The Union of Quantum Field Theory and Non-equilibrium Thermodynamics

Thesis by
Anthony Bartolotta

In Partial Fulfillment of the Requirements for the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended May 24, 2018

Anthony Bartolotta
ORCID: 0000-0003-4971-9545

Acknowledgments

My time as a graduate student at Caltech has been a journey for me, both professionally and personally. This journey would not have been possible without the support of many individuals.

First, I would like to thank my advisors, Sean Carroll and Mark Wise. Without their support, this thesis would not have been written. Despite entering Caltech with weaker technical skills than many of my fellow graduate students, Mark took me on as a student and gave me my first project. Mark also granted me the freedom to pursue my own interests, which proved instrumental in my decision to work on non-equilibrium thermodynamics. I am deeply grateful for being provided this priviledge and for his continued input on my research direction. Sean has been an incredibly effective research advisor, despite being a newcomer to the field of non-equilibrium thermodynamics. Sean was the organizing force behind our first paper on this topic and connected me with other scientists in the broader community; at every step Sean has tried to smoothly transition me from the world of particle physics to that of non-equilibrium thermodynamics. My research would not have been nearly as fruitful without his support.

I would also like to thank the other two members of my thesis and candidacy committees, John Preskill and Keith Schwab. Both took time out of their busy schedules to attend my defenses and provide insightful comments on my research.

Throughout my graduate career, I had the support of a wonderful set of collaborators, without whom many projects would not have been possible. I would like to thank: Sean Carroll, Sebastian Deffner, Stefan Leichenauer, Dibyendu Mandal, Jason Pollack, Michael Ramsey-Musolf, and Nicole Yunger-Halpern.

Coming from a particle physics background, I am an interloper in the non-equilibrium thermodynamics community. Despite this fact, I was quickly welcomed by the community. I would like to thank the many participants of the annual "Information Engines at the Frontiers of Nanoscale Thermodynamics" workshop at the Telluride Science Research Conference, for providing such a welcoming environment and making me feel at home in a field that was not my own. While I cannot thank everyone individually, I would like to extend my deep graditude to Sebastian Deffner. Without Sebastian, I never would have met the wider non-equilibrium thermodynamics community. Fur-

thermore, Sebastian has been an indispensible source of career advice and support as I transition out of graduate school.

I would like to thank the wonderful community in the Caltech High-Energy Theory group. Clifford Cheung for answering all of my questions about the analytic properties of field theories, and Carol Silberstein for maintaining a semblance of order in the group. I would also like to thank the many PhD students and postdocs in the group for making my time here far more enjoyable: Ning Bao, Charles Cao, Aidan Chatwin-Davies, Enrico Herrmann, Matt Heydeman, Murat Kologlu, Stefan Leichenauer, Jason Pollack, Grant Remmen, Alexander Ridgway, Chia-Hsien Shen, and Ashmeet Singh. I would especially like to thank my long-time officemate Jason Pollack for putting up with my incessant stream of physics questions, and more importantly, for not completely giving up on me after I accidentally glued myself to the printer.

For keeping me grounded, and constantly foiling my plans as DM, I would like to thank the Friday night dinner and D&D group: Dustin Anderson, Kevin Barkett, Daniel Brooks, Jeremy Brouillet, and Jason Pollack.

For reminding me that there is a world outside of physics and that it is not the only possible topic of conversation, I would like to thank my friends in other departments: Ania Baetica, Thomas Catanach, Niangjun Chen, Albert Chern, Ioannis Filippidis, Yorie Nakahira, and Yong Sheng Soh.

For supporting my decision to pursue a career in science since before I can remember, I would like to thank my parents, Gene and Andrea Bartolotta, and brother, Vince Bartolotta. I would also like to express my deep gratitude to my parents for their continued support in spite of the damage I caused to the house and yard with my "experiments".

Finally, I would like to thank my incredible wife, Yoke Peng Leong. Without her support (and occasional nagging to take a break), graduate school would have been a far different experience. Even in the darkest times of graduate school, Yoke Peng supported me and was a calm voice of reason to my anxiety. I am deeply grateful for her help in getting though graduate school and cannot imagine anyone else I would rather have at my side going forward.

My work was supported in part by the Walter Burke Institute for Theoretical Physics at Caltech and by the United States Department of Energy under Grant DE-SC0011632.

Abstract

Quantum field theory is the language used to describe nature at its most fundamental scales; while thermodynamics is a framework to describe the collective behavior of macroscopic systems. Recent advances in non-equilibrium thermodynamics have enabled this framework to be applied to smaller systems operating out of thermal equilibrium. This thesis is concerned with both quantum field theory and non-equilibrium thermodynamics independently and with their intersection.

First, a purely phenomenological application of quantum field theory is explored in the context of the upcoming Mu2E experiment. This experiment will look for rare decays which would indicate the presence of physics beyond the Standard Model. Using the language of effective field theories, a next-to-leading order analysis of the conversion rate is performed.

The focus then shifts to an apparent paradox in the Bayesian interpretation of statistical mechanics. For a Bayesian observer making measurements of an open system, the Shannon entropy decreases, in apparent violation of the Second Law of Thermodynamics. It is shown that rather than utilizing the entropy, which can decrease under Bayesian updates, the Second Law for a Bayesian observer can be rephrased in terms of a cross-entropy which is always non-negative.

Finally, the intersection of quantum field theory and non-equilibrium thermodynamics is examined. Using quantum work fluctuation theorems, an investigation of how these frameworks can be applied to a driven quantum field theory is performed. For a time-dependent variant of $\lambda\phi^4$, analytic expressions for the work distribution functions at one-loop order are derived. These expressions are shown to satisfy the quantum Jarzynski equality and Crooks fluctuation theorem.

Published Content and Contributions

With the exception of the Introduction, the body of this thesis is adapted from articles that originally appeared in other forms. All articles are available as arXiv preprints. All authors contributed equally.

- A. Bartolotta and M. J. Ramsey-Musolf, *Coherent $\mu - e$ Conversion at Next-to-Leading Order*, arXiv:1710.0212
- A. Bartolotta, S. M. Carroll, S. Leichenauer, and J. Pollack, *Bayesian second law of thermodynamics*, *Phys. Rev. E* **94** (Aug, 2016) 022102, [arXiv:1508.0242]
- A. Bartolotta and S. Deffner, *Jarzynski equality for driven quantum field theories*, *Phys. Rev. X* **8** (Feb, 2018) 011033, [arXiv:1710.0082]

Contents

List of Figures

List of Tables

Introduction

Two of the main threads in modern theoretical physics are the pursuit of theories that describe the most fundamental components of nature, and theories that describe the collective behavior of systems of many such components. The former is the domain of quantum field theory and particle physics, while the latter belongs to statistical mechanics and thermodynamics.

For most of the 19th century, microscopic theories of physics were based entirely on classical mechanics. However, near the turn of the 20th century, a series of experimental results, including the photoelectric effect and Rutherford's discovery of the atomic nucleus, made apparent the need for a new microphysical theory. It was found that the behavior of these systems could be explained by quantizing the allowed energy levels. This was the advent of modern quantum mechanics. Within the domain of atomic scale experiments, quantum mechanics was met with great success. However, as physicists attempted to probe increasingly high energy scales throughout the 20th century, it became necessary to unify quantum mechanics and the theory of special relativity. These unified frameworks yielded quantum field theory, the cornerstone of modern theoretical physics [4–6].

In quantum field theory, particles are described as the excitations of fields. This general framework is flexible enough to describe both the creation and annihilation of particles in accelerators, and the excitation of vibrational and electrical modes in condensed matter systems [7–9]. Across this tremendous range of energy scales, quantum field theory has been rigorously tested. Arguably, the most stringently tested theory in modern physics is the Standard Model of particle physics.

Developed throughout the 1960s [10–14], the Standard Model is a quantum field theory that economically describes all of modern particle physics with twelve fundamental fermions, four gauge bosons, and the Higgs boson. Since the Large Hadron Collider's discovery of the Higgs boson [15, 16], all of the fundamental particles proposed by the Standard Model have been experimentally observed. Possible physics beyond the Standard Model has been strongly constrained by precision measurements of the electron and muon anomalous magnetic dipole moments [17–20], and the non-observation of a neutron electric dipole moment [21–23], flavor changing neutral currents [24–28], or proton decay [29, 30]. Despite its tremendous successes, the Standard Model still has significant shortcomings. These include the failure to account for the observation of

dark matter, the cosmic matter-antimatter asymmetry, or the accelerating expansion of the universe. This is in addition to theoretical shortcomings such as the hierarchy problem. These limitations have resulted in various proposals to extend the Standard Model, though the Standard Model remains the most complete description of nature to date.

In comparison to the progress over the past two centuries in understanding the fundamental theory of nature, the description of the collective behavior of systems, thermodynamics, has languished. Modern thermodynamics has its roots in the industrial revolution of the 19th century; its development necessitated by the invention of steam engines. Thermodynamics provides a framework to understand the relationships between macrocopic quantities of interest such as work, pressure, temperature, and entropy with minimal assumptions on the underlying microscopic dynamics of the system. Famously, Carnot proved an upper limit on the efficiency of any heat engine which depends only on the temperatures of its working reservoirs [31]. However, this bound can only be saturated in the limit of an infinitely slow engine which operates in constant thermal equilibrium. For systems working in finite time, or out of equilibrium, traditional thermodynamics provides no additional insight and is only able to place weak bounds on the system's behavior.

Parallel to the development of thermodynamics, the microscopic approach of statistical mechanics provided a rigorous underpinning for these results [32–34]. As microscopic theories progressed from classical mechanics to quantum mechanics and to quantum field theory, statistical mechanics generally followed suit. However, thermodynamics remained largely unchanged, focusing exclusively on systems in, or near, equilibrium.

Progress in the thermodynamics of non-equilibrium systems was generally slow throughout the 20th century. The primary advancements in the first half of the 20th century were the Onsager reciprocal relations [35], relating thermodynamic forces and flows in systems at local equilibrium; and the fluctuation-dissipation theorem [36, 37], relating thermal fluctuations to response functions for near-equilibrium systems in the linear response regime. It was only in the late 20th century that the first fluctuation theorems were discovered [38–42].

For realistic, finite-time processes, traditional thermodynamics is only able to state that on average the irreversible entropy production is positive, $\langle \Sigma \rangle \geq 0$. This is the well known Second Law of Thermodynamics. By contrast, the detailed fluctuation theorem is an equality which states that the system is exponentially unlikely to undergo a reduction in entropy, $\mathcal{P}(-\Sigma) = \exp(-\Sigma)\,\mathcal{P}(\Sigma)$ [41, 43, 44]. This statement holds for systems driven arbitrarily far from equilibrium and thus allows us to place powerful constraints on the behavior of non-equilibrium systems. Given the power of the classical

fluctuation theorems, the early 21st century has seen a plethora of proposed quantum extensions of the fluctuation theorems [45–64].

This thesis consists of three parts; touching on quantum field theory, non-equilibrium thermodynamics, and their intersection. These are organized as follows:

- Chapter 2 considers a purely phenomenological application of quantum field theory. The upcoming Mu2E experiment will look for coherent muon to electron conversion, a process which is permitted at unobservably small rates in the Standard Model. As such, any observation of coherent conversion will provide a window into physics beyond the Standard Model. A detailed analysis of the conversion process and associated uncertainties is performed at next-to-leading order. This analysis culminates in a simple expression relating the predicted conversion rate to the model-dependent Wilson coefficients through a collection of pre-computed model-independent parameters which include uncertainties.

- Chapter 3 turns its focus to non-equilibrium thermodynamics. The probability distribution over possible configurations of a system in phase space plays a central role in statistical mechanics and thermodynamics. In the Bayesian interpretation of probability, the probability distribution represents the knowledge of an observer. As an observer can potentially make measurements of the system and update their beliefs, the uncertainty in their distribution can decrease which corresponds to a decrease in entropy. This leads to an apparent contradiction of the Second Law and would imply that the arrow of time for a Bayesian observer points into the past. It is shown that rather than utilizing the entropy, which can decrease under Bayesian updates, the Second Law for a Bayesian observer can be rephrased in terms of a cross-entropy which is always non-negative. Modified versions of the classical fluctuation theorems for a Bayesian observer are also derived.

- Finally, Chapter 4 examines the intersection of quantum field theory and non-equilibrium thermodynamics. After the advent of the classical fluctuation theorems, quantum extensions were rapidly proposed. An investigation of how these frameworks can be applied to a driven quantum field theory is performed. For a time-dependent variant of $\lambda\phi^4$, analytic expressions for the work distribution functions at one-loop order are derived. Using these expressions, it can then be shown that the quantum Jarzynski equality and Crooks fluctuation theorem hold at one-loop order independent of the renormalization scale.

Coherent $\mu - e$ Conversion at Next-to-Leading Order

We analyze next-to-leading order (NLO) corrections and uncertainties for coherent $\mu-e$ conversion . The analysis is general, but numerical results focus on ^{27}Al, which will be used in the Mu2E experiment. We obtain a simple expression for the branching ratio in terms of Wilson coefficients associated with possible physics beyond the Standard Model and a set of model-independent parameters determined solely by Standard Model dynamics. For scalar-mediated conversion, we find that NLO two-nucleon contributions can significantly decrease the branching ratio, potentially reducing the rate by as much as 50%. The pion-nucleon σ-term and quark masses give the dominant sources of parametric uncertainty in this case. For vector-mediated conversion, the impact of NLO contributions is considerably less severe, while the present theoretical uncertainties are comparable to parametric uncertainties.

2.1 Introduction

Despite its many successes, the Standard Model (SM) has several phenomenological and theoretical shortcomings. Phenomenologically, the Standard Model provides no explanation for cosmic matter-antimatter asymmetry, the relic density of cold dark matter, or the dark energy associated with cosmic acceleration. The observation of neutrino oscillations requires extending the SM to account for non-zero neutrino masses. Theoretically, the SM suffers from a hierarchy problem, does not explain the quantization of electric charge, and simply parameterizes the vast range of elementary fermion masses, and an associated mixing between flavor and mass eigenstates.

The flavor problem remains, indeed, one of the most vexing. In the charged lepton sector, the presence of flavor mixing among the light neutrinos implies non-vanishing, though unobservably small, rates for flavor non-conserving processes, such as $\mu \to e\gamma$. Scenarios for physics beyond the Standard Model (BSM), however, can allow for significantly larger rates for such processes. The observation of charged lepton flavor violation (CLFV) may, thus, point to one or more of these proposals and shed new light on the flavor problem. This possibility motivates several current and future CLFV searches, such as the MEG experiment at the Paul Scherrer Institute (PSI) that has recently placed a limit of $< 4.2 \cdot 10^{-13}$ on the branching ratio for $\mu \to e\gamma$ [24]; the upcoming Mu2e and COMET experiments at Fermilab and J-PARC, respectively, which will search for CLFV through the process of coherent $\mu - e$ conversion in the presence of a nucleus [25, 26]; and the possible search for $\mu \to 3e$ at PSI. For recent experimental and theoretical reviews, see Refs. [27, 28]

In this study, we focus on the process of coherent $\mu - e$ conversion. The quantity of interest is the branching ratio

$$\mathrm{BR}(\mu - e) = \frac{\mu^- + A(Z,N) \to e^- + A(Z,N)}{\mu^- + A(Z,N) \to \nu_\mu + A(Z-1,N)} \,, \tag{2.1.1}$$

where the denominator is the rate for muon capture on a nucleus with Z protons and N neutrons with $A = Z + N$. The standard model branching ratio for this process is predicted to be of the order $\mathrm{BR}(\mu - e) \approx 10^{-54}$ [65, 66]. At present, the best experimental bounds are from the SINDRUM II collaboration which has constrained $\mathrm{BR}(\mu - e) < 7 \cdot 10^{-13}$ [25, 67]. The next generation experiments, Mu2e and COMET, are expected to improve these bounds by roughly four orders of magnitude, $\mathrm{BR}(\mu - e) \lesssim 5 \cdot 10^{-17}$ [25, 26].

Previous studies of coherent conversion have focused on leading order processes and their uncertainties [68–71]. The primary goal of this work is to extend the analysis of coherent conversion to include next-to-leading order (NLO) corrections and

their uncertainties. We focus primarily on phenomenological, dimension six effective semileptonic operators that may induce this CLFV conversion process. The framework of $SU(2)$ Chiral Perturbation Theory (ChPT) can then be used to relate operators in the phenomenological CLFV Lagrangian written in terms of quarks to the hadronic degrees of freedom relevant for nuclear physics dynamics. As the momentum transfer scale in coherent conversion is set by the muon mass and because the nucleons have no net strangeness, one might expect $SU(2)$ ChPT to be adequate for present purposes. However, CLFV operators involving strange quarks will still contribute to the conversion process. To assess the possible quantitative impact of these operators, we include their leading order contributions via $SU(2)$ flavor singlet terms in the chiral Lagrangian. Doing so is preferable to the use of full $SU(3)$ ChPT as it allows for better control of both theoretical uncertainties and uncertainties introduced by the low energy constants of the chiral Lagrangian as shown in Refs. [71, 72]. We find that the strange quark contributions are generally small compared to other theoretical and parametric uncertainties, as seen in Table 2.1. Thus, the use of $SU(2)$ ChPT in this context should be robust.

The primary results of this investigation are given in Eqs. (2.1.2,2.1.3,2.1.9) and Tables 2.1 and 2.2. We summarize these results here for convenience. The branching ratio for coherent conversion can be written as a sum of four separate amplitudes, one for each spin configuration of the system,

$$\mathrm{BR}(\mu - e)_{\mathcal{A}} = \left(\frac{v}{\Lambda}\right)^4 \left[\left|\tau_{\mathcal{A}}^{(1)}\right|^2 + \left|\tau_{\mathcal{A}}^{(2)}\right|^2 + \left|\tau_{\mathcal{A}}^{(3)}\right|^2 + \left|\tau_{\mathcal{A}}^{(4)}\right|^2\right]. \tag{2.1.2}$$

Here, $\mathcal{A} = S(V)$ indicates a scalar (vector)-mediated conversion process; $v = 246$ GeV, is the Higgs vacuum expectation value (VEV); Λ is the mass scale associated with the BSM CLFV dynamics; and the indices $w \in \{1, 2, 3, 4\}$ denote each unique configuration as defined in Appendix 2.F.

Within each conversion amplitude, it is possible to separate all model-independent parameters from the Wilson coefficients of the specific CLFV theory. Doing so for the case of scalar-mediated conversion yields

$$\begin{aligned}\left|\tau_S^{(w)}\right|^2 = \Bigg|&\alpha_{S,ud}^{(w)}\left(\frac{C_u^{S,L} \pm C_u^{S,R}}{2}\right) + \alpha_{S,ud}^{(w)}\left(\frac{C_d^{S,L} \pm C_d^{S,R}}{2}\right) \\ &+\alpha_{S,s}^{(w)}\left(\frac{C_s^{S,L} \pm C_s^{S,R}}{2}\right) + \alpha_{S,\Theta}^{(w)}\left(\frac{C_\Theta^{S,L} \pm C_\Theta^{S,R}}{2}\right)\Bigg|^2,\end{aligned} \tag{2.1.3}$$

where the $C_q^{S,L}$ ($C_q^{S,R}$) denote the Wilson coefficients for a scalar interaction involving a left- (right-) handed muon interacting with a light quark of flavor $q = (u, d, s)$ as defined in Eq. (2.2.5); where $C_\Theta^{S,L}$ ($C_\Theta^{S,R}$) give the corresponding heavy quark contributions

entering via the energy-momentum tensor; and where positive (negative) signs are used for $w \in \{1,3\}$ ($w \in \{2,4\}$). All model-independent parameters have been absorbed in the definitions of the α's. These parameters are defined in Appendix 2.G and their numerical values are given in Table 2.1.

