurements in this way. Consider first a scenario in which predictive information scales linearly. When averaged over random groups G of size N, we can write $\langle I(X; Y)\rangle_G = NI'$ with I' a constant. To compute prediction synergy, we separate G into two subsets A and B with N_A and N_B neurons respectively, such that $N_A + N_B = N$. Then, averaging over group choices with these constraints we arrive at $\langle S(A : B)\rangle_{G,A,B} = NI' - (N_A + N_B)I' = 0$. Evidently, linear scaling of predictive information averaged across random groups leads to a net zero average prediction synergy. By similar reasoning, if information is concave up from 1 to N neurons, then the average prediction synergy in groups of up to size N is positive. Concave down scaling indicates prediction redundancy. In Fig 4.5C, we show information scaling curves up to 93 neurons for four different prediction intervals. Each curve is normalized by its maximal value in order to facilitate a comparison of their shapes. A clear trend is revealed in which information scaling becomes more strongly concave up as prediction interval is increased, and over an increasing range of group sizes. Hence, prediction synergy, averaged over groups and their subdivisions, becomes relatively much stronger at large prediction intervals.

The dependence of prediction synergy on group size and timescale reveals that information is collectively encoded, with collective effects becoming more important at large intervals. At 17 ms, prediction synergy is indeed present, but only for subdivisions with small groups. So, while Fig 4.5C suggests that there is prediction redundancy between large groups, correlations between small groups and their environments can explain why the overall rate of information scaling of 13 mbits/neuron exceeds the independent estimate of 8 mbits/neuron. This is directly quantified in Fig 4.5B and represented as the large red dot, lying at approximately $(8, 13)$ mbits/neuron. On the other hand, the N-scaling curves for $\tau = 200$ or 400 ms in 4.5C are concave up over a larger range of N, indicating prediction synergy between larger groups of neurons.

Finally, we propose that this trend towards increasingly collective encoding at longer timescales is reflected in correlation structure. We depict this idea schematically in 4.5D, with blue connections representing correlations. To make this description more precise, we will quantify the contributions of single-neuron and cross-correlations to predictive information in the following section. Before moving on to these analyses, we briefly discuss the connection between correlation structure and prediction synergy in a Gaussian toy model.

The model consists of four degrees of freedom: two in the present and two in the future, all jointly Gaussian. These are divided into two subsystems A and B. We parameterize the distribution by pairwise correlations and study the prediction synergy in three limiting regimes. First, complete independence between subsystems leads to zero prediction synergy or redundancy. Second, when system B is a copy of A, one finds complete redundancy. Finally, synergy is maximized by only keeping cross correlations, in which case self-predictive info goes to zero, regardless of the total amount of predictive information present. A more complete analysis with the explicit expression for prediction synergy for any correlation structure are given in Sec. 4.10. While our toy model should not be taken seriously as a description of correlations in neural data, we expect that these basic motifs of redundancy, independence, and cross-correlation should have analogous effects on overall prediction synergy when they are present.

4.8 How do correlations comprise predictive information?

In the last section, we presented evidence that predictive information in retinal responses is encoded collectively, in that the downstream prediction task requires taking into account correlations between neurons. In order to formalize and quantify this, we introduced prediction synergy. At short timescales, small groups of neurons have prediction synergy with random, large groups, and at longer timescales, average prediction synergy is generically large regardless of group sizes. Since positive prediction synergy in a toy model can be associated with strong cross-correlations, we hypothesized (Fig 4.5A) that the correlation structure responsible for predictability changes depending on timescale. To be precise, at short timescales, where prediction synergy only occurs in small groups, insight from our toy model suggests that autocorrelations should be important for prediction, while at longer timescales, cross-correlations should carry the majority of predictive information.

The retina responds to stimuli in a noisy way due to fluctuations which are inherent to its machinery. For physiological reasons, these fluctuations can be correlated across cells and time, leading to what are called "noise-correlations" which appear in addition to stimulus-induced correlations. To capture the contribution of noise correlations to predictability, we make use of the trial structure of the dataset (Figs 4.1E and 4.6A). As discussed in Sec 4.3.1, each stimulus was shown to the salamander retina around 80 times (in the case of the "opticflow" stimulus, 85 times). During each of these repeated trials, the stimulus driving responses in the retina stays exactly the same, so any variation in the responses comes from noisy outcomes of the internal dynamics of the retina. If we summarize all stimulus features and repeatable aspects of internal retinal dynamics by a time index t, this variation

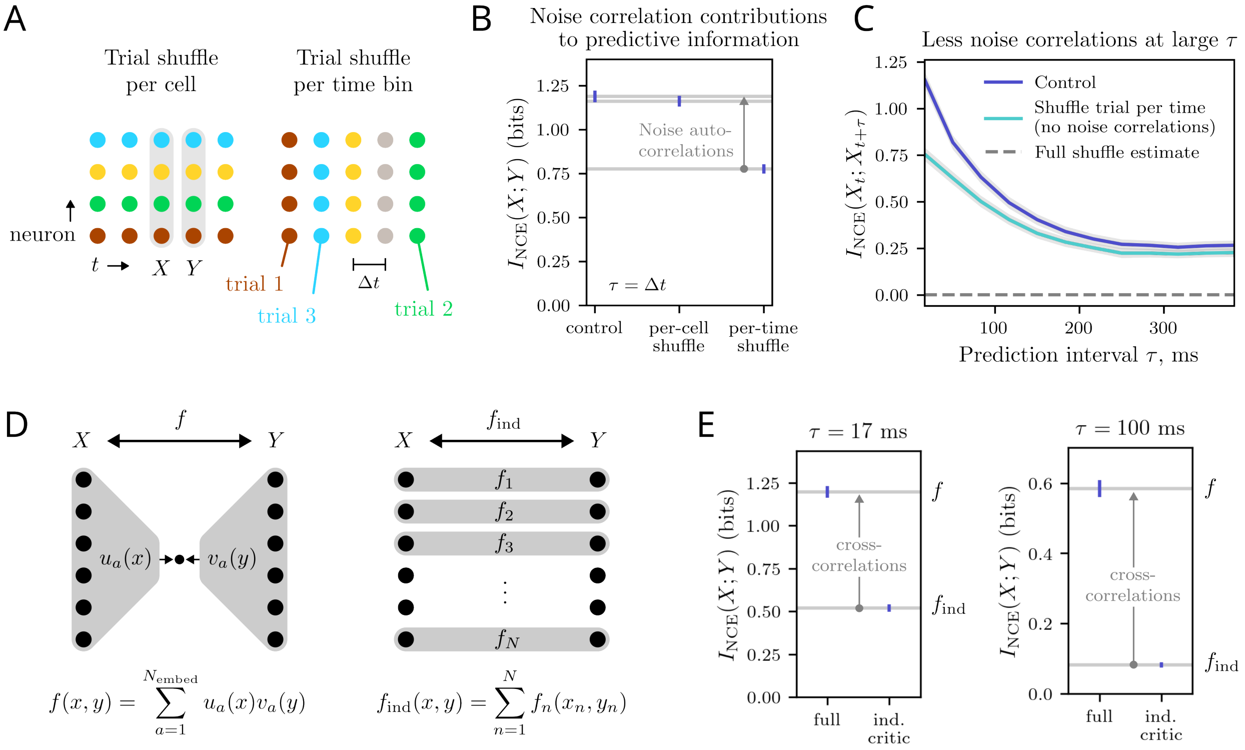

Figure 4.6: Cross-correlations contribute significantly to predictability at short intervals and dominate at large intervals, while autocorrelations only significantly contribute to predictive information at short intervals. A. Each stimulus (one of five 20 s naturalistic movies) was shown to the retina a total of 80-90 times, and this trial structure allows us to probe noise correlations. To remove noise correlations between neurons we can randomly permute the trial in a way that is consistent in time, but changes based on cell. We can similarly remove noise auto-correlations by permuting the trial differently at each time step. Note that both of these methods only change the trial orderings, not the stimulus. B. Predictive information from opticflow stimulus at a single time step, with different noise correlations removed. Noise autocorrelations carry more predictive information than equal-time correlations. Given the limited number of trials it is not possible to resolve noise cross-correlations. C. Predictive information as a function of interval, with and without noise correlations. At longer intervals, noise autocorrelations matter less. D. By removing interactions between neurons in the critic function, we estimate the contributions of cross-correlations at all orders to predictive information. E. Difference in InfoNCE bound between between fully expressive and independent critic functions, at two different prediction intervals. This difference represents the information discarded by a downstream predictor when it ignores cross correlations at all orders.

in response of some collection of neurons under identical, repeated stimulus conditions is captured by a distribution $p(x, y|t)$, where x represents the present state and y represents the future. The whole joint distribution describing statistical dependencies between the present and future, such as we have been analyzing throughout this work, is given by an ergodic average:

$$p(x, y) = \frac{1}{T} \sum_{t=1}^{T} p(x, y|t)$$

One direct way we can probe noise correlations is by removing them from the data artificially, by shuffling. For example, suppose that we denote by $\sigma(n, t, a)$ the state of the n^{th} neuron on the t^{th} time step and in the a^{th} repeat of the stimulus. Then, we could choose for each neuron a different permutation π_n that randomly permutes trials. If we apply this transformation to the data, we get a new dataset $\sigma'(n, t, a) = \sigma(n, t, \pi_n(a))$ which has destroyed part of the noise correlation structure. Then, by doing the same mutual information measurements which we have been doing on these “per-cell” shuffled data, we can estimate the deficits in information due to the absence of these correlations. We also consider “per-time” shuffled trials, which are similar but assign a different permutation to each time step, i.e. $\sigma'(n, t, a) = \sigma(n, t, \pi_t(a))$. These two shuffling methods are depicted graphically in Fig 4.6A.

In Fig 4.6B we show the effects of per-cell and per-time trial shuffling on predictive information at $\tau = 1/60$ s. Strikingly, per-cell trial shuffling does not seem to have a significant impact on predictive information at short timescales. This suggests that noise correlations between neurons do not contribute, and that neurons are conditionally independent given the stimulus. That is, the following statement is empirically true:

$$p(x, y|t) \approx \prod_{n=1}^{N} p(x_n, y_n|t)$$

As a result of this conditional independence, any information carried from the present to future by noise correlations must be comprised of autocorrelations. We measure the magnitude of these effects using the per-time shuffling protocol and give our findings in Fig 4.6C. Out of the roughly 1 total bit of predictive information available at $\tau = 1/60$ s, randomly permuting trial index at every time step removes around 0.35 bits. Explicitly, this measurement tells us 0.35 bits of predictive information are provided by single-neuron noise correlations across time which are not induced by the stimulus, but rather the noisy workings of retinal circuitry. The remaining information must be bound up entirely in stimulus-induced correlations.

For longer prediction intervals, the relative contribution of noise correlations to predictive information begins to fade (Fig 4.6C). This is particularly interesting considering the long timescales present in predictive information (Fig 4.2), since it tells us that these long timescales are actually due to correlations induced by the stimulus. Moreover, this may provide some insight into the phenomenon depicted in Fig 4.3E, wherein the relevant predictive features depend on prediction interval. Since noise correlations contribute strongly to predictive information at short intervals, and since these correlations are a single-neuron effect, their disappearance at longer timescales is consistent with our hypothesis that the balance of individual to collective correlations shifts towards collective at late times.

To be clear, the fact that neurons are conditionally independent given the stimulus and their past does not mean that one should think of prediction as being a single-neuron effect. The importance of collective effects to predictability is confirmed by the presence of widespread prediction synergy. This conditional independence given stimulus only refers to noise correlation structure, which is variation in addition to the repeatable, essentially deterministic aspects of how neurons encode stimuli.

To quantify the contribution of cross-correlations to predictive information, one has to account for not only pairwise interactions but also all higher-order relationships, and in principle at all orders beyond linear. Since doing so explicitly would be prohibitively complicated, we instead consider the decline in performance of a model in which all such interactions are disallowed. This approach leverages the fact that InfoNCE is a lower-bound of the Barber-Agakov bound; its optimization can be conceptualized as variational inference within a class of models. As discussed in 4.3.2, our variational estimate for a joint distribution $p(x, y)$ is represented by a so-called critic function $f(x, y)$. This neural network only represents interactions between x and y, and does not encode the marginal distributions. So far we have taken the critic function to be a fully expressive neural network, able to capture any relevant correlations. Now we consider an independent critic function ansatz, given as a sum of neural networks which each take input only from the state of a single neuron (x_n for the present and y_n for the future):

$$f_{\text{ind}}(x, y) = \sum_{n=1}^{N} f_n(x_n, y_n)$$

In the context of the variational estimate $q(x, y)$ to the true distribution $p(x, y)$, we have

$$q(y_1, \ldots, y_N | x_1, \ldots, x_N) = \frac{p(y_1, \ldots, y_N)}{Z(x_1, \ldots, x_N)} \prod_{n=1}^{N} \exp(f_n(x_n, y_n))$$

In essence, this ansatz models a population where all equal time correlations and auto-correlations are preserved (the marginal distributions $p(x)$ and $p(y)$ are empirical, and are unchanged), but in which each neuron can only directly inform its own future. In other words, all interactions between different neurons are left out of the model. Our two anzates are depicted graphically in Fig 4.6D.

In Fig 4.6E, we provide the estimates for $I_{\text{NCE}}(X; Y)$ obtained with f_{ind} at prediction intervals $\tau = \Delta t$ and $\tau = 100$ ms, with X and Y representing the whole neural population under opticflow stimulus. For comparison, $I_{\text{NCE}}(X; Y)$ evaluated using a fully expressive critic ansatz f is also given. Because the independent model value is not maximal, we cannot directly interperet it as an accurate estimate of mutual information between any two sets of variables, unlike in the case of measurements performed on trial-shuffled data, in which the actual joint distribution was changed. However, we can still find an information theoretic interpretation of this quantity. Consider the difference between $I_{\text{NCE}}(X;Y)[f] - I_{\text{NCE}}(X;Y)[f_{\text{ind}}]$. After optimizing, we can estimate $I_{\text{NCE}}(X;Y)[f] \approx I(X;Y)$ and for the independent critic estimate we can approximate $I_{\text{NCE}}(X;Y)[f_{\text{ind}}] \approx I_{\text{BA}}(X;Y)[f_{\text{ind}}]$. Therefore

$$I_{\text{NCE}}(X;Y)[f] - I_{\text{NCE}}(X;Y)[f_{\text{ind}}] \approx \langle D_{\text{KL}}[p(y|x)||q(y|x)]\rangle_x \ .$$

This difference in performance $\langle D\rangle$ represents the additional information about Y which can be extracted from X upon accounting for some missing correlations. This can be seen considering $q(y|x)$ through the lens of the maximum entropy method (Jaynes [1957]). Once trained, the independent critic takes the form of an exponential family which constrains all autocorrelations at all orders, but has the maximum entropy under this constraint. If cross-correlations are present in data that are not captured by this independent model, then the entropy of q can be further reduced by at most $\langle D\rangle$ by including new terms that reproduce these correlations. Hence, if a population of downstream predictors has a sub-optimal model q of the upstream activity, it can reduce its uncertainty about future responses by a quantity of $\langle D\rangle$ by taking into account correlations between neurons. At $\tau = \Delta t$, observing cross-correlations provides about 0.7 bits, or 58% of the available predictive information, while at $\tau = 100$ ms cross-correlations provide about 0.5 bits, or 84% of the total. This test directly confirms our hypothesis about the relative importance of collective effects as a function of prediction interval. In principle, a refinement of this method could examine the contributions of cross-correlations order-by-order to give a detailed picture of how prediction synergy arises at various timescales.

4.9 Discussion

The ubiquitous problem of prediction can be formalized in a general and powerful way using information theory. Applications of the information bottleneck and related frameworks in particular have yielded deep insights into the structure of predictive computation and its role in biology (Palmer et al. [2015], Chalk et al. [2018], Wang et al. [2022]). In the early vertebrate visual processing stream, IB was used to confirm that the retina optimally predicts its inputs, as it must for the organism to react quickly enough in dangerous situations. Further work examined the possibility that optimal prediction continues as a motif at successive layers of visual processing, so that for example neurons receiving projections from RGCs might optimally encode information about the future states of their inputs (Sederberg et al. [2018]). However, these investigations have been limited to small sets of neurons, due to the intractability of computing information quantities involving joint distributions from larger populations. Here we accessed this many-cell regime by leveraging a variational inference technique which maximizes a lower bound on mutual information and thereby provides an estimate. By simultaneously compressing the neural state and searching over the retained features during variational inference, this method also yields a solution to an IB-like problem. Using this, we solve for the collective variables in a population of 93 RGCs which are optimally predictive of future outputs. We find that all predictive information in this population is encoded collectively in a few delocalized linear features. Overall, our results lend support to the idea that downstream neurons may be able to perform optimal prediction of retinal outputs at the population level. We moreover shed light on basic properties of the neural code under complex naturalistic stimulus, further providing a basis for understanding the downstream prediction computation in realistic scenarios.