Important for this work are the relative magnitudes of the LO, NLO one-loop, and NLO two-nucleon contributions for the scalar-mediated amplitudes. Each contribution contains a common factor of

$$\sqrt{\frac{m_\mu}{\omega_{\rm capt}}}\left(\frac{m_\mu}{4\pi v}\right)^2 = 0.5563 \pm 0.0005\ , \tag{2.1.4}$$

where $\omega_{\rm capt}$ is the muon capture rate. For u- and d-quarks, the LO contribution is obtained from the pion-nucleon σ-term

$$\alpha^{(1)}_{S,ud}({\rm LO}) = \sqrt{\frac{m_\mu}{\omega_{\rm capt}}}\left(\frac{m_\mu}{4\pi v}\right)^2 \frac{\sigma_{\pi N}}{2\hat{m}}\left(I^{(1)}_{S,p} + I^{(1)}_{S,n}\right) = 65 \pm 11\ , \tag{2.1.5}$$

where $\hat{m}$ is the average of u- and d-quark current masses, $\sigma_{\pi N}$ is the pion-nucleon σ-term, and the $I^{(1)}_{S,N}$ are integrals involving the overlap of incoming and outgoing lepton wave functions with the distributions of nucleons N. The definition of $I^{(1)}_{S,N}$ in terms of overlap integrals can be found in Appendix 2.F.

The NLO one-loop contribution is given by

$$-\alpha^{(1)}_{S,ud}({\rm NLO\ loop}) = \sqrt{\frac{m_\mu}{\omega_{\rm capt}}}\left(\frac{m_\mu}{4\pi v}\right)^2 \left(\frac{3B_0 m_\pi \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}\right) \Delta^{(1)}_S = 2.71 \pm 0.30\ , \tag{2.1.6}$$

where $B_0 = 2.75 \pm 0.11$ GeV normalizes the scalar source in the chiral Lagrangian (See Section 2.3 below); m_π and $\mathring{f}_\pi$ are the pion mass and LO pion decay constant; $\mathring{g}_A$ is the LO nucleon axial coupling; and

$$\Delta^{(1)}_S = \left(\tilde{I}^{(1)}_{S,p} + \tilde{I}^{(1)}_{S,n}\right) - \left(I^{(1)}_{S,p} + I^{(1)}_{S,n}\right) = 3.96 \pm 0.39\ , \tag{2.1.7}$$

with the $\tilde{I}^{(1)}_{S,N}$ denoting additional overlap contributions associated with the one-loop amplitudes. The latter depend on the momentum transfer $|\vec{q}|$ to the outgoing electron. The appearance of the difference between the $\tilde{I}^{(1)}_{S,N}$ and $I^{(1)}_{S,N}$ reflects the vanishing of the one-loop amplitudes in the $|\vec{q}| \to 0$ limit. Note that for finite $|\vec{q}|$, $\alpha^{(1)}_{S,ud}$(NLO loop) is finite in the $m_q \to 0$ limit; the explicit m_π appearing in the prefactor of Eq. (2.1.6) is compensated by a $1/m_\pi$ in $\Delta^{(1)}_S$.

The NLO two-nucleon contribution generates a significantly larger correction, given by

$$-\alpha^{(1)}_{S,ud}({\rm NLO\ NN}) = \sqrt{\frac{m_\mu}{\omega_{\rm capt}}}\left(\frac{m_\mu}{4\pi v}\right)^2 \left(\frac{3B_0 K_F \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}\right) f^{SI}_{\rm eff}\left(I^{(1)}_{S,p} + I^{(1)}_{S,n}\right) = 18.8^{+1.6}_{-9.5}\ , \tag{2.1.8}$$

where K_F is the nuclear Fermi momentum and $f_{\rm eff}^{SI} = 1.05^{+0.07}_{-0.53}$ is obtained by performing a one-body Fermi Gas average of the two-nucleon amplitude over a spin- and isospin-symmetric core. Note that both the NLO loop and NLO two-nucleon contributions enter with an opposite sign compared to the LO amplitude, thereby reducing the sensitivity to the $C_q^{S,L}$. The impact of the two-nucleon term may be particularly severe, with a reduction of up to $\sim 25\%$ (50%) of the LO amplitude (rate), although the uncertainty in that estimate is also significant. A similar decomposition applies to the relative magnitudes of the $\alpha_{S,ud}^{(w)}$. We discuss the details leading to these results in the subsequent sections of the paper.

In the case of vector-mediated CLFV, the conversion amplitudes are given by,

$$\left|\tau_V^{(w)}\right|^2 = \left|\alpha_{V,u}^{(w)}\left(\frac{C_u^{V,L} \pm C_u^{V,R}}{2}\right) + \alpha_{V,d}^{(w)}\left(\frac{C_d^{V,L} \pm C_d^{V,R}}{2}\right)\right|^2 . \tag{2.1.9}$$

Once again, the positive signs are used for $w \in \{1,3\}$, while the negative signs are used for $w \in \{2,4\}$. The model-independent α's are defined in Appendix 2.G and their numerical values are given in Table 2.2. The coherent vector amplitudes receive no NLO contributions via either loops or two-nucleon amplitudes. In the latter instance, the result is well-known from the analysis of meson-exchange contributions to the nuclear electromagnetic current. The leading non-trivial corrections to the charge operator appear at NNLO, whereas the three-current receives NLO contributions. The latter, however, is not a coherent operator, so we do not consider the analogous current for the vector-mediated conversion process.

Numerical results for the model-independent parameters $\alpha_{S,ud}^{(1)}$ *etc.* are given in Tables 2.1 and 2.2. As noted above, the NLO two-nucleon contributions may significantly degrade the sensitivity to the scalar-mediated interactions, whereas the vector-mediated sensitivities are unaffected to this order. We also note that the dominant sources of uncertainty in the scalar mediated branching ratio comes from the LO and NLO two-nucleon terms. The LO uncertainties are limited by the determination of the nucleon sigma-terms and quark masses. At NLO, the one-body Fermi Gas averaging of the two-nucleon term is the dominant source of uncertainty. This is again in contrast to the case of vector mediated conversion, for which the parametric and nuclear uncertainties are of the same order of magnitude as one expects for the NNLO contributions which are not explicitly computed in this work.

This paper is organized as follows: in order to facilitate the reader's following the primary logic of our study, we relegate significant material to a number of Appendices that accompany the various sections. In Section 2.2, we introduce the low-energy phenomenological effective CLFV Lagrangian and discuss the corresponding Wilson coefficients. Section 2.3 and the accompanying Appendices 2.A and 2.B review the

Parameter	Value	LO Contribution	NLO Loop	NLO Two-Nucleon
$\alpha^{(1)}_{S,ud}$	43^{+15}_{-12}	65±11	−2.71±0.30	$-18.8^{+9.5}_{-1.6}$
$\alpha^{(1)}_{S,s}$	3.71 ± 0.93	3.71±0.93	—	—
$\alpha^{(1)}_{S,\Theta}$	8.43 ± 0.13	8.43±0.13	—	—
$\alpha^{(2)}_{S,ud}$	32^{+11}_{-8}	47.1±8.3	−1.96±0.22	$-13.6^{+6.9}_{-1.2}$
$\alpha^{(2)}_{S,s}$	2.69 ± 0.67	2.69±0.67	—	—
$\alpha^{(2)}_{S,\Theta}$	6.11 ± 0.10	6.11±0.10	—	—
$\alpha^{(3)}_{S,ud}$	-32^{+8}_{-11}	−47.4±8.3	1.96±0.22	$13.7^{+1.2}_{-7.0}$
$\alpha^{(3)}_{S,s}$	-2.70 ± 0.68	−2.70±0.68	—	—
$\alpha^{(3)}_{S,\Theta}$	-6.15 ± 0.10	−6.15±0.10	—	—
$\alpha^{(4)}_{S,ud}$	-43^{+12}_{-15}	−65±11	2.68±0.29	$18.7^{+1.6}_{-9.5}$
$\alpha^{(4)}_{S,s}$	-3.70 ± 0.93	−3.70±0.93	—	—
$\alpha^{(4)}_{S,\Theta}$	-8.41 ± 0.13	−8.41±0.13	—	—

Table 2.1. Table of branching ratio parameters for scalar-mediated conversion

Parameter	Value
$\alpha^{(1)}_{V,u}$	12.25±0.13
$\alpha^{(1)}_{V,d}$	12.23±0.27
$\alpha^{(2)}_{V,u}$	−9.65±0.11
$\alpha^{(2)}_{V,d}$	−9.63±0.21
$\alpha^{(3)}_{V,u}$	−9.68±0.11
$\alpha^{(3)}_{V,d}$	−9.67±0.21
$\alpha^{(4)}_{V,u}$	12.19±0.13
$\alpha^{(4)}_{V,d}$	12.18±0.27

Table 2.2. Table of branching ratio parameters for vector-mediated conversion

formalism of ChPT. We apply this framework to scalar-mediated CLFV in Section 2.4, deriving the LO and NLO matching of the phenomenological CLFV operators onto the low-energy hadronic interactions at the one- and two-nucleon level. The one-body average of the two-nucleon interaction is discussed in Section 2.5 and Appendix 2.D. In Section 2.6, we consider the case of vector-mediated CLFV. The focus then turns to the sources of theoretical hadronic uncertainties in Section 2.7 and Appendix 2.C. Section 2.8 discusses the calculation of the muon and electron wavefunctions, while Section 2.9 and Appendix 2.E examine uncertainties introduced by the nuclear density distributions. The branching ratio is calculated in Section 2.10 and the accompanying

Appendices 2.F and 2.G, leading to our master formula in Eq. (2.1.2). The impact of the next-to-leading order corrections and uncertainties on the upcoming CLFV experiments is discussed in Section 2.11. We summarize our main results in Section 2.12 and provide Appendix 2.G as a summary of how these results may be utilized.

2.2 Quark-Level CLFV Lagrangian

There are a wide variety of extensions to the Standard Model that allow for CLFV. For an incomplete list of representative models, see, e.g. Refs. [73–83], and for more comprehensive surveys of the literature, see Refs. [28, 84]. Assuming that the process mediating CLFV occurs at a mass scale significantly greater than that of the momentum transfer involved in coherent $\mu - e$ conversion, $q_T^2 \approx m_\mu^2$, it suffices to concentrate on the low-energy effective Lagrangian which includes only SM fields as explicit degrees of freedom.

In principle, one may start with an effective Lagrangian that respects the $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry of the SM. Since our focus is on physics at the hadronic scale and below, we follow other authors [68–70] and work with an effective theory in which only the $\mathrm{SU}(3)_C \times \mathrm{U}(1)_{\mathrm{EM}}$ symmetry is manifest. The lowest dimension conversion operators of interest appear at mass dimension six:

$$\begin{aligned}\mathcal{L}_{\mathrm{CLFV}} = &\sum_{f=\mathrm{u,d,s,c,b,t}} \frac{1}{\Lambda^2}\left[\lambda_f^{S,L}\bar{e}P_L\mu + \lambda_f^{S,R}\bar{e}P_R\mu + \mathrm{h.c.}\right]\bar{q}_f q_f \\ &+ \sum_{f=\mathrm{u,d,s,c,b,t}} \frac{1}{\Lambda^2}\left[\lambda_f^{V,L}\bar{e}\gamma^\nu P_L\mu + \lambda_f^{V,R}\bar{e}\gamma^\nu P_R\mu + \mathrm{h.c.}\right]\bar{q}_f\gamma_\nu q_f.\end{aligned} \tag{2.2.1}$$

In principle, parity odd terms that couple to the pseudoscalar and axial-vector quark currents could be included, but this is not done as these contributions will be suppressed in coherent conversion. We also do not include the dipole operators relevant to $\mu \to e\gamma$ as their contributions to the coherent conversion process are typically suppressed relative to contributions from the scalar and vector interactions in Eq. (2.2.1).

In coherent conversion, the momentum transfer is roughly equal to the muon rest mass. As such, the dominant contributions from heavy quarks arise through loop diagrams. Integrating out the heavy quarks results in an effective gluonic coupling that can be related to the stress energy tensor through the trace anomaly [85]. This

procedure yields the Wilson coefficients

$$C_f^{S,\mathcal{X}} = \lambda_f^{S,\mathcal{X}} - \frac{2}{27}\sum_{h=\mathrm{c,b,t}} \frac{m_f}{m_h}\lambda_h^{S,\mathcal{X}}, \tag{2.2.2}$$

$$C_f^{V,\mathcal{X}} = \lambda_f^{V,\mathcal{X}}, \tag{2.2.3}$$

$$C_\Theta^{\mathcal{X}} = \frac{2}{27}\sum_{h=\mathrm{c,b,t}} \frac{m_N}{m_h}\lambda_h^{S,\mathcal{X}}, \tag{2.2.4}$$

where m_N is the nucleon mass and $\mathcal{X} = L, R$ denotes the muon handedness. The resulting CLFV effective Lagrangian is

$$\begin{aligned}\mathcal{L}_{\mathrm{CLFV}} = &\sum_{f=\mathrm{u,d,s}} \frac{1}{\Lambda^2}\left[C_f^{S,L}\bar{e}P_L\mu + C_f^{S,R}\bar{e}P_R\mu + \mathrm{h.c.}\right]\bar{q}_f q_f \\ &+ \sum_{f=\mathrm{u,d,s}} \frac{1}{\Lambda^2}\left[C_f^{V,L}\bar{e}\gamma^\nu P_L\mu + C_f^{V,R}\bar{e}\gamma^\nu P_R\mu + \mathrm{h.c.}\right]\bar{q}_f\gamma_\nu q_f \\ &+ \frac{1}{M_N\Lambda^2}\left[C_\Theta^L\bar{e}P_L\mu + C_\Theta^R\bar{e}P_R\mu + \mathrm{h.c.}\right]\Theta_\mu^\mu.\end{aligned} \tag{2.2.5}$$

For compactness of notation, we will define the effective CLFV currents

$$J_f = C_f^{S,L}\bar{e}P_L\mu + C_f^{S,R}\bar{e}P_R\mu + \mathrm{h.c.}, \tag{2.2.6}$$

$$J_f^\nu = C_f^{V,L}\bar{e}\gamma^\nu P_L\mu + C_f^{V,R}\bar{e}\gamma^\nu P_R\mu + \mathrm{h.c.}, \tag{2.2.7}$$

$$J_\Theta = C_\Theta^L\bar{e}P_L\mu + C_\Theta^R\bar{e}P_R\mu + \mathrm{h.c.}, \tag{2.2.8}$$

which couple to the quark scalar current, quark vector current, and trace of the stress energy tensor respectively.

The Lagrangian in (2.2.5) enables a model independent analysis of different theories with high-scale CLFV. However, it will be used to describe CLFV processes involving light quarks at the energy scales where QCD is non-perturbative and the relevant degrees of freedom are nucleons and mesons. The appropriate framework for doing this is ChPT.

2.3 Chiral Power Counting and Chiral Lagrangians

ChPT is the low-energy effective field theory of QCD [86]. At low energies QCD becomes confining, which makes perturbative calculations with quarks and gluons intractable. Rather than using quarks and gluons as the fundamental degrees of freedom,

ChPT replaces them with the bound states of mesons and baryons. Beyond these dynamical fields, ChPT can also include external source fields. These external sources will be used to incorporate the effective CLFV operators.

Starting from (2.2.5), one may use ChPT to relate the CLFV currents to an effective theory with multiple unknown LECs that must be matched onto experimental results. As is done in Appendix 2.A, it can be shown that these LECs are related to known nuclear matrix elements that appear in standard ChPT. The scalar and vector CLFV currents then appear in the chiral Lagrangian in an analogous manner to the quark mass and electromagnetic insertions, respectively. However, as the CLFV currents do not scale with the quark mass they are assigned chiral order $\mathcal{O}(1)$. While $\mathcal{O}(1)$ in chiral power counting, the CLFV operators are still small in the sense that they correspond to high-scale physics and thus we may restrict our attention to terms with only a single CLFV insertion.

The inclusion of baryons in the chiral Lagrangian introduces additional complications in power counting beyond leading order. One well established method for dealing with these difficulties is Heavy Baryon Chiral Perturbation Theory (HBChPT) [87]. This method requires a choice of reference velocity V_μ such that the decomposition of a nucleon's momentum, $P_\mu = m_N V_\mu + k_\mu$, yields a value of k_μ that is small compared to the chiral scale. For present purposes, the reference velocity is chosen to be $V_\mu = (1, 0, 0, 0)$ in the rest-frame of the target nucleus. As a result, the magnitude of the residual three-momentum will be of the same order as the nuclear Fermi momentum, $|\vec{k}| \approx K_F \sim \mathcal{O}(q)$.

As noted in Section 2.2, the momentum transfer scale for coherent conversion is set by the muon mass, $m_\mu \approx 106\,\text{MeV}$, which is comparable to the strange quark mass, $m_s \approx 92\,\text{MeV}$. Consequently, one should explicitly include the strange quark in the effective theory. On the other hand, the momentum transfer scale is not much greater than the strange quark mass and the nucleons have no net strangeness. Therefore, one might expect that the contributions of CLFV operators containing strange quark fields will be significantly smaller than the contributions of those coupling through the up and down quarks. If so, it may be advantageous to use $SU(2)$ ChPT with the leading order contributions of the strange quark operators treated as additional singlets under the flavor symmetry rather than resorting to $SU(3)$ ChPT. As has been demonstrated previously [71, 72], chiral $SU(2)$ allows for better control of both theoretical uncertainties and uncertainties introduced by the low-energy constants (LECs) of the chiral Lagrangian than is possible with chiral $SU(3)$. *A priori*, the choice of chiral $SU(2)$ is not necessarily justified. However, the smallness of the strange quark contribution is borne out numerically in the results of Tables 2.1 and 2.2, justifying this approach.

Including the strange quark singlets in the chiral Lagrangian introduces an addi-

tional set of LECs that must be matched onto experimental results. The full set of relevant building blocks for the chiral Lagrangian and the complete chiral Lagrangian can be found in Appendix 2.A.

Lastly, chiral power counting for complete Feynman diagrams needs to be examined: in particular, how chiral power counting applies to multi-nucleon diagrams. One convenient power counting scheme only depends on the vertices and topological properties of the diagram [88–90]. An operator from the purely pionic sector $\mathcal{L}_{\pi\pi}^{(n)}$ is assigned the effective chiral power $\epsilon = n - 2$ while operators from the pion-nucleon Lagrangian $\mathcal{L}_{\pi N}^{(n)}$ are given $\epsilon = n - 1$. This effective chiral power is lower than the chiral order of the Lagrangian because the scaling of the propagators associated with a vertex must now be included with the vertex. This allows any diagram to be assigned an effective chiral order based on the following rule,

$$\nu = 4 - A - 2C + 2L + \sum_i V_i \epsilon_i + \epsilon_{CLFV}, \tag{2.3.1}$$

where A is the number of external nucleons, C is the number of connected parts of the diagram, L is the number of loops, V_i is the number of vertices with effective chiral power ϵ_i, and ϵ_{CLFV} is the effective chiral power of the CLFV operator used.

2.4 Scalar-Mediated Conversion

For the case of scalar-mediated conversion, the CLFV vector currents can be eliminated leaving the Lagrangian

$$\mathcal{L}_{\pi\pi}^{(0)} = \frac{\mathring{f}_\pi^2}{4} Tr\left[\chi\left(U^\dagger + U\right)\right], \tag{2.4.1}$$

$$\mathcal{L}_{\pi N}^{(0)} = \bar{N}\left[\bar{c}_5\left(\chi\left(U + U^\dagger\right) - \frac{1}{2}Tr\left[\chi\left(U + U^\dagger\right)\right]\right) + \bar{c}_1 Tr\left[\chi\left(U + U^\dagger\right)\right] + d_1^S \chi_s\right] N. \tag{2.4.2}$$

The operators χ and χ_s encapsulate insertions of the CLFV currents while the coefficients $\bar{c}_1$, $\bar{c}_5$, and d_1^S are LECs that must be matched onto experimental data, see Appendix 2.A. The constant $\bar{c}_1$ is related to the nucleon mass in the isospin-symmetric limit, the constant $\bar{c}_5$ corresponds to the tree-level, isospin-breaking difference in the proton and neutron masses, and d_1^S is related to the strange quark contribution to the nucleon mass.

There are only two types of scalar insertion vertices that contribute at LO or NLO: insertion on a pion line from $\mathcal{L}_{\pi\pi}^{(0)}$ with effective chiral power $\epsilon_{CLFV} = -2$ and insertion on a nucleon line from $\mathcal{L}_{\pi N}^{(0)}$ with $\epsilon_{CLFV} = -1$. It should be stressed that these are

the effective chiral powers used with the power counting scheme in (2.3.1) and do not correspond to how these terms in the Lagrangian scale with the power of m_π or small momentum q. There are additional types of vertices at the same chiral order, but these will involve an even number of extra pions connected to the vertex; as such, these vertices can only contribute to diagrams at NNLO and beyond.

There are four possible diagrams that may contribute at LO and NLO. These can be divided into three categories:

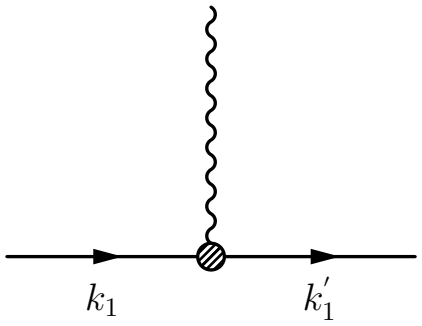
1. Single Nucleon, Tree-Level: The diagram of interest can be seen in Figure 2.1a. This consists of just the scalar insertion from $\mathcal{L}^{(0)}_{\pi N}$ on a single nucleon line that enters at effective chiral order $\nu = 3 - 3A$.

2. Single Nucleon, One-Loop: There are two possible diagrams that involve a pionic loop and a single nucleon. One diagram, Figure 2.1b, consists of a single pion-nucleon vertex where the pion lines make a loop with the scalar insertion. The other, shown in Figure 2.1c, is a sunset diagram with two pion-nucleon vertices where the scalar insertion happens on the internal pion line. Both of these diagrams involve the insertion of a CLFV operator from $\mathcal{L}^{(0)}_{\pi\pi}$ and enter at effective chiral order $\nu = 4 - 3A$. The diagram with a purely pionic loop, Figure 2.1b, must vanish because the scalar insertion is symmetric in flavor indicies while the pion-nucleon vertex is anti-symmetric.

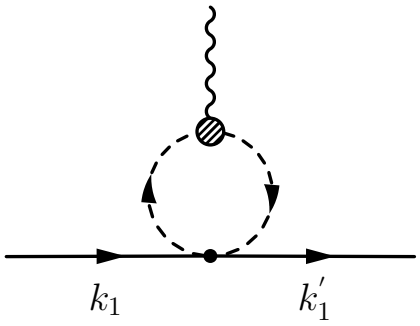
3. Two-Nucleon, Tree-Level: The diagram of interest can be seen in Figure 2.1d. Two nucleons exchange a pion and the scalar insertion occurs on the internal pion line. The CLFV vertex is from $\mathcal{L}^{(0)}_{\pi\pi}$ and thus this diagram enters at effective chiral order $\nu = 4 - 3A$.

It should be noted that these four diagrams have been analyzed previously in the context of dark matter direct-detection with an $SU(3)$ chiral Lagrangian [90]. The present formulation differs primarily in the use of an $SU(2)$ chiral Lagrangian to allow better control of uncertainties and a different treatment of the two-nucleon contribution. The present results were derived independently and agree with those of Ref. [90] in the limit of chiral $SU(2)$. We also note that recent studies of two-nucleon scalar currents relevant to dark matter-nucleus scattering have been performed in Refs. [91–93] using chiral effective field theory.

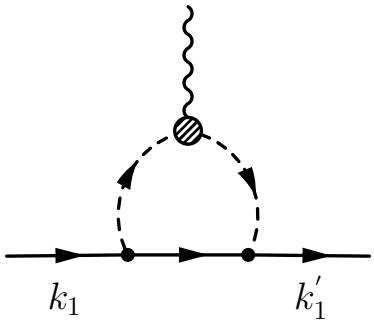
The diagrams involving only a single nucleon can be readily evaluated. Taken



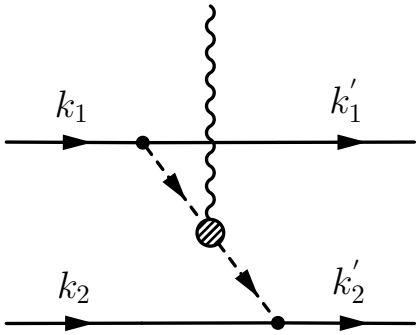
(a) Leading order diagram consisting of a tree level insertion of a CLFV vertex.



(b) Next-to-leading order diagram with a purely pionic loop and single nucleon.



(c) Next-to-leading order sunset diagram with an internal pion and single nucleon.



(d) Next-to-leading order diagram that involves the exchange of a pion between two nucleons.

Figure 2.1. The set of Feynman diagrams that contribute to coherent $\mu - e$ conversion through NLO in a scalar-mediated model of CLFV. The fermionic and scalar lines correspond to nucleons and pions respectively. The shaded vertex represents an insertion of a CLFV operator. For diagrammatic simplicity, the leptonic line is not featured but would connect to the CLFV vertex.

together, these three diagrams result in an effective nucleon-level CLFV Lagrangian,

$$\begin{aligned}\mathcal{L}_{1-N} = \bar{N}\Bigg[&(2\bar{c}_1 - \bar{c}_5)\,Tr\left[\chi\right] + 2\bar{c}_5\chi + d_1^S\chi_S + \frac{1}{\Lambda^2}J_\theta \\ &-\frac{3B_0 m_\pi \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2\Lambda^2}\left(J_u + J_d\right)\left(\frac{2+X_\pi}{\sqrt{X_\pi}}\text{arccot}\left(\frac{2}{\sqrt{X_\pi}}\right) - 1\right)\Bigg] N,\end{aligned} \tag{2.4.3}$$

where the quantity $X_\pi = (\vec{q}_T)^2/m_\pi^2$ depends on the three momentum transferred to the nucleus, $\vec{q}_T$. The contribution from the stress-energy tensor has also been included in this effective Lagrangian.

The Lagrangian (2.4.3) can be further simplified by relating the LECs from $SU(2)$ ChPT to the contributions of the quark condensates to the proton and neutron mass. The difference between the proton and neutron mass is a NNLO effect that arises from isospin-symmetry breaking [94, 95]. Thus, we may take $\bar{c}_5 = 0$ as this is the LEC responsible for the mass splitting. The remaining LEC, $\bar{c}_1$, may be expressed at NLO accuracy in terms of f_u^N (f_d^N), the fraction of the nucleon mass due to the u- (d-) quark condensate, as

$$-4B_0\bar{c}_1 = \frac{m_N f_u^N}{m_u} = \frac{m_N f_d^N}{m_d}. \tag{2.4.4}$$

Similarly, as shown in Appendix 2.B, the unknown LEC for the strange operator, d_1^S, can be matched onto the nucleon mass contribution from the strange quark condensate, f_s^N. Rewritting χ and χ_S in terms of the effective CLFV currents (see Appendix 2.A), one finds the effective Lagrangian

$$\begin{aligned}\mathcal{L}_{1-N} = \frac{1}{\Lambda^2}\bar{N}\Bigg[&\frac{m_N f_u^N}{m_u}J_u + \frac{m_N f_d^N}{m_d}J_d + \frac{m_N f_s^N}{m_s}J_s + J_\theta \\ &-\frac{3B_0 m_\pi \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}(J_u + J_d)\left(\frac{2+X_\pi}{\sqrt{X_\pi}}\text{arccot}\left(\frac{2}{\sqrt{X_\pi}}\right) - 1\right)\Bigg] N.\end{aligned} \tag{2.4.5}$$

However, as we are working to NLO in $SU(2)$ ChPT, it is more appropriate to parameterize the effective Lagrangian in terms of the isospin-symmetry breaking parameter $\xi = \frac{m_d - m_u}{m_d + m_u}$. We will also introduce the isospin-symmetric quantities $\hat{m} = \frac{m_u + m_d}{2}$, the isospin averaged quark mass, and $\sigma_{\pi N}$, the pion-nucleon sigma-term.

As has been shown previously in the literature [71, 72], significant care must be taken to disentangle three flavor uncertainties when providing the chiral expansion for f_u^N and f_d^N. These chiral expansions are known through NNLO [72]. As the present analysis of coherent $\mu - e$ conversion only extends to NLO, one finds

$$m_N f_q^N = \frac{1}{2}\sigma_{\pi N}(1 \mp \xi). \tag{2.4.6}$$

In this expression, q is a placeholder index for the u- (d-) quark condensate which is given by the negative (positive) sign. In terms of the isospin average quark mass, the u- (d-) quark mass is given by the negative (positive) sign in $m_q = \hat{m}(1 \mp \xi)$. It is then straightforward to show using (2.4.6) that

$$\frac{m_N f_q^N}{m_q} = \frac{\sigma_{\pi N}}{2\hat{m}}. \tag{2.4.7}$$

Making use of (2.4.7), one may rewrite (2.4.5) to arrive at the final effective Lagrangian for the one-nucleon sector

$$\mathcal{L}_{1-N} = \frac{1}{\Lambda^2}\bar{N}\left[\frac{\sigma_{\pi N}}{2\hat{m}}\left(J_u + J_d\right) + \frac{\sigma_{sN}}{m_s}J_s + J_\theta \right. \\ \left. -\frac{3B_0 m_\pi \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}\left(J_u + J_d\right)\left(\frac{2+X_\pi}{\sqrt{X_\pi}}\text{arccot}\left(\frac{2}{\sqrt{X_\pi}}\right) - 1\right)\right] N. \tag{2.4.8}$$

In this expression, we have defined the strange-nucleon sigma-term $\sigma_{sN} = m_N f_s^N$.

The two-nucleon sector only includes a single tree-level diagram. This yields the effective two-nucleon Lagrangian

$$\mathcal{L}_{2-N} = -\frac{B_0 \mathring{g}_A^2}{\mathring{f}_\pi^2 \Lambda^2}\left(J_u + J_d\right)\frac{1}{\left(q_1^2 - m_\pi^2\right)\left(q_2^2 - m_\pi^2\right)}\sum_a \left(\bar{N}_1' S\cdot q_1 \tau_a N_1\right)\left(\bar{N}_2' S\cdot q_2 \tau_a N_2\right). \tag{2.4.9}$$

The quantities $q_1 = k_1 - k_1'$ and $q_2 = k_2 - k_2'$ are defined as the difference between the initial and final momenta of the two nucleons. The Lagrangians (2.4.8) and (2.4.9) closely mirror the results from the $SU(3)$ chiral Lagrangian [90].

2.5 Approximate One-body Interaction

The effective Lagrangian (2.4.9) explicitly involves two external nucleons. Consequently, one requires the many-body wavefunctions for the initial and finial nuclei to calculate decay rates with this term. Carrying out such a complete, many-body computation goes beyond the scope of the present study. Nevertheless, in order to estimate the possible magnitude and relative sign of the two-nucleon contribution, we perform an average of the interaction over all core nucleons. In this approximation, it is assumed that every nucleon except for one valence nucleon is part of a spin-symmetric nuclear core. For the spatial wavefunction, the core nucleons can be approximated as being a degenerate Fermi gas. Such a distribution is fully characterized by its Fermi energy, E_F, or alternatively the Fermi momentum, K_F. For our purposes, it suffices to assume a common Fermi momentum for neutrons and protons. Isospin-breaking corrections should be of order $(N-Z)/A$. For earlier applications of this procedure to electroweak properties of nuclei, see, *e.g.*, Refs. [96–98].

After making these approximations and summing over all contributions from the core nucleons, the spin-dependent and spin-independent parts of the resulting effective

interaction can be expressed in momentum space as

$$\mathcal{L}_{\text{eff}} = -\frac{3B_0K_F\mathring{g}_A^2}{64\pi(2\pi)^3\mathring{f}_\pi^2\Lambda^2}\,(J_u+J_d) \cdot \bar{N}(k_f)\left[f^{SI}(\vec{q}_T,\vec{k})\mathbb{1} - f^{SD}(\vec{q}_T,\vec{k})i\vec{\sigma}\cdot\left[\left(\frac{\vec{q}_T}{K_F}\right)\times\left(\frac{\vec{k}}{K_F}\right)\right]\right]N(k_i), \quad (2.5.1)$$

where the Pauli matrices are given by $\vec{\sigma}$, $\vec{k} = \frac{1}{2}\left(\vec{k}_i + \vec{k}_f\right)$ is the average of the initial and final nucleon three-momentum, and $\vec{q}_T = \vec{k}_f - \vec{k}_i$ is the three-momentum transferred to the nucleon. As this is an approximate one-body operator, $\vec{q}_T$ is the same as the three-momentum transferred to the nucleus. The complicated dependence of the effective Lagrangian on $\vec{q}_T$ and $\vec{k}$ is encapsulated in the dimensionless functions $f^{SI}(\vec{q}_T,\vec{k})$ and $f^{SD}(\vec{q}_T,\vec{k})$. The full analytic forms of these functions are given in Appendix 2.D.

For purposes of performing our numerical estimate, it is desirable to approximate these functions by constants. Doing so ensures that the effective interaction (2.5.1) remains local in position space, allowing seamless inclusion with (2.4.8) as a single effective interaction. As is demonstrated in Appendix 2.D, the dimensionless functions $f^{SI}(\vec{q}_T,\vec{k})$ and $f^{SD}(\vec{q}_T,\vec{k})$ can be well approximated by the constants $f_{\text{eff}}^{SI} = 1.05 \pm 0.07$ and $f_{\text{eff}}^{SD} = 0.81 \pm 0.12$ respectively. The uncertainties in these constants include both the experimental uncertainties in the Fermi momentum of ^{27}Al and the anticipated errors induced by approximating the functions $f^{SI}(\vec{q}_T,\vec{k})$ and $f^{SD}(\vec{q}_T,\vec{k})$ by constants.

It is still necessary to include the errors induced by the core-averaging procedure itself. As a first pass, one may estimate these errors by examining previously studied cases in the literature where both core-averaged quantities and numerical many-body results were calculated. Analyzing previous results for the nuclear anapole moment [97], we infer that the core-averaging procedure may introduce an uncertainty of 30% to 50% when the core is treated as a Fermi gas without short range correlations. It should also be noted that the core-averaged quantities generically over estimate the many-body contribution. Thus, we may conservatively take $f_{\text{eff}}^{SI} = 1.05^{+0.07}_{-0.53}$ and $f_{\text{eff}}^{SD} = 0.81^{+0.15}_{-0.42}$.

It is entirely possible, of course, that the results of a complete many-body computation would yield a result that falls outside of the aforementioned estimate. While the simplest single particle shell model description of ^{27}Al is a $1d_{5/2}$ proton hole in ^{28}Si, there is significant configuration mixing with two-particle excitations into the higher-lying $s_{1/2}$ and $d_{3/2}$ orbitals[1]. On the other hand, the results of elastic, magnetic electron scattering appear to agree well with the $1d_{5/2}$ proton hole configuration

[1]We thank C. Johnson for a discussion of this point as well as for a numerical assessment using the Brown-Richter USDB interaction [99].

description [100][2]. Clearly, a detailed many-body computation using the two-body operator derived here will be needed for a definitive, quantitative assessment of the NLO two-body contribution.

As the conversion process is coherent, the spin-independent part of (2.5.1) couples equally to all nucleons while the spin-dependent part is only relevant for unpaired nucleons. In the nuclear shell model, ^{27}Al has only one unpaired proton. This results in a relative $1/A$ suppression of the spin-dependent term. This term may then be neglected as its contributions are comparable in magnitude to NNLO terms not considered in this analysis.

Returning to position space and combining this approximate one-body interaction with the effective Lagrangian for the single nucleon sector yields the approximate effective interaction for scalar-mediated conversion,

$$\mathcal{L}_{\text{scalar}} = \frac{1}{\Lambda^2}\bar{N}\left[\frac{\sigma_{\pi N}}{2\hat{m}}\left(J_u + J_d\right) + \frac{\sigma_{sN}}{m_s}J_s + J_\theta - \frac{3B_0 K_F \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}\left(J_u + J_d\right) f_{\text{eff}}^{SI} \right.$$
$$\left. - \frac{3B_0 m_\pi \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}\left(J_u + J_d\right)\left(\frac{2+X_\pi}{\sqrt{X_\pi}}\text{arccot}\left(\frac{2}{\sqrt{X_\pi}}\right) - 1\right)\right] N. \quad (2.5.2)$$

We emphasize that the NLO loop and two-nucleon contributions enter with the opposite sign relative to the LO single nucleon terms, a feature reflected by the numerical results given in Table 2.1.

2.6 Vector-Mediated Conversion

For the case of vector-mediated conversion, the scalar CLFV operators in the effective Lagrangian from Appendix 2.A can be removed. The vector CLFV operators enter the pion-nucleon Lagrangian at order $\mathcal{L}_{\pi N}^{(0)}$ but do not enter the purely pionic Lagrangian until $\mathcal{L}_{\pi\pi}^{(1)}$. This is because the vector CLFV current cannot couple to the scalar field except through a derivative. Thus pion loop and two-nucleon diagrams for vector-mediated CLFV will enter at NNLO instead of NLO as happened for scalar-mediated CLFV. Therefore, the only relevant diagrams will be tree-level insertions of the vector current. Replacing derivates with explicit factors of nucleon momentum, the CLFV

[2]We thank T. W. Donnelly for alerting us to these results.

Lagrangian may be rewritten as

$$\mathcal{L}_{\text{vector}} = \bar{N}_f \left[\left(V^\mu + \frac{(k_f + k_i)^\mu}{2M_N} - \frac{V \cdot (k_f + k_i)}{2M_N} V^\mu \right) \left(v + v^{(s)} \right)_\mu \right. \\ \left. - \frac{i}{M_N} \epsilon^{\mu\nu\rho\sigma} V_\rho S_\sigma \left(k_f - k_i \right)_\mu \left(\left(1 + \mathring{k}_V \right) v + \left(1 + \mathring{k}_s \right) v^{(s)} + \mu_s v_s^{(s)} \right)_\nu \right] N_i, \tag{2.6.1}$$

where we have used the relation $[S^\mu, S^\nu] = i\epsilon^{\mu\nu\rho\sigma} V_\rho S_\sigma$, see Ref. [87], and identified the unknown LEC for the strange sector with the nucleon's strangeness magnetic moment, as demonstrated in Appendix 2.B.

The second set of terms that appear in the Lagrangian are spin-dependent while the first set are spin-independent. As discussed in Section 2.5, the spin-dependent terms are suppressed by a factor of $1/A$ and it suffices to retain only the coherent, spin-independent terms. The Lagrangian also has terms of the form $V \cdot \left(\frac{k_{(i,f)}}{M_N} \right)$. As the external nucleons will be on shell, these terms are suppressed and actually enter at NNLO instead of NLO. Thus, these terms can be dropped from the effective nucleon CLFV Lagrangian leaving

$$\mathcal{L}_{\text{vector}} = \bar{N}_f \left[\left(V + \frac{k_f + k_i}{2M_N} \right) \cdot \left(v + v^{(s)} \right) \right] N_i. \tag{2.6.2}$$

This Lagrangian depends not only on the magnitude of $(k_f + k_i)_\mu$ but also its direction. By parity-symmetry, the spatial components of $(k_f + k_i)_\mu$ must vanish, but this still leaves the component $(k_f + k_i)_0$, the sum of nucleon kinetic energy. However, in the rest frame of the nucleus, this is equal to $V \cdot (k_f + k_i)$ which enters at NNLO as mentioned before. As a result, the final Lagrangian for vector-mediated conversion through NLO is just given by

$$\mathcal{L}_{\text{vector}} = \bar{N} \left[V \cdot \left(v + v^{(s)} \right) \right] N. \tag{2.6.3}$$

2.7 Hadronic Uncertainties

The effective one-nucleon Lagrangians for scalar-mediated conversion, (2.5.2), and vector-mediated conversion, (2.6.3), introduce a variety of physical parameters that must be matched onto experimental results. These include the light quark masses, pion decay constant, and nucleon axial-vector coupling, among others.