In a 93 cell population, even once activity is binarized, the space of possible states is enormous. Fortunately, there are several simple reasons to expect that a good coarse-graining scheme exists. First, since the neural code is sparse—neurons are much more likely to be silent—many states will never occur. Next, neural responses are correlated. These are correlations induced both by stimulus and by physiology, and both allow us to make statements about what some neurons are likely to do given what some others are doing. Finally, the code must ultimately be learnable, otherwise the brain could never decode visual signals. A number of studies have investigated coarse-graining in the retina, and all describe procedures which reduce the state space while preserving relevant features, though the definition of relevance varies somewhat between them. Our main contribution is to demonstrate a compression scheme which preserves predictive information in a relatively very large population (93 RGCs). This allows us to take large-scale collective effects into account and more accurately represent the problem presented to downstream neurons. By varying the target prediction interval, we find that compressibility extends to long timescales (500 ms), and that there are both generalized and timescale-specific predictive features present. Regardless of timescale, these features are delocalized, and capture information which is collectively encoded.

From the perspective of predictive downstream neurons, is it better to treat signals from neurons individually or collectively? We find that individual effects contribute most at short timescales, and collective effects are crucial at all intervals. First, we quantify the prediction synergy between different subsets of neurons. Single neurons almost always provide more predictive information to a collective than they have about themselves, and at moderate to long prediction intervals, we see finite prediction synergy between large subsets. This means that that the predictive information in two combined subsets is greater the sum of predictive information in those subsets separately. In order to see further into the future, downstream neurons therefore need to incorporate inputs from more neurons. Making use of the repeated stimulus structure of our dataset, we find that equal-time cell-cell noise correlations do not significantly contribute to predictive information. Moreover, noise autocorrelations are significant at short timescales, then decrease in importance as interval is increased. We also directly evaluate the information contributions of interactions between neurons at all orders by restricting the variational ansatz for the present-future conditional distribution to single-neuron effects. The difference in performance between this restricted model and a fully expressive one reveals that even at the shortest prediction interval, roughly half of the predictive information received by a downstream neuron comes from correlations between different RGCs. We conclude that although optimal prediction of retinal outputs at long timescales can be done with only a few linear features, these features need to be collective.

By variationally solving a predictive information bottleneck problem on data taken from vertebrate retina, we have demonstrated that downstream prediction of a large population RGCs is plausible. If applied to data from these downstream neurons, these analyses could also reveal their performance relative to the optimum. Further work could constrain the coarse-graining maps and critic ansatz to a family of interpretable, mechanistic models of retina and downstream neurons and reveal how predictive computation might be enabled or hindered by physiology, similarly to (Wang et al. [2022]). More generally, given the ubiquitous importance of prediction as a biological function, we also anticipate that this method could be useful in finding computationally relevant coarse-graining schemes in other complex and biological systems.

4.10 Calculation details

4.10.1 InfoNCE implementation details

As discussed in the introduction, most of our mutual information estimates are given by the InfoNCE lower bound

$$I_{\mathrm{NCE}}(X;Y)(\theta) = \frac{1}{B}\left\langle \sum_{j=1}^{B} f(x_j, y_j|\theta) - \log\left(B^{-1} \sum_{i=1}^{B} e^{f(x_i, y_j|\theta)} \right) \right\rangle_{\prod_{i=1}^{B} p(x_i, y_i)},$$

Where the "critic function" $f(x,y|\theta)$ takes the form

$$f(x,y|\theta) = \sum_{a=1}^{N_{\text{embed}}} u_a(x|\theta)v_a(y|\theta)\,.$$

Here, v_a and u_a are feed forward multilayer perceptron (MLP) neural networks and θ are all of the model parameters. Both u and v have 2 hidden layers with 32 neurons, and output to the embedding dimension of size $N_{\text{embed}} = 30$. Hidden layers have a tanh nonlinearity. The input dimensions depend on the specific calculation. For example, when coarse-graining to $K = 4$ meta-neurons to predict the whole population state in a single future time bin, we would have $u : \mathbb{R}^4 \to \mathbb{R}^{30}$ and $v : \mathbb{R}^{93} \to \mathbb{R}^{30}$.

To train, we used the Adam optimizer with a learning rate of .009, batch sizes of $B = 800$, and held out half the data for testing. All reported mutual information estimates are from the held-out test sets. We explored effects of dropout on weights within these MLPs but found that early stopping with no dropout was the most effective regularization. We chose to stop training after each data point in the test set had been used by the optimizer 16 times. In both testing and training, the first 20 frames (1/3 of a second) from each response set were removed as these contain transients from stimulus switching. For testing, we constructed 500 batches of size $B = 500$. Typically, this whole train/test procedure would be done on the order of 10's of times, producing means across batches, training initializations, and random train/test designations. Uncertainty for mutual information estimates incorporated variance both due to different initializations as well as variance across batches.

4.10.2 Prediction synergy toy model

To understand how correlation structure may affect prediction synergy, we briefly consider a Gaussian toy model. Two sub-systems A and B have "present" states $x = (x_A, x_B)$, and future states (y_A, y_B). Because all variables are all jointly Gaussian, we can specify all parameters by choosing first- and second-order correlation functions. We choose all means to be zero, that is for $\alpha = A, B$, $\langle x_\alpha \rangle = 0 = \langle y_\alpha \rangle$. Next, we fix the scale of fluctuations so that all $\langle x_\alpha^2 \rangle = \langle y_\alpha^2 \rangle = 1$. With only pairwise correlations between degrees of freedom left, we choose three parameters, a, b, and c, representing autocorrelations, equal-time A-B correlations, and across-time A-B correlations, or cross-correlations, respectively. Note that the only choices of (a, b, c) which are valid are those such that all correlation matrices are positive definite. The expression for prediction synergy in this toy model is

$$S(A:B) = \log \frac{(1-a^2)^2(1-b^2)^2}{((1+b)^2-(a+c)^2)((1-b)^2-(a-c)^2)} \tag{4.6}$$

As a point of reference, note that the predictive information of either subsystem is given by

$$I(X_A; Y_A) = I(X_B; Y_B) = -\log(1-a^2)$$

Where for $\langle [x_A, y_A]^T [x_A, y_A] \rangle$ to be positive definite we must have $a^2 < 1$.

Several limits of 4.6 are especially illuminating. The easiest limit is that of independence between the subsystems, in which $b \to 0$ and $c \to 0$. While this can be seen to yield $S(A : B) = 0$ from the expression above, the deeper reason is that when A and B are statistically independent we must have $I(X;Y) = I(X_A;Y_A) + I(X_B;Y_B)$, and hence $S(A:B) = 0$.

Next, consider a model where B is a direct copy of A. Since x_A has the same relationship with y_A as it does with y_B, we must have $a = c$. Further, x_A must relate to x_B in the same way it relates to itself, so we take $b = 1 - \epsilon$. In the limit $\epsilon \to 0$, this yields $S(A : B) = \log(1-a^2) = -I(X_A;Y_A)$. That is, the prediction synergy is negative, indicating redundancy. Moreover, the extent of this redundancy is given by the total predictive information of a single subsystem, in agreement with our interpretation of this this model as consisting of two redundant copies.

As a final limit, we consider the removal of all correlations except cross-correlations. In this case, $a = 0$ implies that $I(x_A; y_A) = 0$, meaning each subsystem observed alone has no information about its future. Taking $b \to 0$ as well, we find that $S(A : B) = -2\log(1 - c^2)$. From the definition of S, we also see that $S(A : B) = I(X; Y)$. In this limit, all predictive information is encoded collectively, in that observation of both subsystems is required to extract it. It is interesting to note that the expression $-2\log(1 - c^2)$ is the same as in the case of independent systems except under the replacement $c \to a$. The basic reason for this is that this limit also describes two independent subsystems, but x_A should be grouped with y_B and x_B with y_A, meaning our choice of subsystems does not capture this independence.

CHAPTER 5

PRECISION MEASUREMENT OF TRIBOCHARGING IN ACOUSTICALLY LEVITATED SUB-MILLIMETER GRAINS

AGK, Melody X. Lim, Heinrich M. Jaeger

Department of Physics and James Franck Institute, The University of Chicago,
929 E 57th St., Chicago, Illinois 60637, USA

5.1 Abstract

Contact electrification of dielectric grains forms the basis for a myriad of physical phenomena. However, even the basic aspects of collisional charging between grains are still unclear. Here we develop a new experimental method, based on acoustic levitation, which allows us to controllably and repeatedly collide two sub-millimeter grains and measure the evolution of their electric charges. This is therefore the first tribocharging experiment to provide complete electric isolation for the grain-grain system from its surroundings. We use this method to measure collisional charging rates between pairs of grains for three different material combinations: polyethylene-polyethylene, polystyrene-polystyrene, and polystyrene-sulfonated polystyrene. The ability to directly and noninvasively collide particles of different constituent materials, chemical functionality, size, and shape opens the door to detailed studies of collisional charging in granular materials.

5.2 Introduction

In industry, electrostatic charging underpins manufacturing techniques such as powder coating (Bailey [1998]), but can cause catastrophic explosions in the handling of fine powdered materials (Abbasi and Abbasi [2007]). The buildup of charge due to repeated collisions between small particles is thought to be responsible for lightning in volcanic ash clouds (Brook et al. [1974]), the electrification of grains in sand storms (Stow [1969]), and potentially also the very early stages of the formation of planetesimals from interstellar dust (Blum and Wurm [2008], Jungmann et al. [2018]). Several controlling parameters for particle charging have been measured, notably the effects of particle size (Forward et al. [2009], Waitukaitis et al. [2014]), atmospheric conditions and external electric fields (Zhang et al. [2015], Pähtz et al. [2010]), surface hydrophobicity (Lee et al. [2018]), frequency of contact (Shinbrot et al. [2018]), and kinetic energy prior to impact (Poppe et al. [2000]). However, the underlying mechanism for charge exchange remains an area of debate, particularly between insulators, which have very low charge mobility.

Several candidates for the charge carrying species have been suggested, including electrons in trapped surface states (Lowell and Truscott [1986], Liu and Bard [2008], Lacks and Levandovsky [2007]), ions in atomically thin water layers (McCarty and Whitesides [2008], Lee et al. [2018], Harris et al. [2019]), and mechanoradicals produced during contact (Baytekin et al. [2011, 2012]). Elucidating the fundamental mechanism of collisional charging between grains thus calls for systematic, quantitative experiments. The most common experimental approach utilizes drop tests (Mehrotra et al. [2007], Lee et al. [2015], Waitukaitis et al. [2014], Jungmann et al. [2018]), wherein a collection of grains is filmed and studied through the course of a free-fall, or in Faraday cup experiments (LaMarche et al. [2009], Sowinski et al. [2009]) in which only net charges of collections of particles can be measured. These experiments typically involve a large number of particles, which results in both particle-wall and particle-particle collisions, hindering access to the basic physics of a single collision. Recent experimental techniques have made precise measurements of the impact charging of a single submillimeter particle with a fixed substrate (Watanabe et al. [2006], Xie et al. [2016], Lee et al. [2018], Haeberle et al. [2018]). There remains, however, a need to track the evolution of the charging process over repeated, highly controlled collisions between a pair of grains.

Here, we introduce such a method by combining high-speed videography and acoustic levitation, allowing for the contact-free manipulation of a wide variety of constituent materials, particle sizes, and shapes (Lee et al. [2018], Lim et al. [2019b,a]). We dynamically control the location of stable levitation positions within the acoustic field, generating controlled collisions between a pair of particles. The charges on the particles can then be measured entirely non-invasively using the acoustic field, isolating the issues of granular charging to the repeated collisions between the particles. Our experimental protocol takes an important step towards capturing the dependence of charging on size, shape, material, spin, and collisional energy.

Preliminary data demonstrates the utility of this approach with respect to tribocharging between a pair of grains of the same material. We show that pairs of polyethylene grains do not exchange significant charge over approximately one hundred collisions. In contrast, pairs of polystyrene particles exchange charge at a relatively constant rate of $\approx 20\,000\ e$/collision (in units of the elementary charge $e = 1.6 \times 10^{-19}$ C). Sulfonating the surface of one of the polystyrene particles increased this charging rate by a factor of 10.

Our results also suggest a general method to manipulate the location and number of potential minima in an acoustic trap. Previous designs for the transport of materials using acoustic levitation utilise highly structured acoustic interference patterns, through either highly coordinated inputs to a series of independently driven transducers (Foresti et al. [2013], Courtney et al. [2014], Marzo et al. [2015], Baresch et al. [2016], Marzo et al. [2018]), or geometric patterns on the reflector and transducer surfaces (Melde et al. [2016], Wang et al. [2016]). In contrast, our method requires only one transducer, with a single electrical drive, and the actuated motion of a boundary, to produce acoustic traps that can be reconfigured in real time.

5.3 Experimental Setup

The basic idea for the experiment is to actuate collisions between a pair of particles using ultrasound, separate them to their original positions, and then to extract the charge of the particles from their resonant oscillatory motion inside the trap. Thus the measurement sequence typically involves the following steps: 1) separately levitating two particles in the acoustic trap, 2) measuring their charge by applying an oscillating electric field, and 3) colliding the particles such that they return to their initial positions in the acoustic trap. After each collision, or after a series of such collisions, we measure the charge on each particle by repeating step (2). This sequence of events is repeated under computer control, allowing us to precisely track the charge transferred during the collision between a pair of grains.

5.3.1 Overview

Figure 5.1(a) displays a schematic of the experimental setup. A function generator (Agilent 3322a) and high-voltage amplifier (AA Labs A-301HS, gain 20) drive an ultrasound transducer at 41kHz in air (speed of sound $c = 343$ m/s, wavelength $\lambda \approx 8$ mm) via piezoelectric disks (peak-to-peak voltage 360 V). An aluminum plate, spaced a distance h beneath the transducer, acts as a reflector. Adjusting the spacing between the reflector and the base of the transducer to half the sound wavelength ($\lambda/2$) produces a standing wave with a single pressure node at $\lambda/4$, in which particles can be levitated. We confine the particles to a one-dimensional track along the diameter of the transducer by machining a small channel ($l \times w \times d$ =3.15×50×0.31 mm^3) in the reflector, and adjusting the distance between transducer and the top of the channel to match the resonance condition.