The values of $\sigma_{\pi N}$, $\hat{m}$, and ξ in addition to the other low energy parameters that appear in (2.5.2) and (2.6.3) can be determined by making use of lattice QCD results. Modern $N_f = 2{+}1$ lattice QCD simulations provide realistic insight into several of these

parameters with uncertainties that are smaller than their experimental counterparts. The low energy constants $\mathring{f}_\pi$ and B_0 along with the three light quark masses can be taken from the world average of lattice QCD results published by FLAG [101]. Similar world averages have been performed for both σ_{sN} and $\sigma_{\pi N}$ [102, 103]. We observe that more recent lattice determinations of $\sigma_{\pi N}$ [104–107] lead to somewhat smaller values than obtained in Ref. [103]. On the other hand, phenomenological analyses of πN scattering yield a larger value that is in tension with the recent lattice results [108, 109]. The older result given in Ref. [103], based on a fit to earlier lattice results, lies between the recent lattice and phenomenological determinations. For present purposes, then, we will utilize this value and the quoted uncertainty (see Table 2.3), recognizing that a more robust determination will require further effort by the lattice QCD and phenomenological communities.

While lattice QCD simulations do provide better uncertainties for some quantities, others are best taken from experimental results. The pion and nucleon masses presented by the Particle Data Group are known to an exceptional degree of precision [110]. Similarly, the nuclear axial-vector coupling, g_A, has been determined with high precision in ultra-cold neutron studies [111]. It should be stressed, however, that the experimentally observed values of the nucleon pole masses and nucleon axial-vector coupling are not quite the same as the objects that appear in the HBChPT Lagrangian. This is because the parameters in the HBChPT Lagrangian are the tree-level values taken in the chiral limit. Despite this difference, the experimental and chiral values only differ at NNLO and can thus be treated as equivalent for present purposes.

The full collection of low energy constants and their sources is summarized in Table 2.3 of Appendix 2.C along with the set of parameters that are derived from these constants.

2.8 Wavefunctions of the Muon and Electron

Calculation of the coherent $\mu - e$ conversion rate requires knowledge of the wavefunctions for the bound muon and outgoing electron. Once captured by a nucleus, the muon relaxes to its ground state on a time scale much shorter than its mean lifetime. As such, one only needs to consider the captured muon in its ground state. The outgoing electron, however, is in a scattering state of fixed energy. These scattering states are highly relativistic as the electron receives nearly all of the decaying muon's energy, up to higher order corrections from nuclear recoil. To properly describe the wavefunction of the electron the Dirac equation must be used.

While the nucleus and electron or muon technically form a two-body system; reduced mass effects enter at NNLO and therefore the nucleus can be treated as a static source of a central potential. Following standard conventions [68, 69, 112], the time-independent Dirac equation in a spherically symmetric potential may be expressed as

$$W\psi = \left[-i\gamma_5\sigma_r\left(\partial_r + \frac{1}{r} - \frac{\beta}{r}K\right) + V(r) + m\beta\right]\psi, \tag{2.8.1}$$

where

$$\begin{aligned}\beta &= \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & -\mathbb{1}_2 \end{pmatrix}, \gamma_5 = \begin{pmatrix} 0 & \mathbb{1}_2 \\ \mathbb{1}_2 & 0 \end{pmatrix}, \sigma_r = \begin{pmatrix} \hat{r}\cdot\vec{\sigma} & 0 \\ 0 & \hat{r}\cdot\vec{\sigma} \end{pmatrix}, \\ K &= \begin{pmatrix} \vec{\sigma}\cdot\vec{l} + \mathbb{1}_2 & 0 \\ 0 & -\left(\vec{\sigma}\cdot\vec{l} + \mathbb{1}_2\right) \end{pmatrix}.\end{aligned} \tag{2.8.2}$$

In these expressions, the energy and mass of the particle are given by W and m respectively. The operator K has been introduced for convenience as it commutes with the Hamiltonian while $\vec{\sigma}\cdot\vec{l}$ does not. This operator also has the useful property that $K^2 = \hat{J}^2 + \frac{1}{4}$. Letting κ denote the eigenvalue of K and $j(j+1)$ denote that of $\hat{J}^2$, it follows that $\kappa = \pm\left(j + \frac{1}{2}\right)$.

As the operators J^2, J_z, and K commute with the Hamiltonian and each other, it is possible to work in a basis of states that have definite energy and eigenvalues for these operators. The two-component spinors in this basis will be denoted by $\chi^{\mu}_{\kappa}(\theta,\phi)$ where μ is the eigenvalue of J_z. This then allows the wavefunction to be decomposed as

$$\psi = \begin{pmatrix} g_\kappa(r)\,\chi^{\mu}_{\kappa}(\theta,\phi) \\ i f_\kappa(r)\,\chi^{\mu}_{-\kappa}(\theta,\phi) \end{pmatrix}, \tag{2.8.3}$$

where $g(r)$ and $f(r)$ are real valued functions. Expressed in terms of $g(r)$ and $f(r)$, the Dirac equation can be rewritten as the system of coupled differential equations

$$\frac{d}{dr}\begin{pmatrix} g \\ f \end{pmatrix} = \begin{pmatrix} -\frac{\kappa+1}{r} & W - V(r) + m \\ -(W - V(r) - m) & \frac{\kappa-1}{r} \end{pmatrix}\begin{pmatrix} g \\ f \end{pmatrix}. \tag{2.8.4}$$

These coupled equations can then be solved numerically using the shoot-and-match procedure [113].

As the muon is in its ground state, its wavefunction is normalized using the usual scheme

$$\int d^3x \psi^{(\mu)\dagger}_{\kappa',\mu'}(x)\,\psi^{(\mu)}_{\kappa,\mu}(x) = \delta_{\mu',\mu}\,\delta_{\kappa',\kappa}. \tag{2.8.5}$$

The electron, however, is described by a scattering state which require a different normalization scheme. Because the wavefunction takes continuous energy eigenvalues these states are normalized as

$$\int d^3x \psi^{(e)\dagger}_{\kappa',\mu',E'}(x)\,\psi^{(e)}_{\kappa,\mu,E}(x) = 2\pi\delta(E' - E)\,\delta_{\mu',\mu}\,\delta_{\kappa',\kappa}. \tag{2.8.6}$$

2.9 Nuclear Density Distributions

Beyond the wavefunctions of the muon and electron, it is also necessary to determine the distribution of protons and neutrons in the nucleus of ^{27}Al. These distributions directly enter the calculation of the decay rate and the proton density distribution indirectly impacts the muon and electron wavefunctions by virtue of determining the electric potential in the vicinity of the nucleus.

As the proton is electrically charged, its nuclear density distributions have been thoroughly explored through electron scattering experiments [114]. These experiments have determined the nuclear charge density distribution of many nuclei to high precision in a model-independent manner [115]. One such model-independent decomposition of the nuclear charge density distribution is the Fourier-Bessel expansion. Using this expansion, the distribution is given by the piecewise function

$$\rho_p(r) = \begin{cases} \sum_n a_n j_0\left(\frac{n\pi r}{R}\right) & r \leq R \\ 0 & r > R \end{cases}. \tag{2.9.1}$$

There are a variety of ways to normalize this distribution, the scheme $\int 4\pi r^2 \rho(r)\, dr = Z$ will be used here. In (2.9.1), the parameter R acts as a cutoff radius for the charge and the set of parameters a_n correspond to independent components of the charge density distribution. While the distribution is cut off at $r = R$, the distribution is defined such that it goes to zero in a continuous manner. The experimentally determined values of these parameters for ^{27}Al are given in Table 2.4 of Appendix 2.E.

While the Fourier-Bessel parameters of Table 2.4 are given without individual uncertainties, the uncertainty in the root-mean-square charge radius is known. Experimentally, $\langle r^2 \rangle_p^{1/2} = 3.035 \pm 0.002\,\text{fm}$., which corresponds to a relative uncertainty of less than .1% [115]. As this uncertainty is far smaller than the already neglected NNLO contributions, the parameters in Table 2.4 can be treated as exact for current purposes.

The neutron has no electrical charge and it is correspondingly much more challenging to precisely measure the neutron density distribution. One experimental technique uses measurements from pionic-atoms which allows for indirect determination of the neutron density from the isospin dependence of the pion-nucleon interaction [114]. Due to the limitations of this data, the neutron density distribution is usually parameterized in terms of the two-parameter Fermi distribution rather than the model-independent Fourier-Bessel expansion [116]. The two-parameter Fermi distribution is given by

$$\rho_n(r) = \frac{\rho_0}{1 + e^{\frac{r-c}{z}}}. \tag{2.9.2}$$

The thickness parameter, z, and radial parameter, c, describe the shape of the neutron density distribution while ρ_0 is a normalization factor. This factor will be chosen such that $\int 4\pi r^2 \rho\left(r\right) dr = A - Z$.

The neutron thickness parameter, z, is usually taken to be equal to the proton thickness parameter for the same nucleus, assuming a two-parameter Fermi distribution for the protons. Coming from the proton distribution, z has a negligible experimental uncertainty but a difficult to quantify systematic uncertainty. Treating this thickness parameter as fixed, it is possible to determine the experimental value and uncertainties of the radial parameter [116]. Furthermore, the systematic errors associated with fixing the thickness parameter from the proton distribution can be estimated [116]. These systematic errors can be incorporated in the uncertainty in the radial parameter as is done in Table 2.5 of Appendix 2.E.

2.10 Calculation of the Branching Ratio

The primary quantity of experimental interest is the branching ratio for coherent $\mu - e$ conversion. Expressed in terms of the coherent conversion rate, $\Gamma_{\mu-e}$, and the muon capture rate for the target nucleus, ω_{capt}, the branching ratio is given by

$$\text{BR}(\mu - e) = \frac{\Gamma_{\mu-e}}{\omega_{\text{capt}}}. \tag{2.10.1}$$

To calculate the coherent conversion rate, it will be convenient to treat the CLFV Lagrangians (2.5.2) and (2.6.3) as a series of operators acting along the lepton and nucleon lines of the generic form

$$\mathcal{L}_{CLFV} = \frac{1}{\Lambda^2} \sum_j \bar{e} \mathcal{O}_{L,j} \mu \; \bar{N} \mathcal{O}_{N,j} N. \tag{2.10.2}$$

It will be necessary to introduce effective wavefunctions for the nucleons. The isospin index α will be used to distinguish the proton and neutron wavefunctions as $\psi_\alpha\left(x\right)$. The wavefunctions will be defined such that $\left|\psi_\alpha\left(x\right)\right|^2 = \rho_\alpha\left(x\right)$, where $\rho_\alpha\left(x\right)$ is the nuclear density distribution as defined in Section 2.9. Given these definitions, the wavefunctions are normalized to the nucleon number and not unity. Furthermore, it will be more convenient to work in momentum space and thus one defines the Fourier transformed wavefunctions as

$$\tilde{\psi}_\alpha\left(\vec{k}_N\right) = \int d^3x \; e^{-i\vec{x}\cdot\vec{k}_N} \psi_\alpha\left(x\right). \tag{2.10.3}$$

For the conversion process, the system is initially in a bound state composed of the nucleus and the muon. As the muon is in the ground state, its allowed eigenvalues are

$\kappa_i = -1$ and $\mu_i = \pm\frac{1}{2}$. The eigenvalue of $\kappa_i = -1$ is required because the muon's ground state has angular momentum $l = 0$. The final state consists of the nucleus and an outgoing electron that may take the eigenvalues $\kappa_f = \pm 1$ and $\mu_f = \pm\frac{1}{2}$. Furthermore, the wavefunction of the electron is also parameterized by the energy of the electron far away from the nuclear potential, E_e. Neglecting corrections from nuclear recoil which enter at NNLO, conservation of energy requires $E_e = m_\mu - B_E$ where B_E is the binding energy of the muon bound state.

The conversion rate can then be expressed as a sum of transition probabilities over all possible spin configurations,

$$\Gamma_{\mu-e} = \frac{1}{2}\sum_{\mu_i}\sum_{\mu_f,\kappa_f}\frac{m_\mu^5}{\Lambda^4}\left|\tau(E_e,\mu_i,\mu_f,\kappa_f)\right|^2, \tag{2.10.4}$$

where conservation of energy requires $E_e = m_\mu - B_E$. The conversion amplitude may be written in a dimensionless form as,

$$\tau(E_e,\mu_i,\mu_f,\kappa_f) = \frac{1}{m_\mu^{5/2}}\sum_j\int\frac{d^3k'_N}{(2\pi)^3}\int\frac{d^3k_N}{(2\pi)^3}\left[\widetilde{\psi}^*_{\alpha'}\left(\vec{k}'_N\right)\mathcal{O}^{\alpha,\alpha'}_{N,j}\left(\left|\vec{k}'_N-\vec{k}_N\right|\right)\widetilde{\psi}_\alpha\left(\vec{k}_N\right)\right] \cdot\left[\int d^3x e^{i\left(\vec{k}_N-\vec{k}'_N\right)\cdot\vec{x}}\psi^{(e)\dagger}_{\kappa_f,\mu_f,E_e}(x)\,\mathcal{O}_{L,j}\psi^{(\mu)}_{-1,\mu_i}(x)\right]. \tag{2.10.5}$$

The isospin indices α and α' have been introduced for the hadronic operator as it may have isospin dependence, as occurs in the case of vector-mediated conversion. The summation over the index j corresponds to summing over the contributions of each operator in the CLFV Lagrangian.

The structure of the phase space integrals in (2.10.5) does not depend on the model of CLFV and thus it is straightfoward to numerically evaluate these overlap integrals for each possible operator in the Lagrangians (2.5.2) and (2.6.3). This procedure is detailed in Appendix 2.F, and the numerical values and accompanying uncertainties for the phase space integrals of ^{27}Al are given in Table 2.7 of the same appendix.

As stated previously, there are eight possible spin configurations. However, there is a two-fold symmetry in the choice of overall sign for the spins. This reduces the number of independent configurations to only four. For compactness of notation, an index $w \in \{1,2,3,4\}$ will be used to denote each unique configuration. The relationship between all possible spin configurations and w is given in Table 2.6 of Appendix 2.F. The branching ratio can then be written in terms of four separate amplitudes, one for each configuration, leading to Eq. (2.1.2). The corresponding expression for the $\tau_S^{(w)}$ and $\tau_V^{(w)}$ are given in Eqs. (2.1.3) and (2.1.9), respectively. The expressions Eqs. (2.1.2,2.1.3,2.1.9) and the model independent parameters of Tables 2.1 and 2.2

allow one to start with an arbitrary model of CLFV and calculate in a straightforward manner the coherent conversion branching ratio including NLO contributions and uncertainties. These expressions and their model independent parameters constitute the primary results of this paper and their use is summarized in Appendix 2.G.

2.11 Discussion and Analysis

Having expressed $\mathrm{BR}(\mu - e)$ in terms of products of CLFV model-dependent Wilson coefficients and model-independent SM factors, we now discuss the implications in terms of sensitivity to various CLFV scenarios. We first consider the case of scalar-mediated conversion. The model-independent parameter $\alpha^{(w)}_{S,ud}$ is given by

$$\alpha^{(w)}_{S,ud} = \sqrt{\frac{m_\mu}{\omega_{\mathrm{capt}}}}\left(\frac{m_\mu}{4\pi v}\right)^2\left[\frac{\sigma_{\pi N}}{2\hat{m}}\left(I^{(w)}_{S,p}+I^{(w)}_{S,n}\right) - \frac{3B_0K_F\mathring{g}_A^2}{64\pi\mathring{f}_\pi^2}f^{SI}_{\mathrm{eff}}\left(I^{(w)}_{S,p}+I^{(w)}_{S,n}\right) - \frac{3B_0m_\pi\mathring{g}_A^2}{64\pi\mathring{f}_\pi^2}\Delta^{(w)}_S\right], \tag{2.11.1}$$

where the kinematic factor $\Delta^{(w)}_S$ is defined as

$$\Delta^{(w)}_S = \left(\tilde{I}^{(w)}_{S,p}+\tilde{I}^{(w)}_{S,n}\right) - \left(I^{(w)}_{S,p}+I^{(w)}_{S,n}\right). \tag{2.11.2}$$

As is done in Table 2.1, one can consider the LO, NLO loop, and NLO two-nucleon contributions independently. Consider the ratio of the NLO loop contribution to the LO contribution,

$$\frac{-\alpha^{(w)}_{S,ud}(\text{NLO loop})}{\alpha^{(w)}_{S,ud}(\text{LO})} = \frac{\left(\frac{3B_0m_\pi\mathring{g}_A^2}{64\pi\mathring{f}_\pi^2}\right)\Delta^{(w)}_S}{\frac{\sigma_{\pi N}}{2\hat{m}}\left(I^{(w)}_{S,p}+I^{(w)}_{S,n}\right)} = \left[\frac{2\hat{m}}{\sigma_{\pi N}}\left(\frac{3B_0m_\pi\mathring{g}_A^2}{64\pi\mathring{f}_\pi^2}\right)\right]\cdot\left[\frac{\Delta^{(w)}_S}{\left(I^{(w)}_{S,p}+I^{(w)}_{S,n}\right)}\right]. \tag{2.11.3}$$

We have written the ratio as the product of two terms. The first term only depends on the dimensionful low-energy constants parameterizing the relative strength of the LO and NLO couplings. The second term is kinematic in nature, arising from overlap integrals, and is dependent on the spin configuration. Using the numerical values for the low-energy constants and overlap integrals found in Appendices 2.C and 2.F

respectively, one finds

$$\frac{2\hat{m}}{\sigma_{\pi N}}\left(\frac{3B_0 m_\pi \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}\right) = 0.160 \pm 0.029\ , \tag{2.11.4}$$

$$\frac{\Delta_S^{(w)}}{\left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right)} = \begin{cases} 0.261 \pm 0.026 & w = 1 \\ 0.260 \pm 0.027 & w = 2 \\ 0.260 \pm 0.027 & w = 3 \\ 0.259 \pm 0.025 & w = 4 \end{cases}\ . \tag{2.11.5}$$

As can be seen from (2.11.4), the NLO contribution is small compared to the LO contribution just due to the hierarchy of their dimensionful parameters, exactly as expected from ChPT. However, (2.11.5) shows that the NLO term is additionally suppressed by kinematic considerations. As discussed in Section 2.1, the NLO loop contribution depends on $\Delta_S^{(w)}$ which vanishes in the limit of zero momentum transfer. Due to the relatively low momentum transfer involved in coherent conversion, $|q_T| \approx m_\mu$, this further reduce the size of the NLO loop contribution. Taken together, (2.11.4) and (2.11.5) result in the NLO loop contribution being particularly small – roughly 5% of the LO contribution. The NLO loop contribution is sufficiently small that even the parametric uncertainty in the LO contribution is larger than it.

This should be contrasted with the NLO two-body contribution, which is sizable and may appreciably reduce the conversion amplitude. The ratio of the NLO two-nucleon contribution to the LO contribution is

$$\frac{-\alpha_{S,ud}^{(w)}(\text{NLO NN})}{\alpha_{S,ud}^{(w)}(\text{LO})} = \frac{\frac{3B_0 K_F \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2} f_{\text{eff}}^{SI}\left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right)}{\frac{\sigma_{\pi N}}{2\hat{m}}\left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right)} = \frac{2\hat{m}}{\sigma_{\pi N}}\left(\frac{3B_0 K_F \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2}\right) f_{\text{eff}}^{SI} = 0.29^{+0.06}_{-0.16}\ . \tag{2.11.6}$$

While there is significant uncertainty in the value of the two-nucleon contribution due to the one-body averaging procedure of Section 2.5, the two-nucleon contribution is expected to be $15\% - 30\%$ of the LO contribution. As the two-nucleon contribution has the opposite sign of the LO contribution, this can result in the coherent conversion branching ratio decreasing by as much as $25\% - 50\%$. It may seem surprising that the NLO two-nucleon contribution is so much larger than the loop contribution, but this difference is due to the fact that the loop contribution is suppressed for kinematic reasons encapsulated in $\Delta_S^{(w)}$ which are unrelated to the chiral expansion of ChPT. Given the potentially significant impact of the NLO two-nucleon contribution on the sensitivity of $\text{BR}(\mu - e)$ to scalar-mediated interactions, a state-of-the-art many-body computation of this contribution should be performed.