A scale drawing of a cross-section of the aluminum transducer is shown in Fig. 5.1(c). Four piezoelectric disks are bolted between an aluminum cap and transducer. Electrodes of alternating voltage are placed between the piezoelectric disks, such that both the cap and transducer base are grounded. These piezoelectric disks drive the base of the transducer. We

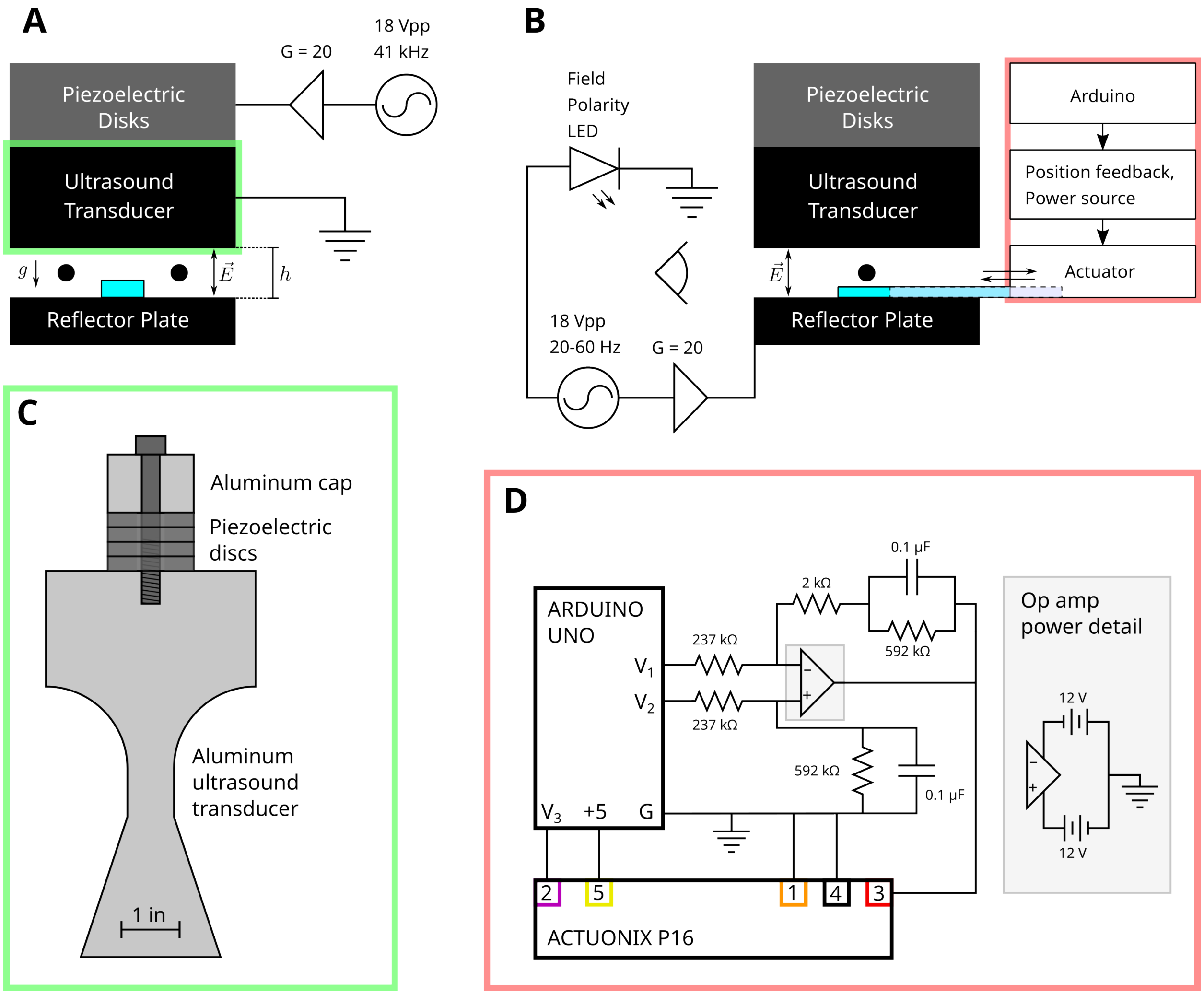

Figure 5.1: The experimental setup. (a) Schematic of the experiment from the perspective captured by the high-speed camera (front). A pair of submillimeter particles (black circles) are levitated halfway between a grounded aluminum transducer and an aluminum reflector. The reflector is mounted on a lab jack (Thorlabs), allowing precision control over the distance between transducer and reflector. Additionally, the reflector is connected to an AC voltage source, providing a vertical electric field across the gap between transducer and reflector. The particles are separated due to acoustic scattering from the acrylic "hand" (blue rectangle) between them, and are backlit and filmed from the front using a high-speed camera. The entire setup is enclosed in an acrylic chamber. (b) Schematic of the experiment from the side, showing the subsystems that control charge measurement (left) and particle collisions (right, red box). (c) Scale-drawing of a cross-section of the (circularly symmetric) transducer. The transducer consists of four piezoelectric disks, bolted between an aluminum cap and the transducer base. (d) Circuit diagram of the control system for the hand (and by extension, the collisions). An Arduino Uno drives a linear actuator by applying a ±5 V potential difference between two outputs, labelled V_1 and V_2. Position feedback from the actuator is then fed back into the Arduino (input V_3), allowing for full control over the position of the hand.

designed the transducer shape using finite element analysis (COMSOL) to amplify the signal from the disks and produce a spatially uniform signal, resulting in a high-amplitude, roughly plane-wave ultrasound signal. The entire assembly has a resonant frequency of approximately 41 kHz. The transducer and reflector are enclosed in an acrylic box with side-walls far from the levitation area (20"×10"×18") in order to mitigate the effect of side-wind perturbations.

We measure the net charge on the levitated particles by applying an AC frequency-swept electric field between the aluminum reflector and (grounded) transducer. This vertical electric field is controlled by a second function generator (BK 4052) and high-voltage amplifier (AA Labs A-301HS, gain 20, shown schematically on the left side of Fig. 5.1(b)), producing a total peak-to-peak voltage of 360 V across the gap between transducer and reflector. At the same time, we connect an LED in parallel with the (unamplified) output of the function generator. This LED produces a visible signal on the surface of the ultrasound transducer when the electric field is positive, allowing for direct visual access to the phase and frequency of the electric field throughout the experiment.

In order to actuate collisions between particles in the acoustic trap, we move additional scattering surfaces within the acoustic field, thus dynamically changing the stable levitation locations within the trap. Specifically, we insert and withdraw a long, thin piece of acrylic (cross section 1.7×7.6 mm^2, length 75 mm, shown in blue in Fig. 5.1(a) and (b)) from the acoustic trap. When the acrylic "hand" is inside the trap, a pair of particles can be levitated on either side of the hand. In contrast, when the hand is removed, the particles each accelerate towards the centre of the trap, collide, and subsequently bounce. Reinserting the hand then separates the particles. This hand is attached to a linear actuator (Actuonix P16), which is in turn controlled by an Arduino Uno with a position feedback circuit (Figs. 5.1(b) and (d)). The Arduino provides a positive (negative) 5 V difference on two pulse-width-modulated outputs to extend (withdraw) the hand from the trap. This signal is subsequently amplified and low-pass filtered, producing a ±12 V analog signal that drives the linear

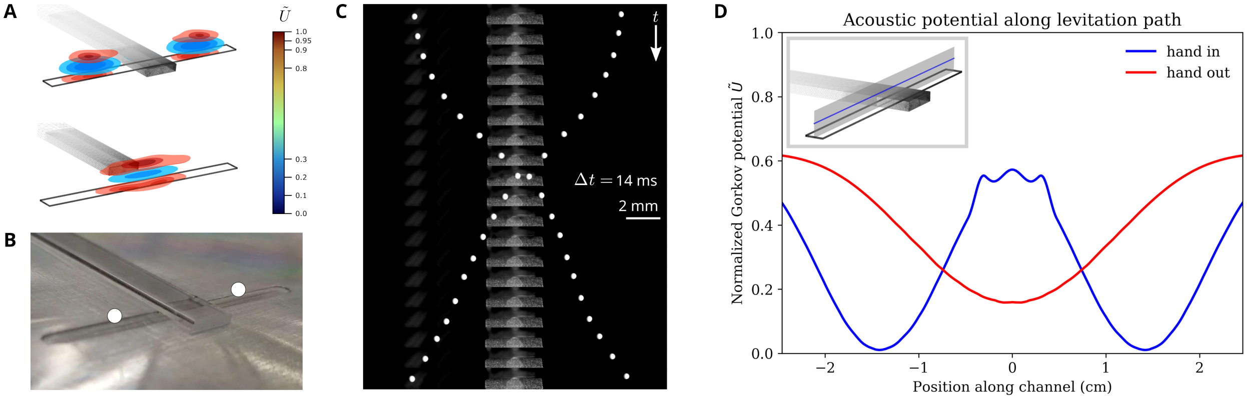

Figure 5.2: Actuating binary collisions using the acoustic field. (a) Time-series of front-lit stills from the experiment, showing a collision between a pair of polyethylene particles (white). The grey rectangle in the center is the hand. (b) Gor'kov potential (scaled by the maximum acoustic potential, and offset such that the minimum acoustic potential corresponds to zero) experienced by a particle as a function of position along channel (blue line in inset). With the hand in, (green curve) two confining wells hold the particles separate, while removal causes these minima to coalesce in the center (red curve). (c) Isosurface contours of the normalized Gor'kov potential for the two configurations: hand in (top) and hand out (bottom). The normalization for all data in (b) and (c) is the same. Subfigures (b) and (c) were generated using finite element simulations (COMSOL).

actuator. Position feedback from the actuator is then fed back to the Arduino, allowing for full control of the extension of the hand.

The entire experiment (charge measurements, collisions, and data recording) is automated using Python, which actuates a collision by signaling the Arduino to extend and then retract the hand from the cavity. The timing of this process depends on the density of the levitated grains, since denser grains accelerate more slowly. For the particles used here (polyethylene and polystyrene), the actuator was set to retract 10 mm at 32.5 mm/s, then rapidly extend to its initial position at the same rate. Throughout the experiment, a high-speed camera (Phantom v12), also controlled using Python, records the motion of the particles (500 frames per second (fps) for the charge measurement, and 2000 fps for the collisions).

5.3.2 Controlled collisions

Figure 5.2(a) shows a time-series of stills from the experiment, revealing the dynamics of a particle-particle collision. At the beginning of the collision, a pair of particles (white) levitate in the acoustic trap. The hand (centre gray rectangle) is retracted, causing the particles to accelerate towards the centre of the acoustic trap. They then collide and rebound. At the same time, the hand is reinserted into the trap, forcing the particles to return to their original positions on either side of the hand.

In order to explain this result quantitatively, we consider the forces on the particles due to the presence and absence of the hand. Particles in the acoustic trap levitate and experience forces due to acoustic scattering. In the limit of particle radius R much smaller than the levitation wavelength ($R \ll \lambda$), this acoustic force is conservative, and can be expressed as the gradient of an acoustic potential. The shape of this potential is determined by the resonance of the cavity, which in turn depends on the geometry of the cavity and the location of scattering objects within it.

Quantitatively, the acoustic potential can be calculated via a perturbation expansion of the acoustic fields due to scattering (Gorkov [1962], Bruus [2012]), such that the acoustic potential U_{rad} on a scatterer with radius R, speed of sound c_p, and material density ρ_p in an inviscid fluid with speed of sound c_0 and density ρ_0 is

$$U = \frac{4\pi}{3} R^3 \rho_0 \left[f_1 \frac{1}{2} c_0^2 \langle p^2 \rangle - f_2 \frac{3}{4} \langle v^2 \rangle \right], \tag{5.1}$$

where angled brackets denote time averages of the pressure p and velocity v. The scattering coefficients f_1 and f_2 are given by

$$f_1 = 1 - \frac{c_p^2 \rho_p}{c_0^2 \rho_0}$$
$$f_2 = \frac{2(\rho_p/\rho_0 - 1)}{2\rho_p/\rho_0 + 1}.$$

The pressure and velocity fields can thus be calculated within any trap geometry using finite element simulations, and then substituted into Eq. 5.1 to predict the acoustic potential. We calculate the acoustic potential for our specific experimental trap geometry, with and without the hand, using COMSOL. In these three-dimensional simulations, we reproduce the experimental conditions (in the frequency domain) by driving the upper boundary with a normal displacement of 1 μm, then establishing perfectly reflecting boundary conditions on a parallel surface. This choice of normal displacement is experimentally reasonable (Andrade et al. [2010]), but has no effect on the Gor'kov potential "shape" since it factors linearly into the acoustic fields. We include the presence of a channel in the reflecting surface, with dimensions matching the experimental conditions (3.15×50×0.31 mm^3). The distance between the upper boundary and the reflector surface is set to $\lambda/2$. At the non-reflecting, lateral boundaries we impose plane wave radiation conditions. We simulate the hand with a perfectly scattering rectangular block, located halfway along the channel (see inset of Fig. 5.2(b) for a diagram).

Figure 5.2(b) illustrates the effect of the hand on the structure of the acoustic field. Plotting the acoustic potential as a function of the lateral position along the channel (blue line in inset of Fig. 5.2(b)) reveals that, when the hand is in the trap, the acoustic potential develops two distinct minima: a pair of particles can be levitated on either side. Alternatively, when the hand is removed from the trap, the two potential minima coalesce into a single minimum, located at the center of the trap. This change in the geometry of the acoustic potential forces particles to accelerate towards the center of the trap and collide.

The shape of the acoustic potential in three dimensions confirms that particles can be stably levitated throughout the process of withdrawing and inserting the hand. Fig. 5.2(c) plots the isosurfaces of normalised acoustic potential $\tilde{U}$ with (top) and without (bottom) the hand. In both cases, the acoustic potential wells are localized within the channel, and retain strong gradients in the vertical direction. Importantly, the acoustic potential wells are also well localized in the horizontal direction, ensuring that the particles remain stably trapped throughout the duration of the collision process.

On occasion, the particles fail to separate properly after collision, due to acoustic or Coulombic interactions, destroying electric isolation. For instance, the particles may stick together in the center or rebound with insufficient kinetic energy, and are then pushed off the track by the hand. In this case, the control program fails to verify the presence of a particle in each minimum after a collision, and immediately ends the experiment. Alternatively, it is possible for one or both of the particles to contact a boundary, such as the bottom plate, then return to its minimum. While this destroys electrical isolation, it is not flagged by the collision verification system. For this reason, videos taken of each collision were manually checked for any such invalidating events.

We note that our method may fail for grains that are either too small or too large. In both cases, our expression for the acoustophoretic potential breaks down. In the small grain limit, particle size approaches the viscous boundary layer thickness δ, and the acoustophoretic force changes significantly (Settnes and Bruus [2012]). In our experiment, $\delta \approx 10\mu\text{m}$. We further note that $\delta \propto \omega^{-1/2}$, so driving a transducer with a higher frequency would push this boundary layer thickness lower and potentially allow for smaller particles. As the particle size increases, it approaches the acoustic radiation wavelength $\lambda \approx 8\,\text{mm}$. In this case, the Gor'kov expression overestimates the potential experienced by the particle. This weakening begins to be noticeable at a radius of around $0.1\lambda \approx 800\,\mu\text{m}$. If one were to increase the transducer driving frequency in order to allow for smaller particles, this upper limit would also decrease, though at a faster rate of ω^{-1}.

5.3.3 Charge measurement

In order to measure the charge, we apply a frequency-swept AC electric field $\vec{E}(t)$ (sweep rate 0.5 Hz/s) across the gap between the transducer and reflector. This electric field oscillates each particle simultaneously. At the same time, the particles are subject to a vertical confining force due to the acoustic field (illustrated in Fig. 5.2(c)). The trajectory of each particle is thus a function of both its charge and the amplitude of the acoustic field.

We ensure that the range of frequencies for the AC electric field includes the particle resonant frequency in the trap by manually setting the frequency limits for the first collision. During each subsequent collision, a Python script analyzes the oscillations of the particles during the previous charge measurement, and extracts the resonant frequency of the particles in the acoustic trap. The frequency range for the subsequent charge measurement is then adjusted such that the resonant frequency of a particle occurs two-thirds of the way through the frequency sweep. For the data shown in Fig. 5.5(a) (a pair of polyethylene grains), the size of the window was 15 Hz; for the data shown in Fig. 5.5(b) and (c), the size of the window was 20 Hz.

A back-lit still from the experiment during a charge measurement is shown in Fig. 5.3(a). From the data, we extract the trajectories of the two particles (Fig. 5.3(b)). We also measure the intensity of the LED connected to the electric field signal generator (Fig. 5.3(c), data in blue). When $\vec{E}(t)$ is positive, the LED shines with an intensity proportional to the amplitude of the electric field. Fitting this signal to the form of the applied frequency sweep (see Appendix B for details) thus allows for the instantaneous measurement of the phase and amplitude of $\vec{E}(t)$ during the motion of the particles.