Lastly, we consider the relative size of parametric uncertainties in scalar-mediated conversion to the theoretical uncertainties which arise from our neglect of NNLO contributions. For the LO contribution, the dominant uncertainty is in determining the quark content of the nucleons. Ignoring factors common to all the model-independent parameters, $\alpha_{S,ud}^{(w)}(\text{LO}) = \frac{\sigma_{\pi N}}{2\hat{m}}\left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right)$. Both the isospin average quark mass and the sum of overlap integrals are known to within $\sim 2\%$, see Tables 2.3 and 2.7 of Appendices 2.C and 2.F, respectively. However, the pion-nucleon sigma-term, $\sigma_{\pi N}$, has a relative uncertainty of $\sim 17\%$, see Table 2.3. This is significantly larger than the NLO loop contribution and is comparable in size to the NLO two-nucleon contribution. Even if the NNLO contributions are comparable in size to the NLO loop contribution and are $\sim 5\%$ of the LO term, significant improvements must be made in the determination of the pion-nucleon sigma-term before the theoretical uncertainty from neglecting NNLO corrections becomes relevant.

We now turn our attention to the case of vector-mediated coherent conversion. As shown in Section 2.6, the NLO contributions to the vector-mediated process are spin-dependent and suppressed by a factor of $1/A$. This suppression makes them comparable in size to the already neglected NNLO contributions. Consequently, the model-independent parameters are completely determined by the leading-order contributions

$$\alpha_{V,u}^{(w)} = \sqrt{\frac{m_\mu}{\omega_{\text{capt}}}}\left(\frac{m_\mu}{4\pi v}\right)^2\left(2I_{V,p}^{(w)} + I_{V,n}^{(w)}\right), \tag{2.11.7}$$

$$\alpha_{V,d}^{(w)} = \sqrt{\frac{m_\mu}{\omega_{\text{capt}}}}\left(\frac{m_\mu}{4\pi v}\right)^2\left(I_{V,p}^{(w)} + 2I_{V,n}^{(w)}\right). \tag{2.11.8}$$

These parameters are known to within $\sim 2\%$ and the dominant uncertainty is from the overlap integrals, which in turn is a reflection of uncertainties in the neutron distribution of ^{27}Al, see Table 2.5 of Appendix 2.E. Of course, these are parametric uncertainties and theoretical uncertainties from the neglect of NNLO terms are not included. Given that the NLO contributions were suppressed, it is difficult to estimate the magnitude of the NNLO contributions. However, one naively expects NNLO corrections in $SU(2)$ HBChPT to contribute at roughly the two percent level, and the NLO loop correction for scalar mediated conversion was found to be five percent of the LO term. Thus, one may conservatively estimate the theoretical uncertainty to be roughly five percent.

2.12 Conclusions

In this work, we have performed an analysis of coherent $\mu - e$ conversion at next-to-leading order and have carefully tracked possible sources of uncertainty. The primary

results of this analysis are the expressions Eqs. (2.1.2,2.1.3,2.1.9) and the corresponding model independent parameters of Tables 2.1 and 2.2. These results are summarized in Appendix 2.G.

Starting with a CLFV Lagrangian of the generic form (2.2.1), one may define the Wilson coefficients (2.2.2)-(2.2.4). It is then straightforward to use Eqs. (2.1.2,2.1.3,2.1.9) and the corresponding model-independent parameters to calculate the branching ratio for coherent conversion at next-to-leading order including uncertainties. Similarly, one can use these expressions to determine the permitted regions of parameter space in the event of a detection or non-detection at the upcoming Mu2E and COMET experiments.

In our analysis of scalar-mediated CLFV, we find that the contributions from the next-to-leading order loop diagram are generally small. However, the contributions from the next-to-leading order two-body diagram have the opposite sign of the leading order contribution and could be up to 30% of its size. This can result in an order one change in the branching ratio for a model of CLFV. For a fixed mediator mass, the sensitivity of the upcoming Mu2E and COMET experiments can be reduced by up to a factor of two.

In the case of scalar-mediated conversion, we find that the dominant source of uncertainty is the determination of the nucleon sigma-terms and quark masses. These uncertainties result in a 30% uncertainty in the amplitude for coherent conversion. This severely limits the ability of a single target detector to discriminate different models of CLFV. Generally, these hadronic uncertainties need to be improved by at least a factor of four before NNLO corrections become relevant. Another significant source of uncertainty comes from the one-body averaging of the two-nucleon effective operator. A more careful treatment of this operator including a full many-body treatment of the nucleus would result in improved uncertainties.

Compared to scalar-mediated conversion, vector-mediated conversion has significantly smaller uncertainties. The dominant source of uncertainty comes from the determination of the neutron distribution in ^{27}Al and this only contributes at the two percent level. This is comparable to the theoretical uncertainties from the neglected NNLO corrections. As such, to improve the precision of the vector-mediated case it will be necessary to calculate the NNLO contributions. This will be technically challenging as it requires a careful treatment of the many-body nuclear wavefunction with spin-dependence.

While the analysis presented here is specific to ^{27}Al, it should be straightforward to extend the present approach to other potential targets. As has been shown in the literature [70], multiple targets will be required in the event of detection to determine the channel of CLFV. Given the large hadronic uncertainties in the branching ratio

for scalar-mediated conversion, the use of multiple targets is highly desirable because it should allow an improved determination of CLFV model parameters over what is naively indicated by the hadronic uncertainties.

2.A Chiral Lagrangian

Ignoring the stress-energy tensor, the quark-level CLFV Lagrangian, (2.2.5), written in terms of the CLFV currents, (2.2.6) and (2.2.7), is given by

$$\mathcal{L}_{\text{CLFV}} = \sum_{f=\text{u,d,s}} \frac{1}{\Lambda^2} J_f \bar{q}_f q_f + \sum_{f=\text{u,d,s}} \frac{1}{\Lambda^2} J_f^\nu \bar{q}_f \gamma_\nu q_f. \tag{2.A.1}$$

HBChPT can then be used to relate (2.A.1) to the physics of nucleons and mesons. The resulting effective theory will have several unknown LECs that can be determined by matching onto experimental determinations of hadronic matrix elements. Through the electromagnetic interaction, the matrix elements for the vector current $\langle N| \bar{q}_f \gamma_\nu q_f |N\rangle$ are known in terms of the Pauli and Dirac or Sachs form factors. For scalar quark currents, the relevant matrix elements are $\langle N| m_f \bar{q}_f q_f |N\rangle$, not $\langle N| \bar{q}_f q_f |N\rangle$. To make contact with the known matrix elements, we introduce factors of the quark mass to rewrite the scalar CLFV term of (2.A.1) as

$$\mathcal{L}_{\text{CLFV}} = \sum_{f=\text{u,d,s}} \left(\frac{J_f}{m_f \Lambda^2} \right) m_f \bar{q}_f q_f. \tag{2.A.2}$$

This has the same form as the operator responsible for insertions of the quark mass. Explicitly including this term in the Lagrangian,

$$\mathcal{L} = \sum_{f=\text{u,d,s}} \left[-1 + \left(\frac{J_f}{m_f \Lambda^2} \right) \right] m_f \bar{q}_f q_f \tag{2.A.3}$$

Given the form of (2.A.3), it is apparent that the scalar CLFV current enters the chiral Lagrangian with the same matrix elements as the quark mass insertion. However, the scalar CLFV current also carries inverse factors of Λ^2 and m_f. Thus, up to these additional factors, the LECs of the effective theory can be expressed in terms of known nuclear matrix elements.

In constructing the Lagrangian for HBChPT, one has dynamical fields corresponding to the pions (π^0, $\pi^\pm$) and nucleons (Ψ_P, Ψ_N), along with insertions of the CLFV currents. These currents and dynamical fields can be organized into a collection of objects with well defined transformation properties under the chiral $SU(2)$ symmetry,

$$\phi = \begin{pmatrix} \pi^0 & \sqrt{2}\pi^+ \\ \sqrt{2}\pi^- & -\pi^0 \end{pmatrix} \qquad v^\mu = \tfrac{1}{\Lambda^2} \cdot \tfrac{J_u^\mu - J_d^\mu}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$U = \exp\left(\tfrac{i\phi}{\mathring{f}_\pi}\right) \qquad v^{(s)\mu} = \tfrac{1}{\Lambda^2} \cdot \tfrac{3}{2}\left(J_u^\mu + J_d^\mu\right)$$
$$u = \exp\left(\tfrac{i\phi}{2\mathring{f}_\pi}\right) \qquad v_s^{(s)\mu} = \tfrac{1}{\Lambda^2} J_s^\mu$$
$$N = \begin{pmatrix} \Psi_P \\ \Psi_N \end{pmatrix} \qquad u_\mu = i\left[u^\dagger\left(\partial_\mu - iv_\mu\right)u - u\left(\partial_\mu - iv_\mu\right)u^\dagger\right]$$
$$\chi = -2B_0 \tfrac{1}{\Lambda^2} \begin{pmatrix} J_u & 0 \\ 0 & J_d \end{pmatrix} \qquad \chi_s = -2B_0 \tfrac{1}{\Lambda^2} J_s$$

In these expressions, $\mathring{f}_\pi$ is the tree-level pion decay constant in the chiral limit and B_0 normalizes the scalar sources. The chiral Lagrangian can then be constructed from these objects by considering all possible combinations that are invariant under chiral $SU(2)$ transformations. These terms can be grouped by chiral order so that the chiral Lagrangian corresponds to a well defined expansion in chiral powers. In our power counting, we will assign the CLFV currents J_f and J_f^ν chiral order $\mathcal{O}(1)$ as they do not scale with the quark mass.

As complete expressions for the chiral Lagrangian beyond NLO can be found in the literature [117–120], only terms that include CLFV operators will be listed here. The relevant CLFV terms present in the pionic Lagrangians are given by,

$$\mathcal{L}_{\pi\pi}^{(0)} = \frac{\mathring{f}_\pi^2}{4} Tr\left[\chi\left(U^\dagger + U\right)\right], \tag{2.A.4}$$

$$\mathcal{L}_{\pi\pi}^{(1)} = \frac{\mathring{f}_\pi^2}{2} Tr\left[i\left(\partial_\mu U^\dagger U + \partial_\mu U U^\dagger\right)v^\mu\right]. \tag{2.A.5}$$

Fixing a reference velocity V^μ for HBChPT, the CLFV terms in the pion-nucleon Lagrangians are,

$$\begin{aligned}\mathcal{L}_{\pi N}^{(0)} = \bar{N}\Bigg[&\frac{1}{2}V^\mu\left(u^\dagger v_\mu u + u v_\mu u^\dagger + 2v_\mu^{(s)}\right) + d_1^S\chi_s \\ &+\bar{c}_5\left(\chi\left(U + U^\dagger\right) - \frac{1}{2}Tr\left[\chi\left(U + U^\dagger\right)\right]\right) + \bar{c}_1\, Tr\left[\chi\left(U + U^\dagger\right)\right]\Bigg] N,\end{aligned} \tag{2.A.6}$$

$$\begin{aligned}\mathcal{L}_{\pi N}^{(1)} = \bar{N}\Big[&-i\frac{1}{2M_N}V^\mu V^\nu\left(\partial_\mu v_\nu + 2v_\nu\partial_\mu + \partial_\mu v_\nu^{(s)} + 2v_\nu^{(s)}\partial_\mu\right)\\ &+ i\frac{1}{2M_N}\left(\partial_\mu v^\mu + 2v_\mu\partial^\mu + \partial^\mu v_\mu^{(s)} + 2v_\mu^{(s)}\partial^\mu\right)\\ &- i\frac{1}{2M_N}\left[S^\mu, S^\nu\right]\left(1+\mathring{k}_V\right)\left(\partial_\mu v_\nu - \partial_\nu v_\mu\right)\\ &- i\frac{1}{2M_N}\left[S^\mu, S^\nu\right]\left(1+\mathring{k}_s\right)\left(\partial_\mu v_\nu^{(s)} - \partial_\nu v_\mu^{(s)}\right)\\ &-i\frac{1}{2M_N}\left[S^\mu, S^\nu\right]d_1^V\left(\partial_\mu v_{s\,\nu}^{(s)} - \partial_\nu v_{s\,\mu}^{(s)}\right)\Big]N,\end{aligned} \tag{2.A.7}$$

where S^μ is the spin operator for HBChPT. In these expressions for the pion-nucleon Lagrangian, the coefficients d_1^V and d_1^S have been introduced. These are new LECs that correspond to strange quark operators that do not normally appear in $SU(2)$ HBChPT. The coefficients $\bar{c}_1$ and $\bar{c}_5$ have also been introduced and should be distinguished from the usual LECs c_1 and c_5 of $SU(2)$ HBChPT. As explained previously, the LECs of the CLFV effective theory differ from the usual matrix elements by a factor of $1/m_q$, see, *e.g.*, (2.4.4).

An additional set of terms of the form $\bar{N}\left(Tr\left[\chi_+\right]iV\cdot\partial + \ldots\right)N$ should also appear in $\mathcal{L}_{\pi N}^{(2)}$. However, for coherent $\mu - e$ conversion, these operators will only appear as insertions on an on-shell nucleon line. For on-shell momenta, $V\cdot k$ is of order $\mathcal{O}\left(q^2\right)$ and thus these operators should be treated as $\mathcal{O}\left(q^3\right)$. Thus, these terms can be neglected.

It is worth noting that $\left[S^\mu, S^\nu\right]\left(\partial_\mu v_{s\,\nu}^{(s)} - \partial_\nu v_{s\,\mu}^{(s)}\right)$ is not the only $SU(2)$ invariant term one could write for a generic isoscalar operator. Naively one could write additional terms involving $v_{s\,\mu}^{(s)}$, however, in addition to being isoscalar $v_{s\,\mu}^{(s)}$ carries strangeness. As the nucleons do not carry net strangeness, the only allowable term at this order is $\left[S^\mu, S^\nu\right]\left(\partial_\mu v_{s\,\nu}^{(s)} - \partial_\nu v_{s\,\mu}^{(s)}\right)$.

2.B Low-Energy Constants of the Isoscalar Strange Operators

The low-energy constants that appear in the effective Lagrangian of Appendix 2.A can be assigned numerical values by making contact with experimental results. This is done by matching analytical expressions for nucleon matrix elements in HBChPT onto those from QCD. However, the normalization schemes for nucleon states in QCD and HBChPT are different. HBChPT treats nucleons as non-relativistic fields with the appropriate non-relativistic normalization while QCD is fully relativistic. The differences between these schemes are of order $\mathcal{O}\left(\frac{k^2}{2m_N}\right)$ and thus only enter at NNLO. As such, these schemes may be treated as equivalent for present purposes.

Most of the LECs appear in standard $SU(2)$ HBChPT and are well known, however, the additional constants introduced by including the isoscalar strange operators must be determined. For the scalar strange operator, there is only one unknown LEC. Comparing terms in the chiral and QCD Lagrangians, there is the equivalency

$$-m_s\bar{q}_s q_s \simeq \bar{N}\left[2B_0 d_1^S m_s + \mathcal{O}\left(q^2\right)\right] N. \tag{2.B.1}$$

Using the matrix element for the contribution of the strange quark condensate to the nucleon mass, $m_N f_s^N = \langle N(0)|\, m_s\bar{q}_s q_s\, |N(0)\rangle$, one immediately arrives at the result

$$\frac{m_N f_s^N}{m_s} = -2B_0 d_1^S. \tag{2.B.2}$$

For the vector strange operator, one must compare the vector current to the electric and magnetic nucleon form factors. Written in terms of the Sachs form factors and only keeping terms through NLO [121],

$$\langle N\left(k'\right)|\bar{q}_s\gamma_\mu q_s|\, N\left(k\right)\rangle = \overline{u}\left(k'\right)\left[\gamma_\mu G_E^S\left(q_T^2\right) + \frac{i\sigma_{\mu\nu}q_T^\nu}{2m_N}\left(G_M^S\left(q_T^2\right) - G_E^S\left(q_T^2\right)\right)\right] u\left(k\right). \tag{2.B.3}$$

These form factors are functions of the three-momentum transfer, q_T^2. As the momentum transfer is much smaller than the nucleon mass, the Sachs form factors can be rewritten as series expansions in the momentum transfer. Because nucleons have no net strangeness, the leading order terms of these expansions are $G_E^S\left(q_T^2\right) = \rho_s \frac{q_T^2}{4M_N}$ and $G_M^S\left(q_T^2\right) = \mu_s + \mathcal{O}\left(q_T^2\right)$ where ρ_s is the strangeness radius and μ_s is the strange magnetic moment [122]. Thus, keeping terms only through NLO the matrix element is given by

$$\langle N\left(k'\right)|\bar{q}_s\gamma_\mu q_s|\, N\left(k\right)\rangle = \overline{u}\left(k'\right)\left[-\frac{1}{M_N}\left[S_\mu, S_\nu\right] q_T^\nu \mu_s\right] u\left(k\right). \tag{2.B.4}$$

This matrix element must then be matched onto the corresponding matrix element for the strange vector current in HBChPT. This vector current can be directly read off of the Lagrangian (2.6.1). This fixes the value of the unknown LEC to be $d_1^V = \mu_s$.

2.C Values of Low-Energy Constants and Physical Quantities

All parameters that depend on renormalization are given in the $\overline{MS}$ scheme at $\mu = 2\,\text{GeV}$ except where otherwise noted. All values taken from the world lattice data [101] make use of results from $N_f = 2+1$ simulations whenever possible.

Quantity	Accepted Value	Source	Notes
$\hat{m}$	$3.373 \pm 0.080\,\text{MeV}$	[101, 123–127]	
m_u/m_d	$0.46 \pm 0.03\,\text{MeV}$	[101, 123–127]	
m_s	$92.0 \pm 2.1\,\text{MeV}$	[101, 123–126, 128]	
f_π	$92.07 \pm 0.99\,\text{MeV}$	[101, 129–131]	
$f_\pi/\mathring{f}_\pi$	1.064 ± 0.007	[101, 126, 130, 132–134]	
Σ	$274 \pm 3\,\text{MeV}$	[101, 126, 127, 132, 133]	
g_A	1.2759 ± 0.0045	[111]	
$\sigma_{\pi N}$	$52 \pm 9\,\text{MeV}$	[103]	
f_s	0.043 ± 0.011	[102]	
v	$246.220\,\text{GeV}$	[110]	
m_π	$138.039\,\text{MeV}$	[110]	Isospin averaged pole mass
m_N	$938.919\,\text{MeV}$	[110]	Isospin averaged pole mass
m_μ	$105.658\,\text{MeV}$	[110]	Pole mass
K_F	$238 \pm 5\,\text{MeV}$	[135]	For $^{27}_{13}$Al, linear interpolation
ω_{capt}	$705.4 \pm 1.3\,\text{ms}^{-1}$	[136]	For $^{27}_{13}$Al
ω_{capt}	$464.30 \pm 0.86\,\text{peV}$	Derived	Unit conversion with $\hbar = 1$
ξ	0.37 ± 0.02	Derived	$\xi \equiv \frac{1-m_u/m_d}{1+m_u/m_d}$
$\mathring{f}_\pi$	$86.5 \pm 1.1\,\text{MeV}$	Derived	$\mathring{f}_\pi \equiv f_\pi \left(\frac{f_\pi}{\mathring{f}_\pi}\right)^{-1}$
B_0	$2.75 \pm 0.11\,\text{GeV}$	Derived	$B_0 = \frac{\Sigma^3}{\mathring{f}_\pi^2}$
σ_{sN}	$40 \pm 10\,\text{MeV}$	Derived	$\sigma_{sN} \equiv f_s m_N$

Table 2.3. Table of low-energy constants

2.D Momentum Dependence of Approximate One-Body Interaction

In Section 2.4, it was shown that a two-nucleon operator enters the effective scalar CLFV Lagrangian at NLO. This two-nucleon operator can be reduced to an effective one-nucleon interaction, (2.5.1), by treating the nuclear core as a degenerate Fermi gas and averaging over core nucleons, as explained in Section 2.5. The dependence of (2.5.1) on both the momentum transfered to the nucleus, q_T, and the average of the initial and final nucleon momenta, k, is encapsulated in the dimensionless functions f^{SI} and f^{SD}. For compactness of notation, define the following dimensionless parameters in terms of the Fermi momentum K_F,

$$\overline{q} = \frac{q_T}{K_F}, \quad \overline{k} = \frac{k}{K_F}, \quad \overline{m} = \frac{m_\pi}{K_F}. \tag{2.D.1}$$

In terms of these dimensionless parameters, the functions f^{SI} and f^{SD} can be expressed as,

$$\begin{aligned} f^{SI}\left(\overline{q},\overline{k}\right) = \frac{1}{2\pi}\int_{-1/2}^{1/2} d\beta \Bigg[& 2\left(1+\frac{-\beta\,\overline{k}\cdot\overline{q}+\beta^2\,\overline{q}^2}{\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}+\beta^2\,\overline{q}^2}\right) \\ & -\left(\frac{4\left(\frac{1}{4}-\beta^2\right)\overline{q}^2+3\overline{m}^2}{\sqrt{\left(\frac{1}{4}-\beta^2\right)\overline{q}^2+\overline{m}^2}}\right)\arctan\left(\frac{2\sqrt{\overline{m}^2+\left(\frac{1}{4}-\beta^2\right)\overline{q}^2}}{\overline{m}^2+\frac{1}{4}\overline{q}^2-2\beta\,\overline{k}\cdot\overline{q}+\overline{k}^2-1}\right) \\ & +\frac{1}{2\sqrt{\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}+\beta^2\,\overline{q}^2}}\cdot\Bigg(1+2\overline{m}^2+\left(\frac{3}{4}-4\beta^2\right)\overline{q}^2-\overline{k}^2 \\ & +2\beta\,\overline{k}\cdot\overline{q}+\beta\frac{\left(1+\frac{1}{4}\overline{q}^2+\overline{m}^2+\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}\right)\left(\overline{k}\cdot\overline{q}-\beta\,\overline{q}^2\right)}{\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}+\overline{q}^2}\Bigg) \\ & \cdot\ln\left(\frac{1+2\sqrt{\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}+\beta^2\,\overline{q}^2}+\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}+\frac{1}{4}\overline{q}^2+\overline{m}^2}{1-2\sqrt{\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}+\beta^2\,\overline{q}^2}+\overline{k}^2-2\beta\,\overline{k}\cdot\overline{q}+\frac{1}{4}\overline{q}^2+\overline{m}^2}\right)\Bigg], \end{aligned} \tag{2.D.2}$$

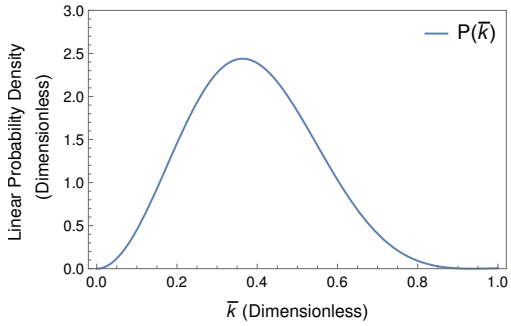
$$f^{SD}\left(\overline{q},\overline{k}\right) = -\frac{1}{2\pi}\int_{-1/2}^{1/2} d\beta \frac{1}{\sqrt{\overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \overline{q}^2}} \cdot \left[\frac{1}{\sqrt{\overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \overline{q}^2}} - \frac{1 + \overline{m}^2 + \overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \frac{1}{4}\overline{q}^2}{\overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \overline{q}^2}\right] \cdot \ln\left[\frac{1 + 2\sqrt{\overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \beta^2\overline{q}^2} + \overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \frac{1}{4}\overline{q}^2 + \overline{m}^2}{1 - 2\sqrt{\overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \beta^2\overline{q}^2} + \overline{k}^2 - 2\beta\,\overline{k}\cdot\overline{q} + \frac{1}{4}\overline{q}^2 + \overline{m}^2}\right]. \tag{2.D.3}$$

To determine the Fermi momentum of $^{27}_{13}$Al, we linearly interpolate between the experimentally measured Fermi momenta of $^{24}_{12}$Mg and $^{40}_{20}$Ca [135]. This results in a Fermi momentum of $K_F = 238 \pm 5\,\text{MeV}$.

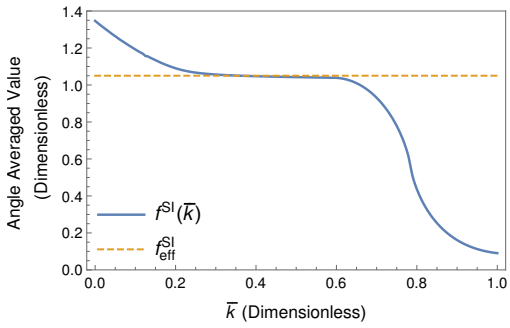
Note that f^{SI} and f^{SD} have an angular dependence due to the presence of $\overline{k}\cdot\overline{q}$. However, f^{SI} and f^{SD} do not vary significantly over the range of possible angular values. As such, the functions can be averaged over all angular values so that they effectively only depend on the magnitude of $\overline{k}$ and $\overline{q}$. Furthermore, it is expected that $|\overline{q}| \approx \frac{m_\mu}{K_F}$ because the process of interest is coherent $\mu - e$ conversion. Thus, f^{SI} and f^{SD} are effectively only functions of $\left|\overline{k}\right|$.

Importantly, the momentum dependence of f^{SI} and f^{SD} is only of interest over the range of momenta common for nucleons in ^{27}Al. Using the model-independent Fourier-Bessel expansion of the proton density distribution [115], the corresponding momentum distribution is shown in Figure 2.2a. It is important to note that Figure 2.2a is a plot of the linear probability density which integrates to unity with respect to $d\left|\overline{k}\right|$.

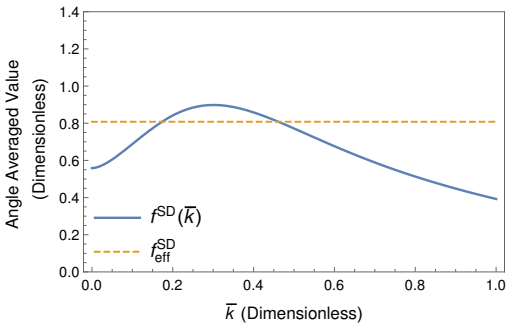
Figures 2.2b and 2.2c show f^{SI} and f^{SD}, respectively, over the same range of momenta with $K_F = 238\,\text{MeV}$. With the goal of arriving at a local approximate interaction in position space, it is necessary to approximate f^{SI} and f^{SD} by constants independent of the nucleon momentum. These constants, f_{eff}^{SI} and f_{eff}^{SD}, are chosen to minimize the weighted RMS error with respect to f^{SI} and f^{SD}. The RMS weights are given by the nucleon momentum distribution. Taking into account both the RMS error and uncertainty in the Fermi momentum, we find $f_{\text{eff}}^{SI} = 1.05 \pm 0.07$ and $f_{\text{eff}}^{SD} = 0.81 \pm 0.12$.



(a) Probability distribution of the magnitude of nucleon momenta in ^{27}Al as a function of the dimensionless momentum.



(b) The angle averaged value of f^{SI} and its constant approximation f^{SI}_{eff} as a function of the dimensionless average momentum.



(c) The angle averaged value of f^{SD} and its constant approximation f^{SD}_{eff} as a function of the dimensionless average momentum.

Figure 2.2. Momentum dependence of the functions f^{SI} and f^{SD}

2.E Nuclear Density Parameters

The proton density distribution of a nucleus can be parameterized in a model-independent manner using the Fourier-Bessel expansion

$$\rho_p(r) = \begin{cases} \sum_n a_n j_0\left(\frac{n\pi r}{R}\right) & r \leq R \\ 0 & r > R \end{cases}. \tag{2.E.1}$$

This distribution is normalized such that $\int 4\pi r^2 \rho(r)\, dr = Z$ and depends on the cutoff radius, R, and the magnitude of the various components, a_n. The values for ^{27}Al are given in Table 2.4 as determined by electron scattering experiments [115].

The neutron density distribution of a nucleus is usually given in terms of the two-parameter Fermi distribution,

$$\rho_n(r) = \frac{\rho_0}{1 + e^{\frac{r-c}{z}}}. \tag{2.E.2}$$

The normalization factor ρ_0 is chosen such that $\int 4\pi r^2 \rho(r)\, dr = A - Z$ while the thickness parameter, z, and radial parameter, c describe the shape of the distribution. The values of these parameters for ^{27}Al are given in Table 2.5 where the experimental and systematic uncertainties have been combined [116].

Table 2.4. Parameters for proton density distribution

Quantity	Accepted Value
$\langle r^2 \rangle_p^{1/2}$ [fm.]	3.035 ± 0.002
R [fm.]	7.0
a_1 $\left[\text{fm.}^{-3}\right]$	$4.3418 \cdot 10^{-1}$
a_2 $\left[\text{fm.}^{-3}\right]$	$6.0298 \cdot 10^{-1}$
a_3 $\left[\text{fm.}^{-3}\right]$	$2.8950 \cdot 10^{-2}$
a_4 $\left[\text{fm.}^{-3}\right]$	$-2.3522 \cdot 10^{-1}$
a_5 $\left[\text{fm.}^{-3}\right]$	$-7.9791 \cdot 10^{-2}$
a_6 $\left[\text{fm.}^{-3}\right]$	$2.3010 \cdot 10^{-2}$
a_7 $\left[\text{fm.}^{-3}\right]$	$1.0794 \cdot 10^{-2}$
a_8 $\left[\text{fm.}^{-3}\right]$	$1.2574 \cdot 10^{-3}$
a_9 $\left[\text{fm.}^{-3}\right]$	$-1.3021 \cdot 10^{-3}$
a_{10} $\left[\text{fm.}^{-3}\right]$	$5.6563 \cdot 10^{-4}$
a_{11} $\left[\text{fm.}^{-3}\right]$	$-1.8011 \cdot 10^{-4}$
a_{12} $\left[\text{fm.}^{-3}\right]$	$4.2869 \cdot 10^{-5}$

Table 2.5. Parameters for neutron density distribution

Quantity	Accepted Value
$\langle r^2 \rangle_n^{1/2}$ [fm.]	3.17 ± 0.11
c [fm.]	3.18 ± 0.19
z [fm.]	0.535

2.F Model Independent Overlap Integrals

In Section 2.10, the coherent conversion rate (2.10.4) was expressed as a sum of transition probabilities over eight possible spin configurations. However, there is a two-fold symmetry in the choice of overall sign for the spins. Thus, there are only four independent configurations. For compactness of notation, an index $w \in \{1,2,3,4\}$ is used to denote each unique configuration. The relationship between all possible spin configurations and w is given in Table 2.6.

For a fixed spin configuration, one can perform the position space integral over the leptonic part of (2.10.5) to arrive at a function of the momentum transfer. For the case of scalar-mediated conversion, these are given in terms of the muon and electron wavefunctions (Section 2.8) by the dimensionless functions

$$Z_S^{(1)}\left(|q_T|\right) = \sqrt{m_\mu}\int dx \frac{1}{2\pi^2}\,|x|^2\, j_0\left(|x|\,|q_T|\right)\left(g_{-1}^{(e)}(x)\,g_{-1}^{(\mu)}(x) + f_{-1}^{(e)}(x)\,f_{-1}^{(\mu)}(x)\right), \tag{2.F.1}$$

$$Z_S^{(2)}\left(|q_T|\right) = \sqrt{m_\mu}\int dx \frac{1}{8\pi}\,|x|^2\, j_1\left(|x|\,|q_T|\right)\left(g_{-1}^{(e)}(x)\,f_{-1}^{(\mu)}(x) - f_{-1}^{(e)}(x)\,g_{-1}^{(\mu)}(x)\right), \tag{2.F.2}$$

$$Z_S^{(3)}\left(|q_T|\right) = \sqrt{m_\mu}\int dx \frac{1}{8\pi}\,|x|^2\, j_1\left(|x|\,|q_T|\right)\left(g_{+1}^{(e)}(x)\,g_{-1}^{(\mu)}(x) + f_{+1}^{(e)}(x)\,f_{-1}^{(\mu)}(x)\right), \tag{2.F.3}$$

$$Z_S^{(4)}\left(|q_T|\right) = \sqrt{m_\mu}\int dx \frac{1}{2\pi^2}\,|x|^2\, j_0\left(|x|\,|q_T|\right)\left(f_{+1}^{(e)}(x)\,g_{-1}^{(\mu)}(x) - g_{+1}^{(e)}(x)\,f_{-1}^{(\mu)}(x)\right). \tag{2.F.4}$$

One can then perform the remaining momentum integrals of (2.10.5) in a model-independent manner. The only CLFV operator that depends on momentum transfer is the arccot term in the NLO loop contribution of (2.5.2). This term will be associated with the overlap integral $\tilde{I}_{S,\alpha}^{(w)}$. All other CLFV operators are independent of momentum transfer and will be accompanied by the overlap integral $I_{S,\alpha}^{(w)}$. Defining the Fourier transformed nucleon density as $\widetilde{\rho}_\alpha(k) = \widetilde{\psi}_\alpha^*(k)\,\widetilde{\psi}_\alpha(k)$, the definitions for these two overlap integrals are given by

$$I_{S,\alpha}^{(w)} = \frac{1}{m_\mu^{5/2}}\int dq_T \int dq_A\,|q_T|^2\,|q_A|^2\, Z_S^{(w)}\left(|q_T|\right)\widetilde{\rho}_\alpha\left(\frac{1}{2}\sqrt{|q_T|^2+|q_A|^2}\right), \tag{2.F.5}$$

$$\tilde{I}_{S,\alpha}^{(w)} = \frac{1}{m_\mu^{5/2}}\int dq_T \int dq_A\,|q_T|^2\,|q_A|^2\, Z_S^{(w)}\left(|q_T|\right)\widetilde{\rho}_\alpha\left(\frac{1}{2}\sqrt{|q_T|^2+|q_A|^2}\right) \cdot \frac{2+X_\pi}{\sqrt{X_\pi}}\mathrm{arccot}\left(\frac{2}{\sqrt{X_\pi}}\right). \tag{2.F.6}$$

In the case of vector-mediated conversion, one instead finds that the leptonic part of (2.10.5) can be reduced to the functions

$$Z_V^{(1)}\left(|q_T|\right) = \sqrt{m_\mu}\int dx \frac{1}{2\pi^2}\left|x\right|^2 j_0\left(|x|\,|q_T|\right)\left(g_{-1}^{(e)}\left(x\right)g_{-1}^{(\mu)}\left(x\right) - f_{-1}^{(e)}\left(x\right)f_{-1}^{(\mu)}\left(x\right)\right), \tag{2.F.7}$$

$$Z_V^{(2)}\left(|q_T|\right) = \sqrt{m_\mu}\int dx \frac{1}{8\pi}\left|x\right|^2 j_1\left(|x|\,|q_T|\right)\left(g_{-1}^{(e)}\left(x\right)f_{-1}^{(\mu)}\left(x\right) + f_{-1}^{(e)}\left(x\right)g_{-1}^{(\mu)}\left(x\right)\right), \tag{2.F.8}$$

$$Z_V^{(3)}\left(|q_T|\right) = \sqrt{m_\mu}\int dx \frac{1}{8\pi}\left|x\right|^2 j_1\left(|x|\,|q_T|\right)\left(g_{+1}^{(e)}\left(x\right)g_{-1}^{(\mu)}\left(x\right) - f_{+1}^{(e)}\left(x\right)f_{-1}^{(\mu)}\left(x\right)\right), \tag{2.F.9}$$

$$Z_V^{(4)}\left(|q_T|\right) = -\sqrt{m_\mu}\int dx \frac{1}{2\pi^2}\left|x\right|^2 j_0\left(|x|\,|q_T|\right)\left(f_{+1}^{(e)}\left(x\right)g_{-1}^{(\mu)}\left(x\right) + g_{+1}^{(e)}\left(x\right)f_{-1}^{(\mu)}\left(x\right)\right). \tag{2.F.10}$$

Unlike the case of scalar-mediated CLFV, no term in the vector-mediated CLFV Lagrangian depends on the momentum transfer. The only type of overlap integral is

$$I_{V,\alpha}^{(w)} = \frac{1}{m_\mu^{5/2}}\int dq_T \int dq_A \left|q_T\right|^2 \left|q_A\right|^2 Z_V^{(w)}\left(|q_T|\right)\widetilde{\rho}_\alpha\left(\frac{1}{2}\sqrt{\left|q_T\right|^2 + \left|q_A\right|^2}\right). \tag{2.F.11}$$

The numerical values of $I_{S,\alpha}^{(w)}$, $\tilde{I}_{S,\alpha}^{(w)}$, and $I_{V,\alpha}^{(w)}$ can readily be calculated using the proton and neutron distributions of Appendix 2.E along with the muon and electron wavefunctions calculated from them (see Section 2.8). The values of $I_{S,\alpha}^{(w)}$, $\tilde{I}_{S,\alpha}^{(w)}$, and $I_{V,\alpha}^{(w)}$ along with their uncertainties are cataloged in Table 2.7.

κ_i	μ_i	κ_f	μ_f	w
-1	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	1
-1	$+\frac{1}{2}$	-1	$+\frac{1}{2}$	1
-1	$-\frac{1}{2}$	-1	$+\frac{1}{2}$	2
-1	$+\frac{1}{2}$	-1	$-\frac{1}{2}$	2
-1	$-\frac{1}{2}$	+1	$+\frac{1}{2}$	3
-1	$+\frac{1}{2}$	+1	$-\frac{1}{2}$	3
-1	$-\frac{1}{2}$	+1	$-\frac{1}{2}$	4
-1	$+\frac{1}{2}$	+1	$+\frac{1}{2}$	4

Table 2.6. Table of spin configurations

Proton Overlap Integral	Value	Neutron Overlap Integral	Value
$I^1_{S,p}$	7.58	$I^1_{S,n}$	7.58±0.24
$I^2_{S,p}$	5.50	$I^2_{S,n}$	5.49±0.17
$I^3_{S,p}$	−5.53	$I^3_{S,n}$	−5.52±0.17
$I^4_{S,p}$	−7.56	$I^4_{S,n}$	−7.55±0.24
$\tilde{I}^1_{S,p}$	9.57	$\tilde{I}^1_{S,n}$	9.55±0.31
$\tilde{I}^2_{S,p}$	6.93	$\tilde{I}^2_{S,n}$	6.92±0.23
$\tilde{I}^3_{S,p}$	−6.96	$\tilde{I}^3_{S,n}$	−6.96±0.23
$\tilde{I}^4_{S,p}$	−9.52	$\tilde{I}^4_{S,n}$	−9.51±0.30
$I^1_{V,p}$	7.35	$I^1_{V,n}$	7.32±0.24
$I^2_{V,p}$	−5.79	$I^2_{V,n}$	−5.76±0.19
$I^3_{V,p}$	−5.81	$I^3_{V,n}$	−5.79±0.19
$I^4_{V,p}$	7.31	$I^4_{V,n}$	7.29±0.24

Table 2.7. Table of overlap integrals

2.G Formula for the Branching Ratio

Given a CLFV Lagrangian of the form (2.2.1), one can define the Wilson coefficients (2.2.2)-(2.2.4). As explained in Section 2.10, the branching ratio for coherent $\mu - e$ conversion can be written as a sum over separate amplitudes for each spin configuration, (2.10.4). Accounting for symmetry in the spin configurations, the index $w \in \{1, 2, 3, 4\}$ indicates the four independent spin configurations of the system as outlined in Table 2.6 of Appendix 2.F. Written as a sum over these four independent configuration, this yields the master equation for the branching ratio, (2.1.2).