In order to determine the charge on the particle from its trajectory in the acoustic field, we start from the equation of motion for the vertical motion of the particle, with mass m

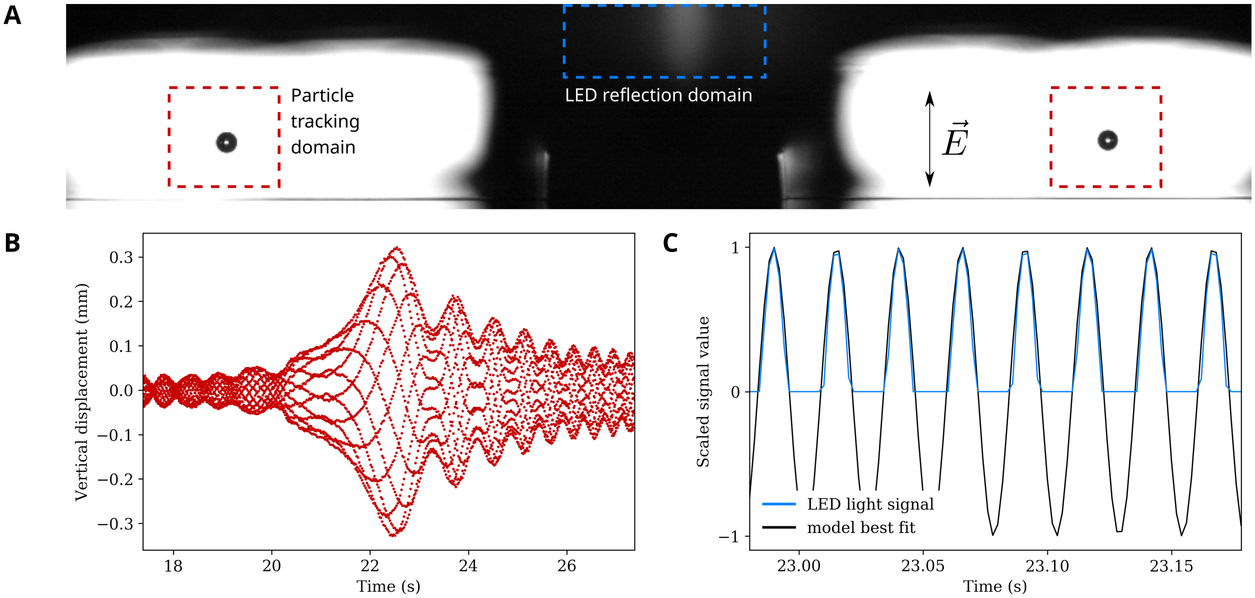

Figure 5.3: Measuring charges on individual particles using high-speed videography. (a) A still from the experiment, showing the (backlit) image captured by the high-speed camera during a charge measurement. A pair of particles (polystyrene, D =0.63 mm) are levitated on either side of the hand (outlined in gray box). The transducer is visible on the top of the image, with the light from the field polarity LED visible in the centre (bright patch). The dotted boxes indicate the separate parts of the image which are later analyzed: the trajectory of the two particles on either side (red boxes), and the intensity of the LED in the central box (blue dotted line). A bright LED corresponds to a positive (upward) electric field. (b) Experimental data for a section of the vertical particle displacement as a function of time, $y(t)$, for a levitated particle in a frequency-swept AC electric field. $y(t)$ here is measured from the median position of the particle, corresponding to its stable levitation position. (c) Experimental data for the brightness of the LED as seen on the surface of the transducer (blue line), normalised by its maximum brightness. The LED is connected in parallel with the electric field produced by the function generator, such that its brightness corresponds to the vertical component of the AC-swept electric field. The brightness of the LED is then fit to a frequency-swept sine wave (black line), recovering the phase and frequency of the electric field during the motion of the particles.

and charge q:

$$m\ddot{y} = -mg - F_d + F_{ay} + qE_0 \sin\left(\omega_E(t)t\right). \tag{5.2}$$

Here mg is the force due to gravity, F_d the air drag F_{ay} the acoustic force in the vertical direction, and E_0 and $\omega_E(t)$ are the amplitude and frequency of the applied frequency-swept electric field.

To derive F_{ay} in Eq. 5.1, we consider the acoustic velocity potential for a standing wave in the y-direction, with the reflector plate set to $y = 0$:

$$\phi(y,t) = -\frac{v_0}{k} \cos ky \sin \omega t\,, \tag{5.3}$$

where v_0 is the maximum acoustic velocity, $k = 2\pi/\lambda$ is the wavenumber, and $\omega = kc$ is the angular frequency of the sound. The acoustic pressure and velocity in the y-direction are thus

$$p = -\rho \frac{\partial \phi}{\partial t} = \rho c v_0 \cos ky \cos \omega t \tag{5.4}$$

and

$$v = \frac{\partial \phi}{\partial y} = v_0 \sin ky \sin \omega t. \tag{5.5}$$

Substituting Eqs. 5.4 and 5.5 into Eq. 5.1, then taking the derivative, yields an expression for F_{ay}:

$$F_{ay} = \frac{5}{8} m v_0^2 k \sin 2ky \,. \tag{5.6}$$

COMSOL simulations confirm that this expression for F_{ay} is accurate even in the presence of the hand and channel. We note that the nonlinearity of this restoring potential is nontrivial, and that the amplitude response of a particle driven near its trap resonance yields rich properties (Fushimi et al. [2018]).

We model the force due to air resistance with the form

$$F_d = 2m\beta_0 \dot{y} + 2m\beta_1 |\dot{y}| \dot{y} \,, \tag{5.7}$$

where the coefficients β_0 and β_1 are fitting parameters to be derived from the measured particle trajectory. In total, there are four fitting parameters to be derived from the particle trajectory: the particle charge q, the acoustic amplitude $a \equiv \frac{5}{8} v_0^2 k$, and the air drag coefficients β_0 and β_1.

The full equation of motion for the charged, acoustically levitated particle in a frequency-swept AC electric field is then

$$\ddot{y} = -g + a \sin 2ky - 2\beta_0 \dot{y} - 2\beta_1 |\dot{y}| \dot{y} + \frac{qE_0}{m} \sin \omega_E(t) t \,. \tag{5.8}$$

Given the trajectory $y(t)$, its derivatives $\dot{y}$ and $\ddot{y}$, and the instantaneous phase and amplitude of the electric field, Eq. 5.8 is linear in the four unknown constants a, β_0, β_1, and q. We thus measure the charge by performing a linear regression on the complete trajectory data from the experiment, and extracting the four unknown parameters from the coefficients of the fit.

5.4 Charge fitting procedure

Our basic strategy is to treat the acceleration terms on the right hand side (RHS)–those that are functions of position, velocity, and time–as independent variables, and treat the total acceleration as the dependent variable. To see how this can be solved with regression, we begin by re-writing Eq. (5.8) in vector representation. We write the set of unknown physical parameters in the form $\vec{\theta} = (a, \beta_0, \beta_1, q)^T$. For a given estimate of $\vec{\theta}$, the equation of motion can be written at each time point t_j as

$$Z_j = \ddot{y}(t_j) + g = \vec{X}(t_j) \cdot \vec{\theta} + A(t_j)\epsilon(t_j) \tag{5.9}$$

where

$$\vec{X}(t_j) \cdot \vec{\theta} = aX_1(t_j) + \beta_0 X_2(t_j) + \beta_1 X_3(t_j) + qX_4(t_j)\,.$$

Here we have moved the acceleration due to gravity g to the left hand side (LHS) because it is known *a priori*. Additionally, because $\vec{\theta}$ is time independent, we must include a time dependent error $\epsilon(t_j)$. According to this construction, each $X_i(t_j)$ is one term on the RHS of Eq. (5.8) corresponding to its coupling θ_i. For example, X_1 corresponds to $\theta_1 = a$, so $X_1(t_j) = \sin 2ky(t_j)$. By including the error scaling in a diagonal matrix $\Omega_{jj} \equiv 1/A(t_j)$, we can rewrite Eq. (5.9) for all t_j simultaneously as

$$\vec{Z} = X^T\vec{\theta} + \Omega^{-1}\vec{\epsilon}. \tag{5.10}$$

The error scaling Ω corrects for the fact that the error variance in the un-scaled case is proportional to the squared amplitude of the trajectory $A(t_j)^2$ (heteroscedasticity). As a result, we statistically weight the data to increase the significance of those data where the oscillation amplitude is small (lower noise)(Fox [1997]). The best linear unbiased estimate for $\vec{\theta}$ can then be found by minimizing ϵ^2, the sum of squared residuals, or equivalently by enforcing $X\Omega\vec{\epsilon}=0$. This yields

$$\vec{\theta}=(X\Omega^2X^T)^{-1}X\Omega^2\vec{Z}\,. \tag{5.11}$$

A successful measurement of $\vec{\theta}$ hinges on a maximally accurate measurement of all elements of X and Z. We outline three corrections to common issues with the data here. First, the acoustic acceleration term of Eq. (5.8) depends sensitively on the absolute height of the particle above the reflector, which cannot be straightforwardly extracted from the video. In order to correct for this unknown offset to the trajectory data, we perform linear regression as in Eq. (5.11), but where the position data is shifted by some amount φ, for some small range of φ. We then take the φ that minimizes the residuals of the regression to be an estimate for the true height offset. For a full discussion of the technical details, including the range over which φ was varied, see Appendix A.

Second, measuring the acceleration of the particle due to the electric field depends on an exact measurement of the instantaneous electric field. Since the charge q is only coupled to the trajectory through the electric field, any measurement error in the electric field leads to a loss of precision in the charge measurement. In the current experiment, there is some error due to a lack of synchronization between the camera and the function generator. We account for this error by fitting the signal from the LED to the input signal from the function generator, with a phase delay ψ and apparent timescale t_s (details in Appendix B). This allows us to circumvent the lack of synchronisation between the camera clock and the function generator clock, thus measuring the true instantaneous electric field. Future experiments could avoid this issue by synchronizing the two clocks, or by synchronizing an oscilloscope with the camera clock and directly measuring the driving voltage.

Third, measuring the damping terms and the LHS of Eq. (5.8) requires taking derivatives of the position data. The trajectory of each particle is approximately harmonic, with a similar

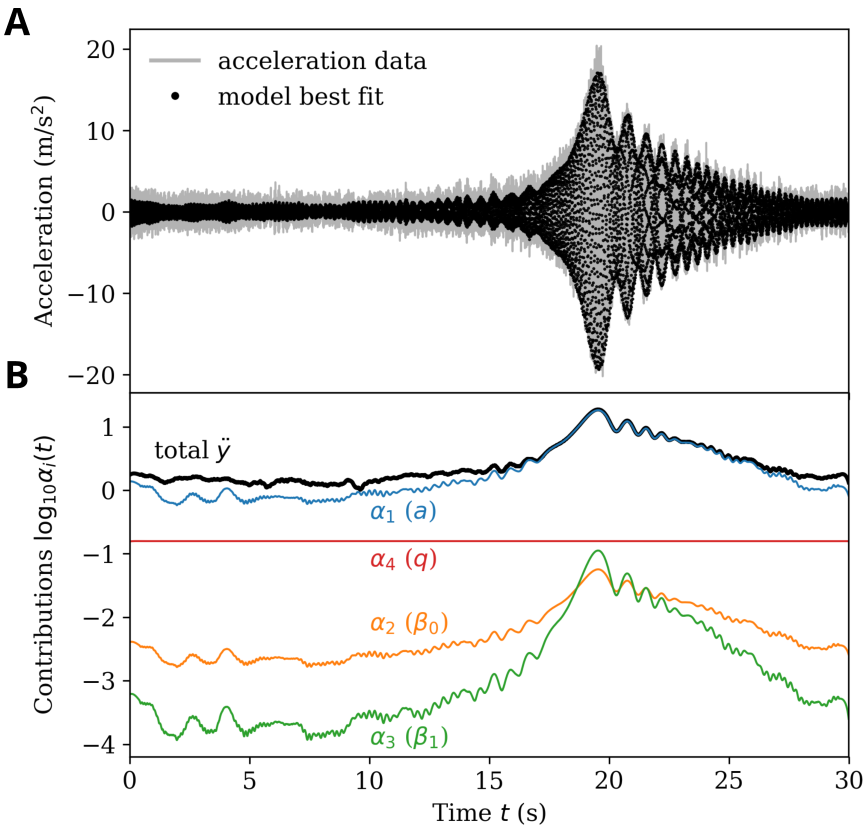

Figure 5.4: Demonstration of equation of motion fitting using Eq. (5.11). (a) Measured acceleration data (gray line) are compared to the best fit equation of motion (black points). (b) Amplitude of the contribution of each term on the right hand side of Eq. (5.8) to the total acceleration (black line) as a function of time. We plot the acoustic acceleration (blue, α_1), electric field acceleration (red, α_4), linear velocity-dependent damping (orange, α_2), and nonlinear damping (green, α_3). The amplitudes of the signals were extracted using a Hilbert transform.

frequency to the frame rate of the camera (the camera frame rate is less than a factor of 10 greater), such that finite difference estimates for the velocity and acceleration underestimate the true values. We correct for this by re-scaling the estimated velocity and acceleration (see Appendix C for details). Filming at a higher frame-rate would reduce the effect of this error.

Figure 5.4(a) compares the (corrected) experimental acceleration data, $\vec{Z}$ (gray lines), with the best fit to the model (Eq. (5.11)), $X^T\vec{\theta}$ (black points). We find generally good agreement over the entire time-series, with the largest errors appearing for large positive accelerations (near $t = 20$ in Fig. 5.4(a)). This excess error appears in general for high-amplitude trajectories, suggesting that the forces on the particle near the channel are stronger than predicted by our ansatz, Eq. (5.6).

In order to ascertain the impact of this excess error, we plot the contributions of the four fitted forces to the totally acceleration data (Fig. 5.4(b)). Throughout the trajectory, the acoustic force (α_1, plotted as a blue line) contributes most significantly to the final result: the acoustic force is measured with the highest certainty for all parts of the reconstructed trajectory. The damping forces (α_2, plotted in orange, and α_3, green) contribute most strongly when the amplitude of oscillation is large. In contrast, the acceleration due to the electric field (α_4, plotted in red) is constant throughout the sweep, and can therefore be measured most accurately in the low amplitude parts of the trajectory, when the contributions from the other forces are proportionally smaller. The deviation of the fitted trajectory from the data at high amplitudes thus has only a relatively small effect on the measured charge. The effect of this excess error on the charge measurement is further reduced by the heteroscedastic weighting of the data points, which statistically favors points with small acceleration amplitude.

5.5 Example data

As a demonstration of the generality of the charge measurement procedure, we collided several types of particle. We began with commercially available polyethylene particles (Cospheric, material density 1 000 kg m^{-3}, diameter 710-850 μm), and polystyrene particles (Norstone, material density 1 050 kg m^{-3}, diameter 620-780 μm). In addition to the bare particles, we also sulfonated the polystyrene particles following the procedure described in Ref. (Coughlin et al. [2013]): 4g polystyrene was added to a vessel with 40 mL pure sulfuric acid. This mixture was stirred at 60 °C for one hour, then removed from heat and rinsed thoroughly with DI water.

Charging time-series for three types of particle-particle collision are shown in Fig. 5.5. In

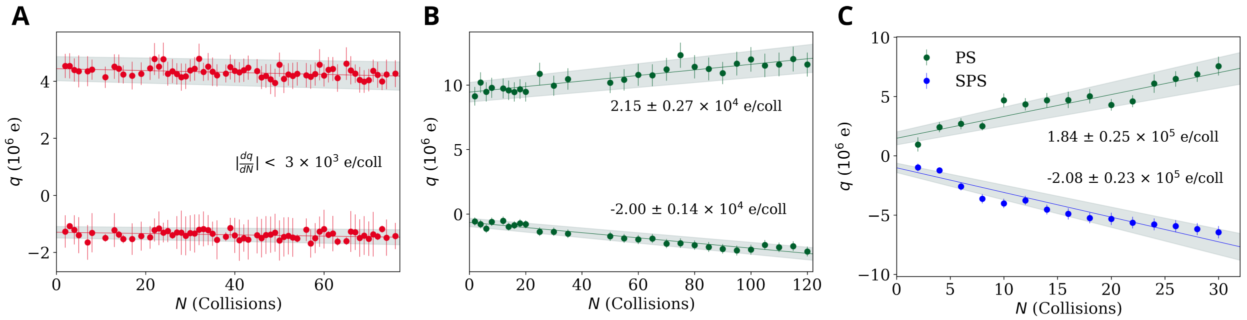

Figure 5.5: Three examples of charging data obtained using our method, with error estimates for charges and charging rate given at 1 standard deviation, as calculated from Eq. (5.13). (a) Data from 76 collisions between two polyethylene (PE) grains, showing no significant charging within experimental uncertainty. (b) Data from 120 collisions between two polystyrene (PS) grains. (3) Data from 30 collisions between two different grains, one polystyrene and one sulfonated polystyrene (SPS).

order to estimate the error associated with the charge fitting, we consider the uncertainty in the charge for a single charge measurement (measuring and fitting the response of a particle to a single AC frequency sweep). From standard weighted least squares regression, the uncertainty Σ_θ in the fitting parameters $\vec{\theta}$ is given by

$$\Sigma_\theta \approx \Sigma_s + (X\Omega^2X^T)^{-1}\frac{\epsilon^2}{N_t}\,. \tag{5.12}$$

This expression includes the uncertainty in the conversion from the measured size of an object to its actual size (pixel scale error Σ_s, see Appendix D for more details), as well as the measurement error in the trajectory itself (ϵ, with total number of data points N_t), propagated through the regression. On the basis of this expression, the uncertainty in the measured charge for a single data point is thus the square root of the matrix element of Σ_θ associated with q:

$$\delta q = \sqrt{(\Sigma_\theta)_{44}}\,. \tag{5.13}$$

Equation 5.13 provides the error on an individual charge measurement after a collision. These errors are plotted as vertical lines around each point in Fig. 5.5. From the charge time-series, we can also extract the rate of charge transfer between a pair of grains, dq/dN. In general, we find that each particle charges by a constant amount with each collision. We thus fit the charging time-series to a line using linear regression (plotted as solid lines in Fig. 5.5), such that the slope of the line gives dq/dN. The standard error of the fit, which combines the scatter in the data with the error on the individual data points, is plotted as a gray shaded area. See Appendix A for a detailed derivation of the standard error of the charging rate.