Each conversion amplitude for a specific spin configuration can then be expressed in terms of Wilson coefficients and a set of model-independent parameters. This is done for scalar-mediated conversion in (2.1.3) and for vector-mediated conversion in (2.1.9). The only model dependent-parameters that appear in these expressions are the Wilson coefficients; all model-independent parameters have been absorbed into the definitions of the α's. Using the definition of $\Delta_S^{(w)}$ from (2.11.2), these are defined as

$$\alpha_{S,ud}^{(w)} = \sqrt{\frac{m_\mu}{\omega_{\text{capt}}}} \left(\frac{m_\mu}{4\pi v}\right)^2 \left[\frac{\sigma_{\pi N}}{2\hat{m}} \left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right) - \frac{3B_0 K_F \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2} f_{\text{eff}}^{SI} \left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right) - \frac{3B_0 m_\pi \mathring{g}_A^2}{64\pi \mathring{f}_\pi^2} \Delta_S^{(w)}\right], \tag{2.G.1}$$

$$\alpha_{S,s}^{(w)} = \sqrt{\frac{m_\mu}{\omega_{\text{capt}}}} \left(\frac{m_\mu}{4\pi v}\right)^2 \frac{\sigma_{sN}}{m_s} \left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right), \tag{2.G.2}$$

$$\alpha_{S,\Theta}^{(w)} = \sqrt{\frac{m_\mu}{\omega_{\text{capt}}}} \left(\frac{m_\mu}{4\pi v}\right)^2 \left(I_{S,p}^{(w)} + I_{S,n}^{(w)}\right), \tag{2.G.3}$$

$$\alpha_{V,u}^{(w)} = \sqrt{\frac{m_\mu}{\omega_{\text{capt}}}} \left(\frac{m_\mu}{4\pi v}\right)^2 \left(2I_{V,p}^{(w)} + I_{V,n}^{(w)}\right), \tag{2.G.4}$$

$$\alpha_{V,d}^{(w)} = \sqrt{\frac{m_\mu}{\omega_{\text{capt}}}} \left(\frac{m_\mu}{4\pi v}\right)^2 \left(I_{V,p}^{(w)} + 2I_{V,n}^{(w)}\right). \tag{2.G.5}$$

The quantities $I_{S,x}^{(w)}$, $\tilde{I}_{S,x}^{(w)}$, and $I_{V,x}^{(w)}$ are the overlap integrals defined in Appendix 2.F and given numerically in Table 2.7. The quantity $f_{\text{eff}}^{SI} = 1.05^{+0.07}_{-0.53}$ characterizes the approximate one-body interaction which is discussed in Section 2.5 and Appendix 2.D. The remaining physical constants are given in Table 2.3 of Appendix 2.C.

As the α parameters are model-independent, they can be calculated in advance and their numerical values and uncertainties are given in Tables 2.1 and 2.2 of Section 2.1. In the case of scalar-mediated conversion, the LO contributions and those of the loop diagram and two-nucleon diagram that enter at NLO can be analyzed separately.

The Bayesian Second Law of Thermodynamics

We derive a generalization of the Second Law of Thermodynamics that uses Bayesian updates to explicitly incorporate the effects of a measurement of a system at some point in its evolution. By allowing an experimenter's knowledge to be updated by the measurement process, this formulation resolves a tension between the fact that the entropy of a statistical system can sometimes fluctuate downward and the information-theoretic idea that knowledge of a stochastically-evolving system degrades over time. The Bayesian Second Law can be written as $\Delta H(\rho_m, \rho) + \langle \mathcal{Q} \rangle_{F|m} \geq 0$, where $\Delta H(\rho_m, \rho)$ is the change in the cross entropy between the original phase-space probability distribution ρ and the measurement-updated distribution ρ_m, and $\langle \mathcal{Q} \rangle_{F|m}$ is the expectation value of a generalized heat flow out of the system. We also derive refined versions of the Second Law that bound the entropy increase from below by a non-negative number, as well as Bayesian versions of integral fluctuation theorems. We demonstrate the formalism using simple analytical and numerical examples.

3.1 Introduction

The Second Law of Thermodynamics encapsulates one of the most important facts about the macroscopic world: entropy increases over time. There are, however, a number of different ways to define "entropy," as well as corresponding controversies over how to best understand the Second Law. In this paper we offer a formulation of the Second Law that helps resolve some of the tension between different approaches, by explicitly including the effects of the measurement process on our knowledge of the state of the system. This Bayesian Second Law (BSL) provides a new tool for analyzing the evolution of statistical systems, especially for small numbers of particles and short times, where downward fluctuations in entropy can be important.

One way to think about entropy and the Second Law, due to Boltzmann, coarse-grains the phase space Γ of a system into macrostates. The entropy of a microstate x is then given by $S = \log \Omega_x$, where Ω_x is the volume of the macrostate to which x belongs. (Throughout this paper we set Boltzmann's constant k_B equal to unity). The coarse-graining itself is subjective, but once it is fixed there is a definite entropy objectively associated with each microstate. Assuming that the system starts in a low-entropy state (the "Past Hypothesis"), the Second Law simply reflects the fact that undirected evolution is likely to take the state into ever-larger macrostates: there are more ways to be high-entropy than to be low-entropy. The Second Law is statistical in the sense that random fluctuations into lower-entropy states, while rare, are certainly possible. In many contexts of interest to modern science, from nanoscale physics to biology, these fluctuations are of crucial importance and the study of "fluctuation theorems" has garnered considerable attention in recent years [39, 41, 42, 137–141].

Another perspective on entropy, associated with Gibbs in statistical mechanics and Shannon [142] in the context of information theory, starts with a normalized probability distribution $\rho(x)$ on phase space, and defines the entropy as $S = -\int dx\, \rho(x) \log \rho(x)$. In contrast with the Boltzmann formulation, in this version the entropy characterizes the state of our knowledge of the system, rather than representing an objective fact about the system itself. The more spread-out and uncertain a distribution is, the higher its entropy. The Second Law, in this view, represents the influence of stochastic dynamics on the evolution of the system, for example due to interactions with a heat bath, under the influence of which we know less and less about the microstate of the system as time passes.

For many purposes, the Gibbs/Shannon formulation of entropy and the Second Law is more convenient to use than the Boltzmann formulation. However, it raises a puzzle: how can entropy ever fluctuate downward? In an isolated system evolving according to Hamiltonian dynamics, the Gibbs entropy is strictly constant, rather than

increasing; for a system coupled to a heat bath with no net energy transfer, it tends to monotonically increase, asymptoting to a maximum equilibrium value. Ultimately, this is because the Gibbs entropy characterizes our knowledge of the microstate of the system, which only diminishes with time.[1]

We can, of course, actually *observe* the system; if we do so, we will (extremely) occasionally notice that it has fluctuated into what we would characterize as a low-entropy state from Boltzmann's perspective. The air in a room could fluctuate into one corner, for example, or a cool glass of water could evolve into a warm glass of water containing an ice cube. To reconcile this real physical possibility with an information-centric understanding of entropy, we need to explicitly account for the impact of the act of measurement on our knowledge of the system. This is the task of Bayesian analysis, which shows us how to update probability distributions in the face of new information [144, 145]. Since the advent of Maxwell's demon, measurement in the context of statistical mechanics has been explored extensively [146]. This has resulted in a body of literature linking information-theoretic quantities to thermodynamic variables [147, 148]. However, such analyses only examine the impact of measurement at the point in time when it is performed. In the present work, we observe that such measurements also contain information about the state of the system at earlier points in time that are hitherto unaccounted for. This results in novel modifications of the Second Law.

The setup we consider consists of a classical system coupled to an environment. The dynamics of the system are stochastic, governed by transition probabilities, either due to intrinsic randomness in the behavior of the system, or to the unpredictable influence of the environment. An experimental protocol is determined by a set of time-dependent parameters, which may be thought of as macroscopic features (such as the location of a piston) controlled by the experimenter. The experimenter's initial knowledge of the system is characterized by some probability distribution; as the system is evolved under the protocol for some period of time, this probability distribution also evolves. At the end of the experiment, the experimenter performs a measurement. Bayes's Theorem tells us how to update our estimates about the system based on the outcome of the measurement; in particular, we can use the measurement outcome to update the final probability distribution, but also to update the *initial* distribution. The BSL is a relation between the original (non-updated) distributions, the updated distributions,

[1] Boltzmann himself also studied a similar formulation of entropy, which he used to prove his H-theorem. The difference is that the H-functional represents N particles in one 6-dimensional single-particle phase space, rather than in a $6N$-dimensional multi-particle phase space. This is not a full representation of the system, as it throws away information about correlations between particles. The corresponding dynamics are not reversible, and entropy increases [143].

and a generalized heat transfer between the system and the environment.

The Second Law contains information about irreversibility; a crucial role in our analysis is played by the relationship between transition probabilities forward in time and "reversed" probabilities backward in time. Consider a situation in which the system in question is an egg, and the experiment consists of holding the egg up and then dropping it. To be precise, the experimental protocol, which we will call the "forward" protocol, is for the experimenter to hold the egg in the palm of her open hand, and then to turn her hand over after a specified amount of time. The initial probability distribution for the particles that make up the egg is one that corresponds to an intact egg in the experimenter's hand. With overwhelming probability the forward protocol applied to this initial state will result in an egg on the floor, broken.

This experiment is clearly of the irreversible type, but we should be careful about why and how it is irreversible. If we reversed the velocities of every particle in the universe, then time would run backward and the egg would reconstitute itself and fly back up into the experimenter's hand. This sort of fundamental reversibility is not what concerns us. For us, irreversibility means that there are dissipative losses to the environment: in particular, there are losses of *information* as the state of the system interacts with that of the environment. This information loss is what characterizes irreversibility. From the theoretical viewpoint, we should ask what would happen if all of the velocities of the broken egg particles were instantaneously reversed, leaving the environment alone. Again with overwhelming probability, the egg would remain broken on the floor. To make sure the time-dependent actions of the experimenter do not affect this conclusion, we should also instruct the experimenter to run her experiment in reverse: she should begin with her palm facing downward while the egg is broken on the floor, and then turn it upward after a certain amount of time. In this example, the effect of reversing the experimental procedure is negligible; the probability that the egg will reassemble itself and hop up into her hand is not zero, but it is extremely small.

The generalization beyond the egg dropping experiment is clear. We have a system, an environment, and an experimenter who executes a forward protocol, which means a macroscopic time-dependent influence on the dynamics of the system. The environmental interactions with the system are deterministic but unknown to the experimenter, and so the system evolves stochastically from her point of view. She assigns probabilities to trajectories the system might take through phase space. We will call these the "forward" probabilities. To isolate potential irreversibility in the system, we consider reversing all of the velocities of the system's particles in its final state, and then executing the "reverse" protocol, which is just the forward protocol backward. The environment still interacts in an unknown way, so the system again evolves stochastically. The probabilities that the experimenter assigns to trajectories in this reversed setup

are called the reverse probabilities.

To get precise versions of the Second Law, we will consider a particular information-theoretic measure of the difference between the forward and reverse probabilities, known as the relative entropy or Kullback-Leibler divergence [149]. The relative entropy of two probability distributions is always non-negative, and vanishes if and only if the two distributions are identical. The relative entropy of the forward and reverse probability distributions on phase space trajectories is a measure of the irreversibility of the system, and the non-negativity of that relative entropy is a precise version of the Second Law.

The inclusion of Bayesian updates as the result of an observation at the end of the protocol leads to the Bayesian Second Law. The BSL can be written in several ways, one of which is:

$$\Delta H(\rho_m, \rho) + \langle \mathcal{Q} \rangle_{F|m} \geq 0. \tag{3.1.1}$$

Here, ρ is the probability distribution without updating, and ρ_m is the updated distribution after obtaining measurement outcome m. $H = -\int \rho_m \log \rho$ is the cross entropy between the two distributions. The cross entropy is the sum of the entropy of ρ_m and the relative entropy of ρ_m with respect to ρ; it can be thought of as the average amount we would learn about the system by being told its precise microstate, if we thought it was in one distribution (the original ρ), but it was actually in another (the updated ρ_m). Like the ordinary entropy, this is a measure of uncertainty: the more information contained in the (unknown) microstate, the greater the uncertainty. However, the cross entropy corrects for our false impression of the distribution. The difference in the cross entropy between the initial and final times is ΔH, and $\langle \mathcal{Q} \rangle_{F|m}$ is the expectation value of a generalized heat transfer between the system and the environment, which contains information about the irreversibility of the system's dynamics. Thus, at zero heat transfer, the BSL expresses the fact that our uncertainty about the system is larger at the time of measurement, even after accounting for the measurement outcome.

The relative entropy is not only non-negative, it is monotonic: if we apply a stochastic (probability-conserving) operator to any two distributions, the relative entropy between them stays constant, or decreases. We can use this fact to prove refined versions of both the ordinary and Bayesian Second Laws, obtaining a tighter bound than zero to the expected entropy change plus heat transfer. This new lower bound is the relative entropy between the initial probability distribution and one that has been cycled through forward and reverse evolution, and therefore characterizes the amount of irreversibility in the evolution.

We also apply our implementation of Bayesian updating to integral fluctuation theorems, extending such theorems to subsets of experimental realizations conditioned on particular measurement outcomes. Lastly, we illustrate the BSL in the context

of some simple models. These include deriving Boltzmann's version of the Second Law within our formalism, and studying the numerical evolution of a randomly driven harmonic oscillator.

3.2 Setup

3.2.1 The System and Evolution Probabilities

We are primarily concerned with dynamical systems that undergo non-deterministic evolution, typically due to interactions with an environment about which the experimenter has no detailed knowledge. The effect of the unknown environment is to induce effectively stochastic evolution on the system; as such, we can only describe the state and subsequent time evolution of the system probabilistically [150]. We are considering classical mechanics, where probabilities only arise due to the ignorance of the experimenter, including ignorance of the state of the environment. Analogous equations would apply more generally to truly stochastic systems, or to stochastic models of dynamical systems.

The state of the system at time t is therefore a random variable X_t taking values in a space of states Γ. We will refer to Γ as "phase space," as if it were a conventional Hamiltonian system, although the equations apply equally well to model systems with discrete state spaces. Because the evolution is non-deterministic, we can only give a probability that the system is in state x at time t, which we write as $P(X_t = x)$. This is a true probability in the discrete case; in the continuous case it is more properly a probability density that should be integrated over a finite region of Γ to obtain a probability, but we generally will not draw this distinction explicitly. For notational convenience, we will often write this probability as a distribution function,

$$\rho_t(x) \equiv P(X_t = x), \tag{3.2.1}$$

which is normalized so that $\int \rho_t(x)\, dx = 1$.

The experimenter has two roles: to manipulate a set of external control parameters defining the experimental protocol, and to perform measurements on the system. All measurements are assumed to be "ideal"; that is, the act of measuring any given property of the system is assumed to induce no backreaction on its state, and we do not track the statistical properties of the measuring device.

We will primarily be studying experiments that take place over a fixed time interval τ. The experimental protocol is fully specified by the history of a set of external control parameters that can change over this time interval, $\lambda_i(t)$. The control parameters λ_i

specify the behavior of various external potentials acting on the system, such as the volume of a container or the frequency of optical tweezers. We will refer to the set $\lambda(t) = \{\lambda_i(t)\}$ of control parameters as functions of time as the "forward protocol."

The forward protocol and the dynamics of the system together determine the forward transition function, π_F, which tells us the probability that the system evolves from an initial state x at $t = 0$ to a final state x' at $t = \tau$:

$$\pi_F(x \to x') \equiv P(X_\tau = x' | X_0 = x; \lambda(t)). \tag{3.2.2}$$

The transition function π_F is a conditional probability, normalized so that the system ends up somewhere with probability one:

$$\int \pi_F(x \to x') dx' = 1. \tag{3.2.3}$$

The forward transition function evolves the initial distribution to the final distribution,

$$\rho_\tau(x') = \int dx\, \rho_0(x) \pi_F(x \to x'). \tag{3.2.4}$$

A central role will be played by the joint probability that the system begins at x and ends up a time τ later at x',[2]

$$P_F(x, x') \equiv P(X_0 = x, X_\tau = x') = \rho_0(x) \pi_F(x \to x'), \tag{3.2.5}$$

which is normalized so that $\int P(x, x')\, dx dx' = 1$. By summing the joint probability over x or x' we obtain the distribution functions $\rho_\tau(x')$ or $\rho_0(x)$, respectively:

$$\begin{aligned} \rho_\tau(x') &= \int P_F(x, x') dx, \\ \rho_0(x) &= \int P_F(x, x') dx'. \end{aligned} \tag{3.2.6}$$

We close this subsection with a brief review on the probabilities of phase-space trajectories. The rules of conditional probability allow us to break up the transition functions based on subdivisions of the time interval $[0, \tau]$. For the special case of a Markov process, we have the identity

$$\begin{aligned} \pi_F(x \to x') = \int [dx]\, & P(X_\tau = x' | X_{t_N} = x_N) \\ & \times P(X_{t_N} = x_N | X_{t_{N-1}} = x_{N-1}) \cdots P(X_{t_1} = x_1 | X_0 = x), \end{aligned} \tag{3.2.7}$$

[2]Here and below we will mostly omit the dependence on the control parameters $\lambda(t)$ from the notation for brevity. They will return in Section 3.2.3 when we discuss time-reversed experiments.

where $[dx]$ is the product of all the dx_k and we choose $t_k = k\tau/(N+1)$. This is familiar as a discretization of the path integral, and in the continuum limit we would write

$$\pi_F(x \to x') = \int_{x(0)=x}^{x(\tau)=x'} \mathcal{D}x(t)\, \pi_F[x(t)]. \tag{3.2.8}$$

The functional $\pi_F[x(t)]$ is a probability density on the space of trajectories with fixed initial position, but with the final position free. This functional has previously been introduced in the literature for both classical and quantum systems [151, 152]. To get a probability density on the space of trajectories with two free endpoints, we just have to multiply $\pi_F[x(t)]$ by the initial distribution $\rho_0(x)$. The result, which we call $P_F[x(t)]$, is the path-space version of the joint distribution $P_F(x, x')$. We will not make heavy use of these path-space quantities below, but the formal manipulations we make with the ordinary transition function and joint distribution can be repeated exactly with the path-space distributions, and occasionally we will comment on the path-space versions of our results.

3.2.2 Measurement and Bayesian Updating

The probability density on phase space can also change through Bayesian updates when a measurement is made: the experimenter modifies her probabilities to account for the new information. We will restrict ourselves to measurements performed at time τ, the end of the experiment, though it is simple to extend the results to more general measurement protocols. The measurement outcome is a random variable M that only depends on the state of the system at time τ, not on the prior history of the system. The measurement is then characterized by the function

$$\begin{aligned} P(m|x') &\equiv P(M = m|X_\tau = x') \\ &= \text{probability of measurement outcome } m \text{ given state } x' \text{ at time } \tau. \end{aligned} \tag{3.2.9}$$

The updated phase space distribution at time τ is obtained by Bayes's rule, which in this case takes the form

$$\rho_{\tau|m}(x') \equiv P(X_\tau = x'|M = m) = \frac{P(m|x')}{P(m)}\rho_\tau(x'). \tag{3.2.10}$$

Here the denominator is $P(m) \equiv \int P(m|y')\rho_\tau(y')dy'$, and serves as a normalization factor.