Within the error of the data, we find that polyethylene particles do not exchange significant charge over the scale of 76 collisions (Fig. 5.5(a)), with measured charging rate dq/dN smaller than $3000\,e$/collision. In contrast, colliding a pair of polystyrene particles (Fig. 5.5(b)) produces a charging rate of $dq/dN \approx 20\ 000\ e$/collision. This charging rate is highly consistent: even when the particles were allowed to collide several times in between charge measurements, the charge time-series follows the same linear trend. Sulfonating one of the polystyrene particles (Fig. 5.5(c)) enhanced the rate at which particles exchanged charge by a factor of almost 10, with measured charging rate of $dq/dN \approx 200000\,e$/collision, suggesting a link between surface chemistry and the propensity to exchange charge.

In our current setup, each data set took several hours to collect. For instance, the data shown in fig 5a depict 76 collisions, over the course of 3 hours. The time taken to collect the data can be divided into the time taken for the actual collisions (a few seconds), frequency sweeping the AC electric field to measure the charges (30-40 seconds), and saving the high-speed video data (90 seconds). This timing is highly dependent on the specific experimental configuration – for instance, decreasing the sweep rate for the electric field would also increase measurement precision, with a corresponding increase in run-time.

Throughout each collision-series, the total charge of the particles is conserved within experimental error, in line with previous findings (Lee et al. [2018]) that levitated particles exchange charge only during collisions (the particles do not exchange charge with the ambient gas). In addition, the lack of charge saturation implies that our particles contact each other at slightly different spots each time, in agreement with previously reported trends (Lee et al. [2018], Harris et al. [2019]). Based on these two observations, we can infer an average surface charge density. By measuring the collisional velocity of the particles during the experiment, we estimate a maximum contact area $A_c \sim 1000\,\mu\text{m}^2$ (see Appendix B for details). This corresponds to an average transferred surface charge density of $3\,e/\mu\text{m}^2$, $20\,e/\mu\text{m}^2$, and $200\,e/\mu\text{m}^2$ respectively for the three data sets shown in Fig. 5.5, which are comparable to previous studies of granular tribocharging (McCarty and Whitesides [2008], Haeberle et al. [2018]).

5.6 Conclusions

We have constructed an experimental method capable of noninvasively triggering repeated controlled collisions between grains, and then measuring their individual electric charge with high precision. Our experiment is the first demonstration of a tribocharging experiment where the grains are completely isolated from their surroundings, aside from the grain-grain contacts, allowing for clean, high-precision access to the basic physics of granular contact charging. This is particularly important in cases where the charging rate is so small that subtle differences in the initial condition of the grains (material, hydrophilicity, surface chemistry) have a significant effect on the overall charging behaviour.

The acoustic-levitation-based technique we demonstrate here is material-independent, and can be extended straightforwardly to other particle types, surface chemistries, shapes, and sizes. Since the dynamics of the particle-particle collision are controlled by the acoustic trap, our setup could also eventually be extended to probe the effect of collisional velocity and spin on collisional charging. In particular, the collisional velocity could be varied by increasing or decreasing the acoustic field amplitude via the transducer driving voltage, and tunable acoustic vortices may be used to control particle spin(Marzo et al. [2018]). Our method to trigger controlled collisions between levitated objects is highly general, and serves as a platform for further studies of non-equilibrium assembly, as well as applications in containerless processing.

5.7 Equilibrium point calculation

When we recover position data from video frames, the absolute height of the particle from the reflector plate cannot be measured simply. The only reliable measurement we can make is relative positions between frames. This is only an issue because of the nonlinear dependence of the acoustic force (5.6) on position, shifting $Y_j \to Y_j + \Delta y$ changes X_{1j}, but none of the other $\vec{X}_i$. To estimate the correct shift Δy to apply to our data, we first approximate the equilibrium position assuming a linearized acoustic potential, then perform a brute-force search in the neighborhood of this guess using the full nonlinear potential. The optimal equilibrium position then uniquely determines the necessary shift Δy. In properly shifted coordinates, the true equilibrium position is given by

$$y_0(a) = \frac{\pi}{2k} - \frac{1}{2k} \arcsin\left(\frac{g}{a}\right) . \tag{5.14}$$

Thus, in principle, shifting $Y_j \to Y_j - \langle Y \rangle_t + y_0(a)$ solves this issue, but doing so causes the equation of motion to become nonlinear in $a = \theta_1$ making standard linear regression inappropriate. Note that $\langle \cdot \rangle_t$ denotes averaging over the time index. To begin, we obtain an estimate for a from the resonance frequency of the particle. Following Lee et al.(Lee et al. [2018]) the instantaneous frequency $f(t)$ and amplitude $A(t)$ of the particle trajectory are found via Hilbert transform. Taking the resonance time to be $\arg\max_t A(t) = t_{res}$, then approximating the acoustic force as linear, we find

$$a \approx \frac{(2\pi f(t_{res}))^2}{2k}\,.$$

To correct this initial guess, we perform linear regression on the trajectory for a set of fixed values $\varphi \equiv (1/2k)\arcsin(g/a)$ in a small neighborhood. We then take φ_0, the value of φ leading to the best regression, as our estimate. Quantitatively, we define $X_1(\vec{Y}, t_j)$ as Eq. (5.6) evaluated on the shifted position data:

$$X_1(\vec{Y}, t_j) = \sin(2k(Y_j - \langle Y \rangle_t) - \varphi_0)\,.$$

When fitting on simulated data, the time domain used in regressions for this preliminary minimization is irrelevant. In real data however, different time domains seem to yield different estimates for φ_0. We elected to always fit over a 2s window immediately preceding t_{res} for a few reasons. First, as we will discuss in Appendix C, velocity and acceleration data need to be re-scaled to reverse errors induced by finite difference effects. This re-scaling approximation is frequency dependent and is most dependable when the amplitude is growing due to driving near resonance. Second, within this regime Fig. 5.4 illustrates that the scale of acceleration due to acoustic force is several decades above all other effects.

5.8 Electric field fitting

In order to couple q to acceleration in our equation of motion, we need to know the electric field in the cavity as a function of time. As shown in Fig. 5.3(a,c), the electric field strength is encoded frame-by-frame as the reflection of an LED on the transducer surface which is visible in the upper portion of the picture. When the LED is on, the $\vec{E}$ field has a positive vertical component which is approximately proportional to its brightness. The challenge we are presented with is fitting a model of the electric field to this data. We know that the electric field is a swept sine with a total sweep time T, initial frequency f_i, and final frequency f_f. The instantaneous frequency is

$$\omega_E(t) = 2\pi\left(\frac{\alpha}{2}t + f_i\right)t, \quad \alpha = \frac{f_f - f_i}{T}\,.$$

This model admits a prediction for the time points τ_n corresponding to the peaks of the electric field, which are the most easily extracted feature of the data. Because $\vec{X}_4$ is the only term coupling the charge q to the data, this fit needs to be extremely precise, and two seemingly small effects must be taken into account. The first is a slight phase shift in the signal which accounts for the finite frame rate: since the function generator is not synced with the camera, the sweep trigger almost always falls at some point in time between two exposures. The second effect is a minuscule error in the scale of time as a result of very slightly different clock speeds between the camera and function generator. We therefore introduce two fit parameters, a phase shift ψ and time scale t_s, which we will vary over in an attempt to minimize the sum of squared errors in observed ($\tilde{\tau}_n$) and calculated (τ_n) peak times:

$$\begin{aligned} \tau_n(\psi, t_s) &= -\frac{f_i}{t_s\alpha} + \frac{1}{t_s}\sqrt{\left(\frac{f_i}{\alpha}\right)^2 + \frac{1}{\alpha}\left(\frac{\psi}{\pi} + \frac{1}{2} + 2n\right)} \\ (\psi, t_s) &= \underset{\psi,\ t_s}{\arg\min} \sum_n (\tilde{\tau}_n - \tau_n(\psi, t_s))^2 \end{aligned} \tag{H15}$$

A typical value for this time scaling in our setup is around $t_s \approx 0.99996$. After carrying out this minimization, we can evaluate $X_4(t)$.

$$X_4(t) = \frac{E_0}{m}\sin(\omega_E(t_s t)t_s t - \psi) \tag{H16}$$

5.9 Finite difference rescaling

Throughout the sweep, each particle's trajectory is approximately harmonic with frequency not too different from the frame rate: $1/f \sim 10\Delta t$. Because these rates are similar, finite difference estimates for velocity and acceleration underestimate the true values. We correct for this by re-scaling the velocity and acceleration according to the error induced at maximal values:

$$\dot{y} \to g_1 \dot{y}, \quad \ddot{y} \to g_2 \ddot{y}\,.$$

For a central first-order difference acting on a sinusoid, the error appears at the zeros of the signal:

$$g_1(\omega) = \frac{\omega \Delta t}{\sin \omega \Delta t}\,.$$

The second order difference introduces error at the peaks of the signal:

$$g_2(\omega) = \frac{\omega^2 \Delta t^2}{2 - 2\cos \omega \Delta t}\,.$$

This rescaling is frequency dependent, and we approximate the response frequency as the driving frequency $\omega_E(t)$. If $\Delta^{(1)}$ and $\Delta^{(2)}$ are our first- and second-order central finite difference operators respectively, then we can define:

$$\begin{aligned} V_j(\vec{Y}) &= g_1(\omega_E(t_j)) \left[\Delta^{(1)} \vec{Y}\right]_j \\ X_2(\vec{Y}, t_j) &= 2V_j \\ X_3(\vec{Y}, t_j) &= 2|V_j|V_j \end{aligned}$$

Similarly,

$$Z_j(\vec{Y}) = g_2(\omega_E(t_j)) \left[\Delta^{(2)}\vec{Y}\right]_j + g$$

As we take $\omega\Delta t \to 0$, we find $g_1 \to 1$ and $g_2 \to 1$. If it can be achieved, this is a preferable condition, as this re-scaling is only an approximation and varies in validity throughout the course of a sweep. This highlights an important consideration for future iterations of this experiment. For experiments with stronger acoustic potentials, a proportionally higher frame rate will be necessary to keep this error in check.

5.10 Error in measured charging rate

Equation (5.12) includes an approximation for Σ_s, the variance contributed by uncertainty in the pixel scale measurement. This does not affect each θ_i equally: due to approximate linearity of X_1 and X_2 on $\vec{Y}$, does not significantly impact θ_1 or θ_2. When s is the pixel scale and σ_s is its uncertainty, we have

$$\Sigma_s = \left(\frac{\sigma_s}{s}\right)^2 \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & \theta_3^2 & \theta_3\theta_4 \\ & & \theta_4\theta_3 & \theta_4^2 \end{bmatrix}. \tag{J17}$$

To express the variances for charging rate fit parameters, we first need to define more terms. Let the charging data be $\vec{q} = (q_1, \ldots, q_{N_q})^T$ with $\delta\vec{q}$ defined similarly. Let C be a $2\times N_q$ matrix such that C_{1j} is the number of collisions preceding the measurement of q_j and $C_{2j} = 1$. Then for our charging rate fit parameters $\vec{\gamma}$, we have

$$\vec{q} = C^T\vec{\gamma} + W^{-1}\vec{\eta}.$$

with $W_{ii} = 1/\delta q_i$. Thus

$$\begin{aligned}\vec{\gamma} &= (CW^2C^T)^{-1}CW^2\vec{q} \\ \Sigma_\gamma &= (CW^2C^T)^{-1}\left(CW + \frac{\eta^2}{N_q - 2}\right) .\end{aligned} \tag{J18}$$

We then calculate the error in charging rate using the diagonals of Σ_γ to find

$$\delta\gamma_i = \sqrt{(\Sigma_\gamma)_{ii}} .$$

5.11 Internal consistency of fitting method

In order to verify our charge measurement method, we used two internal consistency checks. First, to demonstrate the accuracy of our fitting method, we produced simulations of a particle subject to the expected equation of motion and applied our fitting method to the data thus produced. We find that our method extracts the correct charge to within 0.5% error over the range of charge magnitudes observed in our experiment. We elaborate on this below and in Fig 5.6. Another form of internal consistency which we have demonstrated is charge conservation. Since the particles are totally electrically isolated from their environment, and no appreciable net charge transfer occurs with the surrounding air, the charging rates in each trial should be equal and opposite. Within error, this condition is not violated by our data.

Our first internal consistency test addresses possible bias in our fitting method by applying it to simulated data with known input parameters. To this end, we chose a range of electric charge values spanning two orders of magnitude which covered all values observed in experiment. For each charge in this set, we simulated the trajectory of a particle with that charge, subject to our proposed equation of motion. We then applied our fitting method to this simulated data while holding all other experimental parameters (i.e. acoustic accelera-

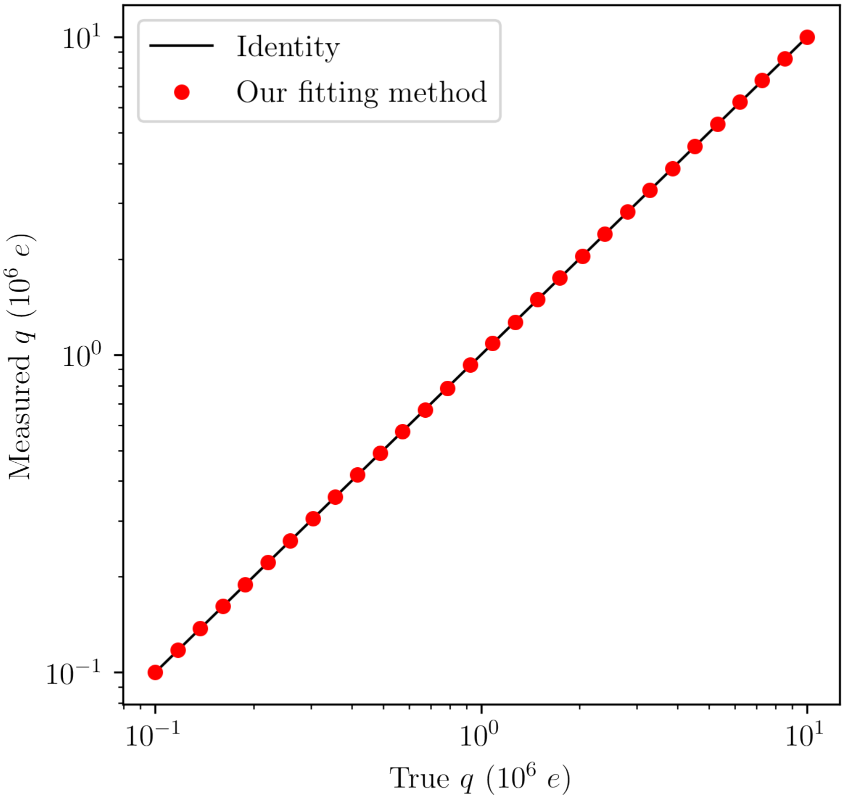

Figure 5.6: Here we plot the fit-determined charge as a function of the true charge (used as an input to the simulation). Simulation consisted of Runge-Kutte 4(5) integration, with physical parameters taken from the trial represented in Fig. 5b. Charge estimates all lie within 0.5% error.

tion, drag) fixed. The resulting measured charges are plotted as function of true charge in Fig 6. We find no evidence of extra biases introduced by our fitting method over the range of charges investigated. We further note that the error observed here is almost entirely due to the fitting step wherein the "zero point" of position is found, as detailed in Appendix A. When the exact height is used, the error is negligible. In the experiment, the error is mostly due to uncertainty in pixel scale, which underlies many of our physical measurements.

5.12 Estimate of the maximum contact area during a collision

We estimate the maximum contact area A_c during a head-on collision between two elastic spheres of radii R_1 and R_2 using Hertzian contact theory.

Upon collision, the potential energy stored in elastic deformation will be(Landau et al. [1975], Timoshenko and Goodier [1970])

$$U = \frac{2}{5}\varepsilon r^{1/2} h^{5/2} \,. \tag{L19}$$

Where $(2R_1 + 2R_2) - h$ is the total distance between the centers of mass of the particles at maximal compression, and the reduced radius r and elastic constant ε are given by

$$\begin{aligned} r &= \frac{R_1 R_2}{R_1 + R_2} \\ \epsilon &= \frac{4}{3}\left(\frac{1-\sigma_1^2}{E_1} + \frac{1-\sigma_2^2}{E_2}\right)^{-1} \quad . \end{aligned} \tag{L20}$$

In this expression, E_i are the Young's moduli and σ_i are the Poisson coefficients. From momentum conservation,

$$h = h_1 + h_2 = h_1\left(1 + \frac{m_1}{m_2}\right) \,, \tag{L21}$$

where h_1 is the depression into particle 1, and h_2 is defined similarly. Substituting this expression for h yields

$$h_1 = \left(1 + \frac{m_1}{m_2}\right)^{-1} \left(\frac{5U}{2\varepsilon} r^{-1/2}\right)^{2/5} . \tag{L22}$$

In the center of mass frame, we can simply substitute the total kinetic energy in for U. Finally, working within the approximation of a small deformation,

$$A_1 \approx 2\pi R_1 h_1 . \tag{L23}$$

In the special case that both particles have the same material parameters and radius R, we find

$$A = 2\pi \left[\frac{15}{16} \frac{mv^2(1-\sigma^2)}{E} R^2\right]^{2/5} . \tag{L24}$$

Using videos of collisions to estimate the kinetic energies of particles prior to the collision and equating this to the total elastic potential energy, we find that $A \sim 1000\mu\mathrm{m}^2$.

CHAPTER 6

CONCLUDING REMARKS

Physics teaches us that simple models are crucial to the endeavor of theoretical science. Not every detail matters for every outcome, and there may be deep and ubiquitous reasons for this fact. Perhaps these reasons are to be found in discussions about limit theorems, universality, and information geometry. Investigations into renormalization group methods for applications outside of physics, for example, suggest that the paths to simplification in large systems are generic and not particular to the theories for which RG was originally built to handle (Brown and Sethna [2003], Raju et al. [2018], Quinn et al. [2022], Berman et al. [2023], Berman and Klinger [2022], Bradde and Bialek [2017], Jona-Lasinio [1975]). At the same time, the apparent diversity and complexity of biology can be bewildering, and it might seem unlikely that simplifying paths exist.

However complex biology may be, it is structured, and therefore not maximally complex. While connections between neurons may seem random, the brain carries out sophisticated computations on its disordered substrate. Despite genetic variation among individuals, morphogenesis leads developing organisms to remarkably consistent, reproducible forms. Even variation itself can acquire functional significance, for example in phenotypic traits distributed across bacterial colonies or in retinal ganglion cell subtypes accounting for responses to a rich vocabulary of visual features. The structures that matter biologically are those that interact with natural selection, by way of biological function and fitness. For this reason, living things are, in part, organized from the top-down. But so too are they built out of physical components and systems which had to evolve from simpler living organisms, and this imposes some amount of bottom-up organization. While we can see that simple models are something to strive for, and even that some amount of coarse-graining should be able to reveal regularities in living systems, the unique interplay of bottom-up and top-down structure makes it difficult to straightforwardly adopt ideas about coarse-graining developed for the physics of nonliving systems. In particular, the choice of a *notion of scale* implicit in any renormalization group calculation is itself difficult to ascertain in living systems. Extra care must be taken to ensure that the information one keeps under coarse-graining has functional, biological relevance.

In an effort to find principles which may lead to good coarse-graining schemes in biology, we studied the renormalization group (RG) and the information bottleneck (IB). In the IB, coarse-graining is optimized to preserve information about some specified set of signals, giving the practitioner control over what is precisely meant by "relevance". Because this can be chosen to represent biological function, the latter has found some success within biology. By formally connecting IB and RG, we argued that the notion of scale used in an RG procedure can be generalized. Under an appropriate choice of relevance variable, scale acquires the meaning of pertinence to how a given biological function is carried out. Because living systems and their parts frequently exhibit many functions, this motivates a property that we term multi-relevance, in which a system exhibits multiple coexisting RG flows, each with a different choice of collective basis. Finally, we carried out IB coarse-graining in data taken from a large population of salamander retinal neurons under the choice of future retinal state as the relevance variable. This choice formalizes the biological function of predictive computation, which is known to be important in the vertebrate visual system (Gollisch and Meister [2010], Palmer et al. [2015]). The collective variables which maximize predictive information in the salamander retina change with prediction interval, corresponding to a shift in importance of collective effects in correlation structure.

These investigations immediately motivate several new directions for inquiry. Firstly, theoretical work is needed to understand the structure of solutions to IB problems in large systems. For example, one can imagine bringing in ideas from large deviation theory (Touchette [2012, 2009]) to get approximate IB solutions in large-data or large-state space limits. Given such an understanding, the IB-RG connection could be used to turn data-driven IB analyses into effective models in the language of RG and many-body theory. This could then further be used to examine ideas about criticality in biological systems (Beggs [2008], Mora and Bialek [2011], Mora et al. [2015], Schwab et al. [2014], Tkačik et al. [2015], Levina et al. [2007], Morrell et al. [2021], Tiberi et al. [2022], Ngampruetikorn et al. [2023]) in a way that connects back to biological function.

IB and RG have also attracted attention as potential organizing frameworks in machine learning and inference in probability distributions on high-dimensional spaces more generally (Alemi et al. [2019], Shwartz-Ziv and Tishby [2017], Kolchinsky et al. [2019], Alemi [2019], Mehta and Schwab [2014], Cotler and Rezchikov [2023b], Berman et al. [2023], Berman and Klinger [2022]). Here, many questions revolve around the surprising efficacy of complicated, possibly over-parameterized models in capturing generalizable structure in data. One very exciting idea is that unsupervised methods such as VAEs and diffusion models may succeed at their tasks for similar reasons, and that RG and IB could be used to explain this. This path has already been explored to a small extent, via investigations that seek direct connections between IB, RG, and foundational ideas about inference (Zellner [1988], Bény and Osborne [2015], Alemi [2019], Berman et al. [2023]).

Another question raised by this work is the precise relationship between multi-relevance and the IB-RG connection. While multi-relevance was originally motivated by a hypothetical IB-RG analysis with several 'good' choices of relevance variable, the property we present arises from a class of mixture family constructions. This does not admit a clear connection back to IB, i.e. it is not clear whether a certain choice of relevance variable could give an appropriate coarse-graining scheme for one mixture component but not the other. In fact, we expect it is the case that the basic idea of multi-relevance, that is, the coexistence of meaningful distinct notions of scale, is more general than the mixture family construction. In work which was not presented here, we explored definitions of multi-relevance which were more in line with the original motivation. However, a central issue in those investigations is the lack of solid definition for a ‘good’ choice of relevance variable. Relatedly, since the relevance variable is chosen by the practitioner, it is difficult to see how such multi-relevance could be an intrinsic property. By contrast, the mixture family construction offers a clearer, more intrinsic notion of multi-relevance. Still, the idea that one might be able to choose a better or worse relevance variable in the application of IB is well worth exploring, as it gets to the very heart of the whole matter of coarse-graining.

REFERENCES

Physics of Life. National Academies Press, Washington, D.C., December 2022. ISBN 978-0-309-27400-5. doi:10.17226/26403. URL `https://www.nap.edu/catalog/26403`.

Tasneem Abbasi and SA Abbasi. Dust explosions–cases, causes, consequences, and control. *Journal of Hazardous Materials*, 140(1-2):7–44, 2007.

Tosif Ahamed, Antonio C. Costa, and Greg J. Stephens. Capturing the continuous complexity of behaviour in Caenorhabditis elegans. *Nature Physics*, 17(2):275–283, February 2021. ISSN 1745-2481. doi:10.1038/s41567-020-01036-8. URL `http://www.nature.com/articles/s41567-020-01036-8`. Number: 2 Publisher: Nature Publishing Group.

Alexander A. Alemi. Variational Predictive Information Bottleneck. *arXiv:1910.10831 [cs, math, stat]*, October 2019. URL `http://arxiv.org/abs/1910.10831`. arXiv: 1910.10831.

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep Variational Information Bottleneck. *arXiv:1612.00410 [cs, math]*, October 2019. URL `http://arxiv.org/abs/1612.00410`. arXiv: 1612.00410.

P. W. Anderson. More Is Different. *Science*, 177(4047):393–396, August 1972. ISSN 0036-8075, 1095-9203. doi:10.1126/science.177.4047.393. URL `https://www.sciencemag.org/lookup/doi/10.1126/science.177.4047.393`.

Marco AB Andrade, Flávio Buiochi, and Julio C Adamowski. Finite element analysis and optimization of a single-axis acoustic levitator. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 57(2):469–479, 2010.

S.M. Apenko. Information theory and renormalization group flows. *Physica A: Statistical Mechanics and its Applications*, 391(1-2):62–77, January 2012. ISSN 03784371. doi:10.1016/j.physa.2011.08.014. URL `https://linkinghub.elsevier.com/retrieve/pii/S037843711100642X`.

C. Bagnuls and C. Bervillier. Exact Renormalization Group Equations. An Introductory Review. *Physics Reports*, 348(1-2):91–157, July 2001. ISSN 03701573. doi:10.1016/S0370-1573(00)00137-X. URL `http://arxiv.org/abs/hep-th/0002034`. arXiv: hep-th/0002034.

Adrian G Bailey. The science and technology of electrostatic powder spraying, transport and coating. *Journal of Electrostatics*, 45(2):85–120, 1998.

I. Balog, A. Rançon, and B. Delamotte. Critical Probability Distributions of the Order Parameter from the Functional Renormalization Group. *Physical Review Letters*, 129(21): 210602, November 2022. doi:10.1103/PhysRevLett.129.210602. URL `https://link.aps.org/doi/10.1103/PhysRevLett.129.210602`. Publisher: American Physical Society.

Diego Baresch, Jean-Louis Thomas, and Régis Marchiano. Observation of a single-beam gradient force acoustical trap for elastic particles: acoustical tweezers. *Physical Review Letters*, 116(2):024301, 2016.

Bilge Baytekin, H Tarik Baytekin, and Bartosz A Grzybowski. What really drives chemical reactions on contact charged surfaces? *Journal of the American Chemical Society*, 134 (17):7223–7226, 2012.

HT Baytekin, AZ Patashinski, M Branicki, Bilge Baytekin, S Soh, and Bartosz A Grzybowski. The mosaic of surface charge in contact electrification. *Science*, 333(6040):308–312, 2011.

John M Beggs. The criticality hypothesis: how local cortical networks might optimize information processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1864):329–343, February 2008. ISSN 1364-503X, 1471-2962. doi:10.1098/rsta.2007.2092. URL `https://royalsocietypublishing.org/doi/10.1098/rsta.2007.2092`.

Jürgen Berges, Nikolaos Tetradis, and Christof Wetterich. Non-perturbative renormalization flow in quantum field theory and statistical physics. *Physics Reports*, 363(4-6):223–386, June 2002. ISSN 03701573. doi:10.1016/S0370-1573(01)00098-9. URL `https://linkinghub.elsevier.com/retrieve/pii/S0370157301000989`.

David S. Berman and Marc S. Klinger. The Inverse of Exact Renormalization Group Flows as Statistical Inference, December 2022. URL `http://arxiv.org/abs/2212.11379`. arXiv:2212.11379 [hep-th].

David S. Berman, Marc S. Klinger, and Alexander G. Stapleton. Bayesian Renormalization. *Machine Learning: Science and Technology*, October 2023. ISSN 2632-2153. doi:10.1088/2632-2153/ad0102. URL `http://arxiv.org/abs/2305.10491`. arXiv:2305.10491 [cond-mat, physics:hep-th].

Gordon J. Berman, William Bialek, and Joshua W. Shaevitz. Predictability and hierarchy in *Drosophila* behavior. *Proceedings of the National Academy of Sciences*, 113(42):11943–11948, October 2016. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.1607601113. URL `http://www.pnas.org/lookup/doi/10.1073/pnas.1607601113`.

Michael J. Berry, Iman H. Brivanlou, Thomas A. Jordan, and Markus Meister. Anticipation of moving stimuli by the retina. *Nature*, 398(6725):334–338, March 1999. ISSN 1476-4687. doi:10.1038/18678. URL `https://www.nature.com/articles/18678`. Number: 6725 Publisher: Nature Publishing Group.

William S. Bialek. *Biophysics: searching for principles.* Princeton University Press, Princeton, NJ, 2012. ISBN 978-0-691-13891-6.

Jürgen Blum and Gerhard Wurm. The growth mechanisms of macroscopic bodies in protoplanetary disks. *Annu. Rev. Astron. Astrophys.*, 46:21–56, 2008.

Serena Bradde and William Bialek. PCA Meets RG. *Journal of Statistical Physics*, 167(3-4):462–475, May 2017. ISSN 0022-4715, 1572-9613. doi:10.1007/s10955-017-1770-6. URL http://link.springer.com/10.1007/s10955-017-1770-6.

Braden A. W. Brinkman. Non-perturbative renormalization group analysis of nonlinear spiking networks, January 2023. URL http://arxiv.org/abs/2301.09600. arXiv:2301.09600 [cond-mat, q-bio].

M Brook, CB Moore, and T Sigurgeirsson. Lightning in volcanic clouds. *Journal of Geophysical Research*, 79(3):472–475, 1974.

Kevin S. Brown and James P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E*, 68(2):021904, August 2003. doi:10.1103/PhysRevE.68.021904. URL https://link.aps.org/doi/10.1103/PhysRevE.68.021904. Publisher: American Physical Society.

Henrik Bruus. Acoustofluidics 7: The acoustic radiation force on small particles. *Lab on a Chip*, 12(6):1014–1021, 2012.

Cédric Bény and Tobias J. Osborne. Renormalisation as an inference problem, April 2014. URL http://arxiv.org/abs/1310.3188. arXiv:1310.3188 [cond-mat, physics:hep-th, physics:quant-ph].

Cédric Bény and Tobias J Osborne. The renormalization group via statistical inference. *New Journal of Physics*, 17(8):083005, August 2015. ISSN 1367-2630. doi:10.1088/1367-2630/17/8/083005. URL https://iopscience.iop.org/article/10.1088/1367-2630/17/8/083005.

Léonie Canet, Hugues Chaté, and Bertrand Delamotte. Quantitative Phase Diagrams of Branching and Annihilating Random Walks. *Physical Review Letters*, 92(25):255703, June 2004. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.92.255703. URL https://link.aps.org/doi/10.1103/PhysRevLett.92.255703.

Léonie Canet, Hugues Chaté, Bertrand Delamotte, Ivan Dornic, and Miguel A. Muñoz. Nonperturbative Fixed Point in a Nonequilibrium Phase Transition. *Physical Review Letters*, 95(10):100601, August 2005. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.95.100601. URL https://link.aps.org/doi/10.1103/PhysRevLett.95.100601.

Léonie Canet, Hugues Chaté, and Bertrand Delamotte. General framework of the non-perturbative renormalization group for non-equilibrium steady states. *Journal of Physics A: Mathematical and Theoretical*, 44(49):495001, December 2011. ISSN 1751-8113, 1751-8121. doi:10.1088/1751-8113/44/49/495001. URL https://iopscience.iop.org/article/10.1088/1751-8113/44/49/495001.

Matthew Chalk, Olivier Marre, and Gašper Tkačik. Toward a unified theory of efficient, predictive, and sparse coding. *Proceedings of the National Academy of Sciences*, 115(1):

186–191, January 2018. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.1711114115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1711114115.

Gal Chechik, Amir Globerson, Naftali Tishby, and Yair Weiss. Information Bottleneck for Gaussian Variables. page 24, January 2005.

Jordan Cotler and Semon Rezchikov. Renormalization group flow as optimal transport. *Physical Review D*, 108(2):025003, July 2023a. doi:10.1103/PhysRevD.108.025003. URL https://link.aps.org/doi/10.1103/PhysRevD.108.025003. Publisher: American Physical Society.

Jordan Cotler and Semon Rezchikov. Renormalizing Diffusion Models, September 2023b. URL http://arxiv.org/abs/2308.12355. arXiv:2308.12355 [hep-lat, physics:hep-th].

Jessica E Coughlin, Andreas Reisch, Marie Z Markarian, and Joseph B Schlenoff. Sulfonation of polystyrene: Toward the "ideal" polyelectrolyte. *Journal of Polymer Science Part A: Polymer Chemistry*, 51(11):2416–2424, 2013.

Charles RP Courtney, Christine EM Demore, Hongxiao Wu, Alon Grinenko, Paul D Wilcox, Sandy Cochran, and Bruce W Drinkwater. Independent trapping and manipulation of microparticles using dexterous acoustic tweezers. *Applied Physics Letters*, 104(15):154103, 2014.

Felix Creutzig and Henning Sprekeler. Predictive Coding and the Slowness Principle: An Information-Theoretic Approach. *Neural Computation*, 20(4):1026–1041, April 2008. ISSN 0899-7667, 1530-888X. doi:10.1162/neco.2008.01-07-455. URL https://direct.mit.edu/neco/article/20/4/1026-1041/7297.

Felix Creutzig, Amir Globerson, and Naftali Tishby. Past-future information bottleneck in dynamical systems. *Physical Review E*, 79(4):041925, April 2009. ISSN 1539-3755, 1550-2376. doi:10.1103/PhysRevE.79.041925. URL https://link.aps.org/doi/10.1103/PhysRevE.79.041925.

Bertrand Delamotte. An Introduction to the Nonperturbative Renormalization Group. *arXiv:cond-mat/0702365*, 852:49–132, 2012. doi:10.1007/978-3-642-27320-9_2. URL http://arxiv.org/abs/cond-mat/0702365. arXiv: cond-mat/0702365.

Domenico Di Sante, Matija Medvidović, Alessandro Toschi, Giorgio Sangiovanni, Cesare Franchini, Anirvan M. Sengupta, and Andrew J. Millis. Deep Learning the Functional Renormalization Group. *Physical Review Letters*, 129(13):136402, September 2022. doi:10.1103/PhysRevLett.129.136402. URL https://link.aps.org/doi/10.1103/PhysRevLett.129.136402. Publisher: American Physical Society.

Xinqiang Ding, Zhengting Zou, and Charles L. Brooks III. Deciphering protein evolution and fitness landscapes with latent space models. *Nature Communications*, 10(1):5644, December 2019. ISSN 2041-1723. doi:10.1038/s41467-019-13633-0. URL https://www.nature.com/articles/s41467-019-13633-0. Number: 1 Publisher: Nature Publishing Group.

N. Dupuis, L. Canet, A. Eichhorn, W. Metzner, J. M. Pawlowski, M. Tissier, and N. Wschebor. The nonperturbative functional renormalization group and its applications. *Physics Reports*, 910:1–114, May 2021. ISSN 03701573. doi:10.1016/j.physrep.2021.01.001. URL `http://arxiv.org/abs/2006.04853`. arXiv: 2006.04853.

H. Erbin, V. Lahoche, and D. Ousmane Samary. Non-perturbative renormalization for the neural network-QFT correspondence. *Machine Learning: Science and Technology*, 3(1): 015027, February 2022. ISSN 2632-2153. doi:10.1088/2632-2153/ac4f69. URL `https://dx.doi.org/10.1088/2632-2153/ac4f69`. Publisher: IOP Publishing.

Daniel S. Fisher. Critical behavior of random transverse-field Ising spin chains. *Physical Review B*, 51(10):6411–6461, March 1995. ISSN 0163-1829, 1095-3795. doi:10.1103/PhysRevB.51.6411. URL `https://link.aps.org/doi/10.1103/PhysRevB.51.6411`.

Michael E Fisher. Renormalization group theory: Its basis and formulation in statistical physics. *Rev. Mod. Phys.*, 70(2):29, 1998.

Daniele Foresti, Majid Nabavi, Mirko Klingauf, Aldo Ferrari, and Dimos Poulikakos. Acoustophoretic contactless transport and handling of matter in air. *Proceedings of the National Academy of Sciences*, 110(31):12549–12554, 2013.

Keith M Forward, Daniel J Lacks, and R Mohan Sankaran. Charge segregation depends on particle size in triboelectrically charged granular materials. *Physical Review Letters*, 102 (2):028001, 2009.

John Fox. *Applied regression analysis, linear models, and related methods.* Sage Publications, Inc, 1997.

T Fushimi, TL Hill, A Marzo, and BW Drinkwater. Nonlinear trapping stiffness of mid-air single-axis acoustic levitators. *Applied Physics Letters*, 113(3):034102, 2018.

Nigel Goldenfeld. *Lectures on Phase Transitions and the Renormalization Group*. CRC Press, 1 edition, March 2018. ISBN 978-0-429-49349-2. doi:10.1201/9780429493492. URL `https://www.taylorfrancis.com/books/9780429962042`.

Nigel Goldenfeld. There's Plenty of Room in the Middle: The Unsung Revolution of the Renormalization Group, June 2023. URL `http://arxiv.org/abs/2306.06020`. arXiv:2306.06020 [cond-mat, physics:hep-th, physics:physics].

Samuel Goldman, Nima Lashkari, Robert G. Leigh, and Mudassir Moosa. Exact Renormalization of Wave Functionals yields continuous MERA, January 2023. URL `http://arxiv.org/abs/2301.09669`. arXiv:2301.09669 [hep-th, physics:quant-ph].

Tim Gollisch and Markus Meister. Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina. *Neuron*, 65(2):150–164, January 2010. ISSN 08966273. doi:10.1016/j.neuron.2009.12.009. URL `https://linkinghub.elsevier.com/retrieve/pii/S0896627309009994`.

Amit Gordon, Aditya Banerjee, Maciej Koch-Janusz, and Zohar Ringel. Relevance in the Renormalization Group and in Information Theory. *Physical Review Letters*, 126(24): 240601, June 2021. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.126.240601. URL `https://link.aps.org/doi/10.1103/PhysRevLett.126.240601`.

LP Gorkov. Forces acting on a small particle in an acoustic field within an ideal fluid. *Sov. Phys. Doklady*, 6:773–775, 1962.

Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Phase diagrams with real-space mutual information neural estimation. *arXiv:2103.16887 [cond-mat]*, April 2021. URL `http://arxiv.org/abs/2103.16887`. arXiv: 2103.16887.

Jan Haeberle, André Schella, Matthias Sperl, Matthias Schröter, and Philip Born. Double origin of stochastic granular tribocharging. *Soft Matter*, 14(24):4987–4995, 2018.

Taiki Haga. *Renormalization Group Analysis of Nonequilibrium Phase Transitions in Driven Disordered Systems*. Springer Theses. Springer Singapore, Singapore, 2019. ISBN 9789811361708 9789811361715. doi:10.1007/978-981-13-6171-5. URL `http://link.springer.com/10.1007/978-981-13-6171-5`.

Isaac A. Harris, Melody X. Lim, and Heinrich M. Jaeger. Temperature dependence of nylon and ptfe triboelectrification. *Phys. Rev. Materials*, 3:085603, Aug 2019. doi:10.1103/PhysRevMaterials.3.085603. URL `https://link.aps.org/doi/10.1103/PhysRevMaterials.3.085603`.

G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, July 2006. doi:10.1126/science.1127647. URL `https://www.science.org/doi/full/10.1126/science.1127647`. Publisher: American Association for the Advancement of Science.

Brad K Hulse, Hannah Haberkern, Romain Franconville, Daniel Turner-Evans, Shin-ya Takemura, Tanya Wolff, Marcella Noorman, Marisa Dreher, Chuntao Dan, Ruchi Parekh, Ann M Hermundstad, Gerald M Rubin, and Vivek Jayaraman. A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife*, 10:e66039. ISSN 2050-084X. doi:10.7554/eLife.66039. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9477501/`.

E. T. Jaynes. Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620–630, May 1957. ISSN 0031-899X. doi:10.1103/PhysRev.106.620. URL `https://link.aps.org/doi/10.1103/PhysRev.106.620`.

Patrick Jentsch and Chiu Fan Lee. Critical phenomena in compressible polar active fluids: Dynamical and functional renormalization group studies. *Physical Review Research*, 5(2): 023061, April 2023. doi:10.1103/PhysRevResearch.5.023061. URL `https://link.aps.org/doi/10.1103/PhysRevResearch.5.023061`. Publisher: American Physical Society.

G. Jona-Lasinio. The renormalization group: A probabilistic view. *Il Nuovo Cimento B Series 11*, 26(1):99–119, March 1975. ISSN 1826-9877. doi:10.1007/BF02755540. URL http://link.springer.com/10.1007/BF02755540.

G. Jona-Lasinio. Renormalization group and probability theory. *Physics Reports*, 352(4-6):439–458, October 2001. ISSN 03701573. doi:10.1016/S0370-1573(01)00042-4. URL https://linkinghub.elsevier.com/retrieve/pii/S0370157301000424.

Felix Jungmann, Tobias Steinpilz, Jens Teiser, and Gerhard Wurm. Sticking and restitution in collisions of charged sub-mm dielectric grains. *Journal of Physics Communications*, 2 (9):095009, 2018.

Leo P. Kadanoff. Scaling laws for ising models near T c. *Physics Physique Fizika*, 2(6): 263–272, June 1966. ISSN 0554-128X. doi:10.1103/PhysicsPhysiqueFizika.2.263. URL https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.2.263.

Sergei V. Kalinin, Ondrej Dyck, Stephen Jesse, and Maxim Ziatdinov. Exploring order parameters and dynamic processes in disordered systems via variational autoencoders. *Science Advances*, 7(17):eabd5084, April 2021. doi:10.1126/sciadv.abd5084. URL https://www.science.org/doi/10.1126/sciadv.abd5084. Publisher: American Association for the Advancement of Science.

Adam G Kline and Stephanie E Palmer. Gaussian information bottleneck and the non-perturbative renormalization group. *New Journal of Physics*, 24(3):033007, March 2022. ISSN 1367-2630. doi:10.1088/1367-2630/ac395d. URL https://iopscience.iop.org/article/10.1088/1367-2630/ac395d.

Maciej Koch-Janusz and Zohar Ringel. Mutual information, neural networks and the renormalization group. *Nature Physics*, 14(6):578–582, June 2018. ISSN 1745-2473, 1745-2481. doi:10.1038/s41567-018-0081-4. URL http://www.nature.com/articles/s41567-018-0081-4.

Artemy Kolchinsky, Brendan D. Tracey, and David H. Wolpert. Nonlinear Information Bottleneck. *Entropy*, 21(12):1181, November 2019. ISSN 1099-4300. doi:10.3390/e21121181. URL https://www.mdpi.com/1099-4300/21/12/1181.

Peter Kopietz, Lorenz Bartosch, and Florian Schütz. *Introduction to the Functional Renormalization Group*, volume 798 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-05093-0 978-3-642-05094-7. doi:10.1007/978-3-642-05094-7. URL http://link.springer.com/10.1007/978-3-642-05094-7.

Daniel J Lacks and Artem Levandovsky. Effect of particle size distribution on the polarity of triboelectric charging in granular insulator systems. *Journal of Electrostatics*, 65(2): 107–112, 2007.

Vincent Lahoche, Dine Ousmane Samary, and Mohamed Tamaazousti. Generalized scale behavior and renormalization group for data analysis. *arXiv:2002.10574 [cond-mat,* *physics:hep-th]*, April 2021. URL http://arxiv.org/abs/2002.10574. arXiv: 2002.10574.

Keirnan R LaMarche, Xue Liu, Shejal K Shah, Troy Shinbrot, and Benjamin J Glasser. Electrostatic charging during the flow of grains from a cylinder. *Powder Technology*, 195 (2):158–165, 2009.

LD Landau, EM Lifshitz, JB Sykes, WH Reid, and Ellis H Dill. *Theory of elasticity: Vol. 7 of Course of Theoretical Physics*. Pergamon Press, 1975.

Victor Lee, Scott R Waitukaitis, Marc Z Miskin, and Heinrich M Jaeger. Direct observation of particle interactions and clustering in charged granular streams. *Nature Physics*, 11(9): 733, 2015.

Victor Lee, Nicole M James, Scott R Waitukaitis, and Heinrich M Jaeger. Collisional charging of individual submillimeter particles: Using ultrasonic levitation to initiate and track charge transfer. *Physical Review Materials*, 2(3):035602, 2018.

Patrick M. Lenggenhager, Doruk Efe Gökmen, Zohar Ringel, Sebastian D. Huber, and Maciej Koch-Janusz. Optimal Renormalization Group Transformation from Information Theory. *Physical Review X*, 10(1):011037, February 2020. ISSN 2160-3308. doi:10.1103/PhysRevX.10.011037. URL http://arxiv.org/abs/1809.09632. arXiv: 1809.09632.

A. Levina, J. M. Herrmann, and T. Geisel. Dynamical synapses causing self-organized criticality in neural networks. *Nature Physics*, 3(12):857–860, December 2007. ISSN 1745-2473, 1745-2481. doi:10.1038/nphys758. URL http://www.nature.com/articles/nphys758.

Melody X Lim, Kieran A Murphy, and Heinrich M Jaeger. Edges control clustering in levitated granular matter. *Granular Matter*, 21(3):77, 2019a. ISSN 1434-7636.

Melody X Lim, Anton Souslov, Vincenzo Vitelli, and Heinrich M Jaeger. Cluster formation by acoustic forces and active fluctuations in levitated granular matter. *Nature Physics*, 15 (5):460, 2019b.

Daniel F. Litim. Optimisation of the exact renormalisation group. *Physics Letters B*, 486 (1-2):92–99, July 2000. ISSN 03702693. doi:10.1016/S0370-2693(00)00748-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0370269300007486.

Daniel F. Litim. Optimized renormalization group flows. *Physical Review D*, 64(10):105007, October 2001. ISSN 0556-2821, 1089-4918. doi:10.1103/PhysRevD.64.105007. URL https://link.aps.org/doi/10.1103/PhysRevD.64.105007.

Belle Liu, Arthur Hong, Fred Rieke, and Michael B. Manookin. Predictive encoding of motion begins in the primate retina. *Nature Neuroscience*, pages 1–12, August 2021. ISSN 1546-1726. doi:10.1038/s41593-021-00899-1. URL https://www.nature

.com/articles/s41593-021-00899-1. Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Motion detection;Neural circuits;Retina;Sensory processing Subject_term_id: motion;neural-circuit;retina;sensory-processing.

Chongyang Liu and Allen J Bard. Electrostatic electrochemistry at insulators. *Nature Materials*, 7(6):505, 2008.

Zhiru Liu and Benjamin H. Good. Dynamics of bacterial recombination in the human gut microbiome, June 2023. URL https://www.biorxiv.org/content/10.1101/2022.08.24.505183v2. Pages: 2022.08.24.505183 Section: New Results.

Zi-Kui Liu, Bing Li, and Henry Lin. Multiscale Entropy and Its Implications to Critical Phenomena, Emergent Behaviors, and Information. *Journal of Phase Equilibria and Diffusion*, 40(4):508–521, August 2019. ISSN 1863-7345. doi:10.1007/s11669-019-00736-w. URL https://doi.org/10.1007/s11669-019-00736-w.

Zi-Kui Liu, Yi Wang, and Shun-Li Shang. Zentropy Theory for Positive and Negative Thermal Expansion. *Journal of Phase Equilibria and Diffusion*, 43(6):598–605, December 2022. ISSN 1863-7345. doi:10.1007/s11669-022-00942-z. URL https://doi.org/10.1007/s11669-022-00942-z.

J Lowell and WS Truscott. Triboelectrification of identical insulators. ii. theory and further experiments. *Journal of Physics D: Applied Physics*, 19(7):1281, 1986.

B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna. Parameter Space Compression Underlies Emergent Theories and Predictive Models. *Science*, 342(6158):604–607, November 2013. ISSN 0036-8075, 1095-9203. doi:10.1126/science.1238723. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1238723.

Asier Marzo, Sue Ann Seah, Bruce W Drinkwater, Deepak Ranjan Sahoo, Benjamin Long, and Sriram Subramanian. Holographic acoustic elements for manipulation of levitated objects. *Nature Communications*, 6:8661, 2015.

Asier Marzo, Mihai Caleap, and Bruce W Drinkwater. Acoustic virtual vortices with tunable orbital angular momentum for trapping of mie particles. *Physical Review Letters*, 120(4):044301, 2018.

Logan S McCarty and George M Whitesides. Electrostatic charging due to separation of ions at interfaces: contact electrification of ionic electrets. *Angewandte Chemie International Edition*, 47(12):2188–2207, 2008.

Amit Mehrotra, Fernando J Muzzio, and Troy Shinbrot. Spontaneous separation of charged grains. *Physical Review Letters*, 99(5):058001, 2007.

Pankaj Mehta and David J. Schwab. An exact mapping between the Variational Renormalization Group and Deep Learning, October 2014. URL http://arxiv.org/abs/1410.3831. arXiv:1410.3831 [cond-mat, stat].

Kai Melde, Andrew G Mark, Tian Qiu, and Peer Fischer. Holograms for acoustics. *Nature*, 537(7621):518, 2016.

Leenoy Meshulam and William Bialek. Statistical mechanics for networks of real neurons, August 2024. URL `http://arxiv.org/abs/2409.00412`. arXiv:2409.00412 [cond-mat].

Leenoy Meshulam, Jeffrey L. Gauthier, Carlos D. Brody, David W. Tank, and William Bialek. Coarse Graining, Fixed Points, and Scaling in a Large Population of Neurons. *Physical Review Letters*, 123(17):178103, October 2019. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.123.178103. URL `https://link.aps.org/doi/10.1103/PhysRevLett.123.178103`.

Thierry Mora and William Bialek. Are Biological Systems Poised at Criticality? *Journal of Statistical Physics*, 144(2):268–302, July 2011. ISSN 0022-4715, 1572-9613. doi:10.1007/s10955-011-0229-4. URL `http://link.springer.com/10.1007/s10955-011-0229-4`.

Thierry Mora, Aleksandra M. Walczak, William Bialek, and Curtis G. Callan. Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences*, 107(12):5405–5410, March 2010. doi:10.1073/pnas.1001705107. URL `https://www.pnas.org/doi/full/10.1073/pnas.1001705107`. Publisher: Proceedings of the National Academy of Sciences.

Thierry Mora, Stéphane Deny, and Olivier Marre. Dynamical Criticality in the Collective Activity of a Population of Retinal Neurons. *Physical Review Letters*, 114(7):078105, February 2015. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.114.078105. URL `https://link.aps.org/doi/10.1103/PhysRevLett.114.078105`.

Mia C. Morrell, Audrey J. Sederberg, and Ilya Nemenman. Latent Dynamical Variables Produce Signatures of Spatiotemporal Criticality in Large Biological Systems. *Physical Review Letters*, 126(11):118302, March 2021. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.126.118302. URL `https://link.aps.org/doi/10.1103/PhysRevLett.126.118302`.

Mia C Morrell, Ilya Nemenman, and Audrey Sederberg. Neural criticality from effective latent variables. *eLife*, 12:RP89337, March 2024. ISSN 2050-084X. doi:10.7554/eLife.89337. URL `https://doi.org/10.7554/eLife.89337`. Publisher: eLife Sciences Publications, Ltd.

Vudtiwat Ngampruetikorn, Ilya Nemenman, and David J. Schwab. Extrinsic vs Intrinsic Criticality in Systems with Many Components, September 2023. URL `http://arxiv.org/abs/2309.13898`. arXiv:2309.13898 [cond-mat, physics:physics, q-bio].

Thomas Pähtz, Hans J Herrmann, and Troy Shinbrot. Why do particle clouds generate electric charges? *Nature Physics*, 6(5):364, 2010.

Stephanie E. Palmer, Olivier Marre, Michael J. Berry, and William Bialek. Predictive information in a sensory population. *Proceedings of the National Academy of Sciences*, 112(22): 6908–6913, June 2015. doi:10.1073/pnas.1506855112. URL https://www.pnas.org/doi/full/10.1073/pnas.1506855112. Publisher: Proceedings of the National Academy of Sciences.

Pedro Pessoa and Ariel Caticha. Exact Renormalization Groups As a Form of Entropic Dynamics. *Entropy*, 20(1):25, January 2018. doi:10.3390/e20010025. URL https://www.mdpi.com/1099-4300/20/1/25. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

Joseph Polchinski. Renormalization and effective lagrangians. *Nuclear Physics B*, 231(2): 269–295, January 1984. ISSN 05503213. doi:10.1016/0550-3213(84)90287-6. URL https://linkinghub.elsevier.com/retrieve/pii/0550321384902876.

Ben Poole, Sherjil Ozair, Aaron van den Oord, Alexander A. Alemi, and George Tucker. On Variational Bounds of Mutual Information, May 2019. URL http://arxiv.org/abs/1905.06922. arXiv:1905.06922 [cs].

Torsten Poppe, Jürgen Blum, and Thomas Henning. Experiments on collisional grain charging of micron-sized preplanetary dust. *Astrophysical Journal*, 533(1):472, 2000.

Jason S. Prentice, Olivier Marre, Mark L. Ioffe, Adrianna R. Loback, Gašper Tkačik, and Michael J. Berry Ii. Error-Robust Modes of the Retinal Population Code. *PLOS Computational Biology*, 12(11):e1005148, November 2016. ISSN 1553-7358. doi:10.1371/journal.pcbi.1005148. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005148. Publisher: Public Library of Science.

Katherine N. Quinn, Michael C. Abbott, Mark K. Transtrum, Benjamin B. Machta, and James P. Sethna. Information geometry for multiparameter models: New perspectives on the origin of simplicity. September 2022. URL http://arxiv.org/abs/2111.07176. arXiv:2111.07176 [cond-mat, physics:physics].

Archishman Raju, Benjamin B. Machta, and James P. Sethna. Information loss under coarse graining: A geometric approach. *Physical Review E*, 98(5):052112, November 2018. ISSN 2470-0045, 2470-0053. doi:10.1103/PhysRevE.98.052112. URL https://link.aps.org/doi/10.1103/PhysRevE.98.052112.

Luisa Ramirez and William Bialek. Compression as a path to simplification: Models of collective neural activity, December 2021. URL http://arxiv.org/abs/2112.14334. arXiv:2112.14334 [cond-mat, q-bio].

Nicole C. Rust and Stephanie E. Palmer. Remembering the Past to See the Future. *Annual review of vision science*, 7:349, July 2021. doi:10.1146/annurev-vision-093019-112249. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9751846/.

Jared Salisbury and Stephanie E. Palmer. Optimal prediction and natural scene statistics in the retina. *arXiv:1507.00125 [q-bio]*, July 2015. URL http://arxiv.org/abs/1507.00125. arXiv: 1507.00125.

Jared M. Salisbury and Stephanie E. Palmer. Optimal Prediction in the Retina and Natural Motion Statistics. *Journal of Statistical Physics*, 162(5):1309–1323, March 2016. ISSN 0022-4715, 1572-9613. doi:10.1007/s10955-015-1439-y. URL http://link.springer.com/10.1007/s10955-015-1439-y.

Andrew M Saxe, Yamini Bansal, Joel Dapello, Madhu Advani, Artemy Kolchinsky, Brendan D Tracey, and David D Cox. On the information bottleneck theory of deep learning. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124020, December 2019. ISSN 1742-5468. doi:10.1088/1742-5468/ab3985. URL https://iopscience.iop.org/article/10.1088/1742-5468/ab3985.

Matthew S. Schmitt, Maciej Koch-Janusz, Michel Fruchart, Daniel S. Seara, and Vincenzo Vitelli. Information theory for model reduction in stochastic dynamical systems, December 2023. URL http://arxiv.org/abs/2312.06608. arXiv:2312.06608 [cond-mat].

Elad Schneidman, Michael J. Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440 (7087):1007–1012, April 2006. ISSN 0028-0836, 1476-4687. doi:10.1038/nature04701. URL http://www.nature.com/articles/nature04701.

David J. Schwab, Ilya Nemenman, and Pankaj Mehta. Zipf's Law and Criticality in Multivariate Data without Fine-Tuning. *Physical Review Letters*, 113(6):068102, August 2014. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.113.068102. URL https://link.aps.org/doi/10.1103/PhysRevLett.113.068102.

Audrey J. Sederberg, Jason N. MacLean, and Stephanie E. Palmer. Learning to make external sensory stimulus predictions using internal correlations in populations of neurons. *Proceedings of the National Academy of Sciences*, 115(5):1105–1110, January 2018. doi:10.1073/pnas.1710779115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1710779115. Publisher: Proceedings of the National Academy of Sciences.

Mikkel Settnes and Henrik Bruus. Forces acting on a small particle in an acoustical field in a viscous fluid. *Physical Review E*, 85(1):016327, 2012.

Troy Shinbrot, Behrooz Ferdowsi, Sankaran Sundaresan, and Nuno AM Araujo. Multiple timescale contact charging. *Physical Review Materials*, 2(12):125003, 2018.

Ravid Shwartz-Ziv and Naftali Tishby. Opening the Black Box of Deep Neural Networks via Information. *arXiv:1703.00810 [cs]*, April 2017. URL http://arxiv.org/abs/1703.00810. arXiv: 1703.00810.

Adrien Six, Maria Encarnita Mariotti-Ferrandiz, Wahiba Chaara, Susana Magadan, Hang-Phuong Pham, Marie-Paule Lefranc, Thierry Mora, Véronique Thomas-Vaslin, Aleksandra M. Walczak, and Pierre Boudinot. The past, present, and future of immune repertoire biology–the rise of next-generation repertoire analysis. *Frontiers in immunology*, 4:413, 2013. URL `https://www.frontiersin.org/articles/10.3389/fimmu.2013.00413/full`. Publisher: Frontiers Media SA.

Noam Slonim and Naftali Tishby. Agglomerative Information Bottleneck. page 7, 2000.

Noam Slonim, Nir Friedman, and Naftali Tishby. Multivariate Information Bottleneck. *Neural Computation*, 18(8):1739–1789, August 2006. ISSN 0899-7667, 1530-888X. doi:10.1162/neco.2006.18.8.1739. URL `https://direct.mit.edu/neco/article/18/8/1739-1789/7079`.

Andrew Sowinski, Fawzi Salama, and Poupak Mehrani. New technique for electrostatic charge measurement in gas–solid fluidized beds. *Journal of Electrostatics*, 67(4):568–573, 2009.

CD Stow. Dust and sand storm electrification. *Weather*, 24(4):134–144, 1969.

Charlotte Strandkvist, Pavel Chvykov, and Mikhail Tikhonov. Beyond RG: from parameter flow to metric flow, November 2020. URL `http://arxiv.org/abs/2011.12420`. arXiv:2011.12420 [cond-mat].

S. P. Strong, Roland Koberle, Rob R. de Ruyter van Steveninck, and William Bialek. Entropy and Information in Neural Spike Trains. *Physical Review Letters*, 80(1):197–200, January 1998. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.80.197. URL `https://link.aps.org/doi/10.1103/PhysRevLett.80.197`.

Dj Strouse and David J. Schwab. The Deterministic Information Bottleneck. *Neural Computation*, 29(6):1611–1630, June 2017. ISSN 0899-7667, 1530-888X. doi:10.1162/NECO_a_00961. URL `https://direct.mit.edu/neco/article/29/6/1611-1630/8273`.

Lorenzo Tiberi, Jonas Stapmanns, Tobias Kühn, Thomas Luu, David Dahmen, and Moritz Helias. Gell-Mann–Low Criticality in Neural Networks. *Physical Review Letters*, 128(16): 168301, April 2022. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.128.168301. URL `https://link.aps.org/doi/10.1103/PhysRevLett.128.168301`.

SP Timoshenko and JN Goodier. *Theory of Elasticity*. McGraw-Hill Book Company, 1970.

Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW)*, pages 1–5, Jerusalem, Israel, April 2015. IEEE. ISBN 978-1-4799-5524-4 978-1-4799-5526-8. doi:10.1109/ITW.2015.7133169. URL `http://ieeexplore.ieee.org/document/7133169/`.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. April 2000. doi:10.48550/arXiv.physics/0004057. URL http://arxiv.org/abs/physics/0004057. arXiv:physics/0004057.

Gašper Tkačik, Olivier Marre, Thierry Mora, Dario Amodei, Michael J Berry II, and William Bialek. The simplest maximum entropy model for collective behavior in a neural network. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03011, March 2013. ISSN 1742-5468. doi:10.1088/1742-5468/2013/03/P03011. URL https://iopscience.iop.org/article/10.1088/1742-5468/2013/03/P03011.

Gašper Tkačik, Olivier Marre, Dario Amodei, Elad Schneidman, William Bialek, and Michael J. Berry. Searching for Collective Behavior in a Large Network of Sensory Neurons. *PLoS Computational Biology*, 10(1):e1003408, January 2014. ISSN 1553-7358. doi:10.1371/journal.pcbi.1003408. URL https://dx.plos.org/10.1371/journal.pcbi.1003408.

Gašper Tkačik, Thierry Mora, Olivier Marre, Dario Amodei, Stephanie E. Palmer, Michael J. Berry, and William Bialek. Thermodynamics and signatures of criticality in a network of neurons. *Proceedings of the National Academy of Sciences*, 112(37):11508–11513, September 2015. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.1514188112. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1514188112.

Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478 (1-3):1–69, July 2009. ISSN 03701573. doi:10.1016/j.physrep.2009.05.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0370157309001410.

Hugo Touchette. A basic introduction to large deviations: Theory, applications, simulations. *arXiv:1106.4146 [cond-mat, physics:math-ph]*, February 2012. URL http://arxiv.org/abs/1106.4146. arXiv: 1106.4146.

Scott R Waitukaitis, Victor Lee, James M Pierson, Steven L Forman, and Heinrich M Jaeger. Size-dependent same-material tribocharging in insulating grains. *Physical Review Letters*, 112(21):218001, 2014.

Siwei Wang, Benjamin Hoshal, Elizabeth de Laittre, Thierry Mora, Michael Berry, and Stephanie Palmer. Learning low-dimensional generalizable natural features from retina using a U-net. *Advances in Neural Information Processing Systems*, 35:11355–11368, December 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/49d608425f1bee2864e034a9e9e1ec9e-Abstract-Conference.html.

Tian Wang, Manzhu Ke, Weiping Li, Qian Yang, Chunyin Qiu, and Zhengyou Liu. Particle manipulation with acoustic vortex beam induced by a brass plate with spiral shape structure. *Applied Physics Letters*, 109(12):123506, 2016.

Yihang Wang, João Marcelo Lamim Ribeiro, and Pratyush Tiwary. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nature Communications*, 10(1):3573, December 2019. ISSN 2041-1723.

doi:10.1038/s41467-019-11405-4. URL http://www.nature.com/articles/s41467-019-11405-4.

Hideo Watanabe, Abdolreza Samimi, Yu Long Ding, Mojtaba Ghadiri, Tatsushi Matsuyama, and Kendal G Pitt. Measurement of charge transfer due to single particle impact. *Particle & Particle Systems Characterization*, 23(2):133–137, 2006.

Franz J. Wegner and Anthony Houghton. Renormalization Group Equation for Critical Phenomena. *Physical Review A*, 8(1):401–412, July 1973. ISSN 0556-2791. doi:10.1103/PhysRevA.8.401. URL https://link.aps.org/doi/10.1103/PhysRevA.8.401.

Christof Wetterich. The average action for scalar fields near phase transitions. *Zeitschrift fr Physik C Particles and Fields*, 57(3):451–469, September 1993a. ISSN 0170-9739, 1434-6052. doi:10.1007/BF01474340. URL http://link.springer.com/10.1007/BF01474340.

Christof Wetterich. Exact evolution equation for the effective potential. *Physics Letters B*, 301(1):90–94, February 1993b. ISSN 03702693. doi:10.1016/0370-2693(93)90726-X. URL http://arxiv.org/abs/1710.05815. arXiv: 1710.05815.

K Wilson. The renormalization group and the expansion. *Physics Reports*, 12(2):75–199, August 1974. ISSN 03701573. doi:10.1016/0370-1573(74)90023-4. URL https://linkinghub.elsevier.com/retrieve/pii/0370157374900234.

Kenneth G. Wilson. Renormalization Group and Critical Phenomena. II. Phase-Space Cell Analysis of Critical Behavior. *Physical Review B*, 4(9):3184–3205, November 1971a. ISSN 0556-2805. doi:10.1103/PhysRevB.4.3184. URL https://link.aps.org/doi/10.1103/PhysRevB.4.3184.

Kenneth G. Wilson. Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture. *Physical Review B*, 4(9):3174–3183, November 1971b. ISSN 0556-2805. doi:10.1103/PhysRevB.4.3174. URL https://link.aps.org/doi/10.1103/PhysRevB.4.3174.

Kenneth G. Wilson and Michael E. Fisher. Critical Exponents in 3.99 Dimensions. *Physical Review Letters*, 28(4):240–243, January 1972. ISSN 0031-9007. doi:10.1103/PhysRevLett.28.240. URL https://link.aps.org/doi/10.1103/PhysRevLett.28.240.

L Xie, N Bao, Y Jiang, K Han, and J Zhou. An instrument for charge measurement due to a single collision between two spherical particles. *Review of Scientific Instruments*, 87(1): 014705, 2016.

Arnold Zellner. Optimal Information Processing and Bayes's Theorem. *The American Statistician*, 42(4):278–280, November 1988. ISSN 0003-1305. doi:10.1080/00031305.1988.10475585. URL https://www.tandfonline.com/do i/abs/10.1080/00031305.1988.10475585. Publisher: ASA Website _eprint: https://www.tandfonline.com/doi/pdf/10.1080/00031305.1988.10475585.

Yanzhen Zhang, Thomas Pähtz, Yonghong Liu, Xiaolong Wang, Rui Zhang, Yang Shen, Renjie Ji, and Baoping Cai. Electric field and humidity trigger contact electrification. *Physical Review X*, 5(1):011002, 2015.

Cheyenne Ziegler, Jonathan Martin, Claude Sinner, and Faruck Morcos. Latent generative landscapes as maps of functional diversity in protein sequence space. *Nature Communications*, 14(1):2222, April 2023. ISSN 2041-1723. doi:10.1038/s41467-023-37958-z. URL https://www.nature.com/articles/s41467-023-37958-z. Number: 1 Publisher: Nature Publishing Group.