If we know the transition function, we can also update the phase space distribution at any other time based on the measurement outcome at time τ. Below we will make use of the updated initial distribution:

$$\rho_{0|m}(x) \equiv P(X_0 = x|M = m) = \frac{\rho_0(x) \int dx'\ \pi_F(x \to x')P(m|x')}{P(m)}. \tag{3.2.11}$$

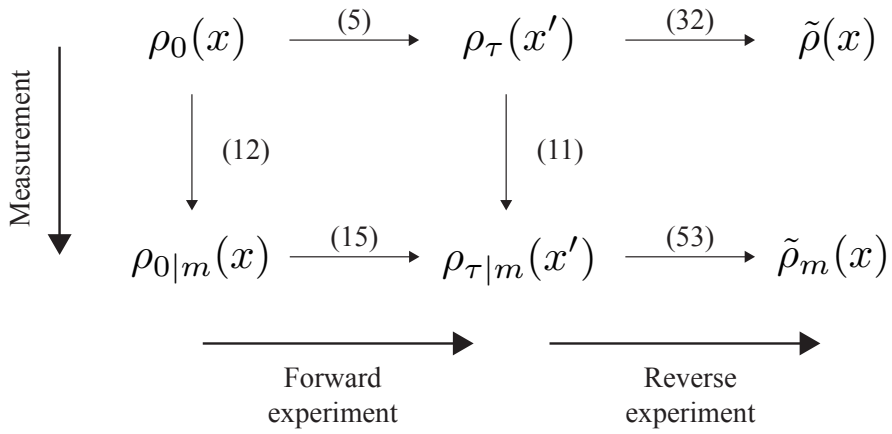


Figure 3.1. Relationships between the various distribution functions we define: the original distribution $\rho_0(x)$, its time-evolved version $\rho_\tau(x')$, their corresponding Bayesian-updated versions $\rho_{0|m}(x)$ and $\rho_{\tau|m}(x')$, and the cycled distributions $\tilde{\rho}(x)$ and $\tilde{\rho}_m(x)$ discussed in Sections 3.3.2 and 3.4.3. Equation numbers refer to where the distributions are related to each other.

This reflects our best information about the initial state of the system given the outcome of the experiment; $\rho_{0|m}(x)$ is the probability, given the original distribution $\rho_0(x)$ and the measurement outcome m at time $t = \tau$, that the system was in state x at time $t = 0$. For example, we may initially be ignorant about the value of an exactly conserved quantity. If we measure it at the end of the experiment then we know that it had to have the same value at the start; this could mean a big difference between ρ_0 and $\rho_{0|m}$, though often the effects will be more subtle. The various distribution functions we work with are summarized in Figure 3.1 and listed in Table 3.1.

Finally, we can update the forward transition functions,

$$\pi_{F|m}(x \to x') \equiv P(X_\tau = x'|X_0 = x, M = m) = \frac{\pi_F(x \to x')P(m|x')}{\int dy'\ \pi_F(x \to y')P(m|y')}, \tag{3.2.12}$$

and the joint distributions,

$$P_{F|m}(x, x') \equiv P(X_0 = x, X_\tau = x'|M = m) = \frac{P(m|x')}{P(m)} P_F(x, x') = \rho_{0|m}(x)\pi_{F|m}(x \to x'), \tag{3.2.13}$$

based on the measurement outcome. As we would expect, the updated transition function evolves the updated distribution from the initial to the final time:

$$\rho_{\tau|m}(x') = \int dx\, \rho_{0|m}(x)\pi_{F|m}(x \to x'). \tag{3.2.14}$$

Distribution	Name	Definition
$\rho_0(x)$	Initial Distribution	3.2.1
$\pi_F(x \to x')$	Forward Transition Function	3.2.2
$\rho_\tau(x')$	Final Distribution	3.2.4
$P_F(x, x')$	Joint Forward Distribution	3.2.5
$P(m\|x)$	Measurement Function	3.2.9
$\rho_{\tau\|m}(x')$	Updated Final Distribution	3.2.10
$\rho_{0\|m}(x)$	Updated Initial Distribution	3.2.11
$\pi_{F\|m}(x \to x')$	Updated Forward Transition Function	3.2.12
$P_{F\|m}(x, x')$	Updated Joint Forward Distribution	3.2.13
$\pi_R(\overline{x'} \to \overline{x})$	Reverse Transition Function	3.2.17
$P_R(x, x')$	Joint Reverse Distribution	3.2.18
$P_{R\|m}(x, x')$	Updated Joint Reverse Distribution	3.2.18
$\tilde{\rho}(x)$	Cycled Distribution	3.3.10
$\tilde{\rho}_m(x)$	Updated Cycled Distribution	3.4.17

Table 3.1. List of named probability distributions and their defining equations. These are grouped according to whether they are updated and/or time-reversed.

It may seem odd to update the transition functions based on measurements, since in principle the original transition functions were completely determined by the stochastic dynamics of the system and this is a desirable property that one would like to preserve. For this reason, the unupdated transition functions will play a special role below, while the updated ones are only used as an intermediate quantity in algebraic manipulations.

To illustrate these definitions, consider a simple toy model: a collection of N independent classical spins, each of which has a fixed probability to flip its state at each timestep. In this model it is most intuitive to work with a distribution function defined on macrostates (total number of up spins) rather than on microstates (ordered sequences of up/down spins).

The distribution functions relevant to our analysis are illustrated for this toy model with $N = 100$ spins in Fig. 3.2. To make the effects of evolution and updating most clear, we start with a bimodal initial distribution $\rho_0(x)$, uniform on the intervals $0 \leq x < 10$ and $90 < x \leq 100$. The system is evolved for a short time τ, not long enough to attain the equilibrium distribution, which would be a binomial centered at $x = N/2 = 50$. The final distribution $\rho_\tau(x')$ therefore has two small peaks just above and below $x' = 50$. We then perform a measurement, which simply asks whether most of the spins are up or down, obtaining the answer "mostly down." This corresponds to

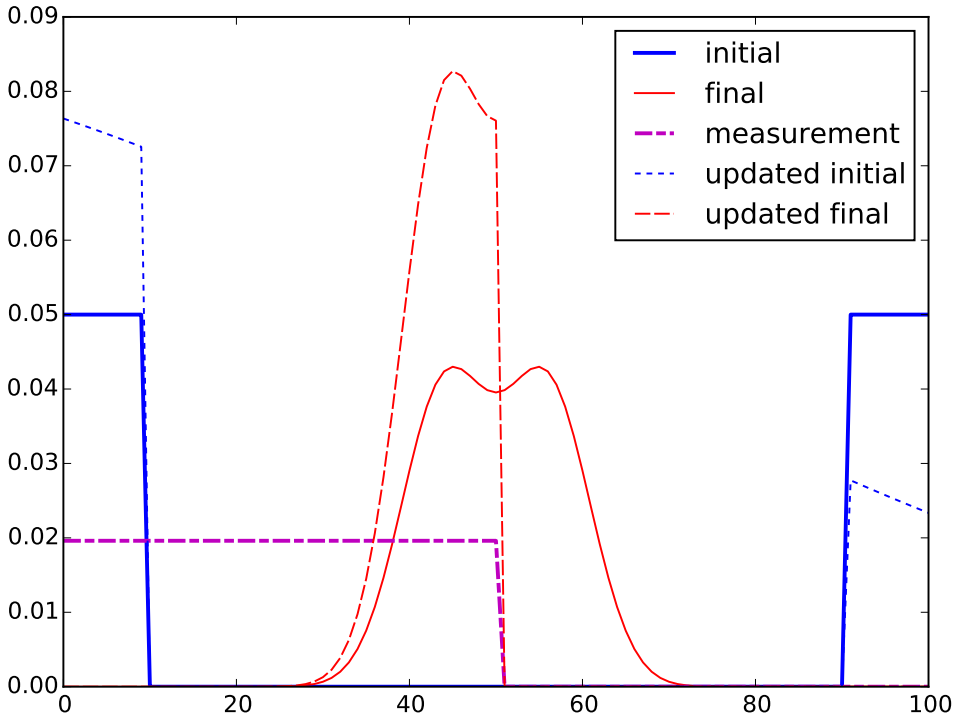


Figure 3.2. The various distribution functions illustrated within a toy model of 100 independent spins with a fixed chance of flipping at every timestep. The distributions are normalized functions on the space of the total number x of up-spins. We consider an initial distribution (thick solid blue line) that is equally split between the intervals $x < 10$ and $90 < x$. The system is evolved for enough time to come close to equilibrium but not quite reach it, as shown by the final distribution (thin solid red line). A measurement is performed, revealing that less than half of the spins are up (dot-dashed purple line). We can therefore update the post-measurement final distribution (dashed red line). The corresponding updated initial distribution (dotted blue line) is similar to the original initial distribution, but with a boost at low x and a decrease at high x.

a measurement function

$$P(m|x) = \begin{cases} 1 \text{ if } x \leq 50, \\ 0 \text{ if } x > 50. \end{cases} \tag{3.2.15}$$

In Fig. 3.2 we have plotted the normalized version $P(m|x)/P(m)$. From this we can construct the updated final and updated initial distributions, using (3.2.10) and (3.2.11).

The updated final distribution is just the left half of the non-updated final distribution, suitably renormalized. The updated initial distribution is a re-weighted version of the non-updated initial distribution, indicating that there is a greater probability for the system to have started with very few up spins (which makes sense, since our final measurement found that the spins were mostly down). This toy model does not have especially intricate dynamics, but it suffices to show how our evolution-and-updating procedure works.

3.2.3 The Reverse Protocol and Time Reversal

The Second Law contains information about the irreversibility of the time-evolution of the system, so to derive it we need to specify procedures to time-reverse both states and dynamics. Specifically, we will define an effectively "time-reversed" experiment that we can perform whose results can be compared to the time-forward experiment. As discussed in the Introduction, the point here is not to literally reverse the flow of time upon completion of the time-forward experiment (which would just undo the experiment), but to isolate the effects of dissipative processes, like friction, which result from complicated interactions with the environment.

For a state x, we denote by $\overline{x}$ the time-reversed state. In a ballistic model of particles, $\overline{x}$ is just the same as x with all of the particle velocities reversed. We are only talking about the velocities of the particles that make up the system, not the environment. In practice, an experimenter is not able to control the individual velocities of all of the particles in the system, so it may seem pointless to talk about reversing them. It will often be possible, however, to set up a time-reversed probability distribution $\overline{\rho}(x) \equiv \rho(\overline{x})$ given some procedure for setting up $\rho(x)$. For instance, if the system has a Maxwellian distribution of velocities with zero center-of-mass motion, then the probability distribution on phase space is actually time-reversal invariant.

Time reversal of dynamics is simpler, primarily because we have only limited experimental control over them. The system will have its own internal dynamics, it will interact with the environment, and it will be influenced by the experimenter. In a real experiment, it is only the influence of the experimenter that we are able to control, so our notion of time reversal for the dynamics is phrased purely in terms of the way the experimenter decides to influence the system. The experimenter influences the system in a (potentially) time-dependent way by following an experimental protocol, $\lambda(t)$, which we have called the "forward protocol." The forward protocol is a sequence of instructions to carry out using some given apparatus while the experiment is happening. We therefore define a "reverse protocol," which simply calls for the experimenter to execute the instructions backward. In practice, that involves time-reversing the control parameters (*e.g.*, reversing macroscopic momenta and magnetic fields) and running

them backwards in time, sending

$$\lambda_i(t) \to \bar{\lambda}_i(\tau - t). \tag{3.2.16}$$

For simplicity we will generally assume that the control parameters are individually invariant under time-reversal, so we won't distinguish between λ and $\bar{\lambda}$. The non-trivial aspect of the reverse protocol is then simply exchanging t with $\tau - t$. If the control parameters are time-independent for the forward protocol, then there will be no difference between the forward and reverse protocols. This kind of experiment involves setting up the initial state of the system and then just waiting for a time τ before making measurements.

Recall that the transition functions π_F for the system were defined assuming the experimenter was following the forward protocol. The reverse protocol is associated with a set of reverse transition functions π_R. We define π_R in analogy with (3.2.2) as

$$\pi_R(\overline{x'} \to \overline{x}) \equiv P(X_\tau = \overline{x} | X_0 = \overline{x'}; \overline{\lambda}\,(\tau - t)), \tag{3.2.17}$$

normalized as usual so that $\int \pi_R(\overline{x'} \to \overline{x})\, d\overline{x} = 1$.

We will also need a time-reversed version of the joint distribution P_F. As before, let $\rho_0(x)$ denote the initial distribution, and let $\rho_{\tau|m}(x)$ and $\rho_\tau(x)$ denote the distributions at time τ after following the forward protocol with and without Bayesian updates due to measurement, respectively. Then, following (3.2.5) and (3.2.13), define

$$\begin{aligned} P_R(x, x') &\equiv \rho_\tau(x')\pi_R(\overline{x'} \to \overline{x}), \\ P_{R|m}(x, x') &\equiv \rho_{\tau|m}(x')\pi_R(\overline{x'} \to \overline{x}). \end{aligned} \tag{3.2.18}$$

Although the reverse transition functions π_R are written as functions of the time-reversed states $\overline{x}$ and $\overline{x'}$, it is straightforward to apply the time-reversal map on these states to obtain the left-hand side purely as a function of x and x'.

It is helpful to think of these reverse joint probabilities in terms of a brand new experiment that starts fresh and runs for time τ. The initial distribution for this experiment is given by the final distribution coming from the forward experiment (with or without updates), and the experiment consists of time-reversing the state, executing the reverse protocol, and then time-reversing the state once more.

Our formalism should be contrasted with the typical formulation of a reverse experiment found in the literature. The initial distribution for the reverse experiment is frequently taken to be the equilibrium distribution for the final choice of control parameters [42]. The present method is more similar to the formalism of Seifert [150] in which an arbitrary final distribution, $p_1(x_t)$, is considered.

Note that in the definition of $P_{R|m}$, unlike in (3.2.13) above, the conditioning on m does not affect the transition function π_R. This is because, from the point of view of the reverse experiment, the measurement happens at the beginning. But π_R is a conditional probability which assumes a particular initial state (in this case $\overline{x'}$), and so the measurement m does not provide any additional information that can possibly affect the transition function. Also note the ordering of the arguments as compared with P_F in (3.2.5): the initial state for the reversed experiment is the second argument for P_R, while the initial state for the forward experiment is the first argument in P_F. Finally, we record the useful identity

$$\frac{P_{F|m}(x, x')}{P_{R|m}(x, x')} = \frac{P_F(x, x')}{P_R(x, x')}, \tag{3.2.19}$$

assuming both sides are well-defined for the chosen states x and x'.

3.2.4 Heat Flow

The Crooks Fluctuation Theorem [41] relates forward and reverse transition functions between equilibrium states to entropy production. It can be thought of as arising via coarse-graining from the "detailed fluctuation theorem," which relates the probabilities of individual forward and backward trajectories to the heat generated along the path through phase space [42, 137]. Outside the context of equilibrium thermodynamics, this relationship can be thought of as the definition of the "heat flow":

$$Q[x(t)] \equiv \log \frac{\pi_F[x(t)]}{\pi_R[\overline{x}(\tau - t)]}. \tag{3.2.20}$$

The quantity $Q[x(t)]$ can be equated with the thermodynamic heat (flowing out of the system, in this case) in situations where the latter concept makes sense. (More properly, it is the heat flow in units of the inverse temperature of the heat bath, since $Q[x(t)]$ is dimensionless.) However, $Q[x(t)]$ is a more general quantity than the thermodynamic heat; it is well-defined whenever the transition functions exist, including situations far from equilibrium or without any fixed-temperature heat bath.

In a similar manner, we can use the coarse-grained transition functions (depending on endpoints rather than the entire path) to define the following useful quantity,

$$\mathcal{Q}(x \to x') \equiv \log \frac{\pi_F(x \to x')}{\pi_R(\overline{x'} \to \overline{x})}. \tag{3.2.21}$$

This quantity $\mathcal{Q}$, the "generalized heat flow," is intuitively a coarse-grained version of the change in entropy of the environment during the transition $x \to x'$ in the forward experiment, though it is well-defined whenever the appropriate transition functions

exist. Similar concepts of coarse-grained entropy production have been explored previously in the literature [153, 154]. It is this generalized heat flow that will appear in our versions of the Second Law and the Bayesian Second Law.

3.3 Second Laws from Relative Entropy

All of the information about forward and reversed transition probabilities of the system is contained in the joint forward probability distribution $P_F(x, x')$ and reverse distribution $P_R(x, x')$, defined in (3.2.5) and (3.2.18), respectively. The effects of a Bayesian update on a measurement outcome m are accounted for in the distributions $P_{F|m}(x, x')$ and $P_{R|m}(x, x')$, given in (3.2.13) and (3.2.18). The most concise statements of the Second Law therefore arise from comparing these distributions.

3.3.1 The Ordinary Second Law from Positivity of Relative Entropy

The relative entropy, also known as the Kullback-Leibler divergence [149], is a measure of the distinguishability of two probability distributions:

$$D(p\|q) \equiv \int dx\, p(x) \log \frac{p(x)}{q(x)} \geq 0. \tag{3.3.1}$$

In a rough sense, $D(p\|q)$ can be thought of as the amount of information lost by replacing a true distribution p by an assumed distribution q. Relative entropy is nonnegative as a consequence of the concavity of the logarithm, and only vanishes when its two arguments are identical. In this sense it is like a distance, but with the key property that it is asymmetric in p and q, as both the definition and the intuitive description should make clear.

The relative entropy has been used in previous literature to quantify the information loss due to the stochastic evolution of a system. This has been achieved by analyzing path-space or phase-space distributions at a fixed time [42, 155, 156]. In a similar manner, we compute the relative entropy of the forward probability distribution with respect to the reverse one. However, we think of $P_F(x, x')$ and $P_R(x, x')$ each as single distributions on the space $\Gamma \times \Gamma$, so that

$$D(P_F\|P_R) = \int dx dx'\, P_F(x, x') \log \frac{P_F(x, x')}{P_R(x, x')}. \tag{3.3.2}$$

Into this we can plug the expressions (3.2.5) and (3.2.18) for P_F and P_R, as well as the relations (3.2.6) between those distributions and the single-time distributions $\rho_0(x)$

and $\rho_\tau(x')$, to obtain

$$D(P_F\|P_R) = \int \rho_0(x)\pi_F(x\to x')\left(\log\frac{\rho_0(x)}{\rho_\tau(x')} + \log\frac{\pi_F(x\to x')}{\pi_R(\overline{x'}\to\overline{x})}\right)dxdx' \tag{3.3.3}$$

$$= S(\rho_\tau) - S(\rho_0) + \int dxdx'\ P_F(x,x')\mathcal{Q}(x\to x'). \tag{3.3.4}$$

Here S is the usual Gibbs or Shannon entropy,

$$S(\rho) \equiv -\int \rho(x)\log\rho(x)\,dx, \tag{3.3.5}$$

and $\mathcal{Q}$ is the generalized heat flow defined by (3.2.21) above. The first two terms in (3.3.4) constitute the change in entropy of the system, while the third term represents an entropy change in the environment averaged over initial and final states. This expansion of (3.3.4) is essentially a restatement of the Kawai-Parrondo-van den Broeck equality [155]. We will introduce the notation $\langle\cdot\rangle_F$ to denote the average of a quantity with respect to the probability distribution P_F,

$$\langle f\rangle_F \equiv \int dxdx'\ P_F(x,x')f(x,x'). \tag{3.3.6}$$

The positivity of the relative entropy (3.3.2) is therefore equivalent to

$$\Delta S + \langle\mathcal{Q}\rangle_F \geq 0, \tag{3.3.7}$$

with equality if and only if $P_F = P_R$. This is the simplest form of the Second Law; it states that the change in entropy of the system is bounded from below by (minus) the average of the generalized heat $\mathcal{Q}$ with respect to the forward probability distribution.

The result (3.3.7) is an information-theoretical statement; in the general case we should not think of S as a thermodynamic entropy, or $\langle\mathcal{Q}\rangle_F$ as the expectation value of a quantity which can be measured in experiments. To recover the thermodynamic Second Law, we must restrict ourselves to setups in which temperature, heat flow, and thermodynamic entropy are all well-defined. In this case, we can interpret $\langle\mathcal{Q}\rangle_F$ as the expected amount of coarse-grained heat flow into the environment. "Coarse-grained" here refers to the difference between the endpoint-dependent $\mathcal{Q}(x\to x')$ and the fully path-dependent $Q[x(t)]$ introduced above. By considering the relative entropy of the forward path-space probability $P_F[x(t)]$ with respect to the reverse one $P_R[x(t)]$, we can recover the ordinary Second Law with the ordinary heat term, obtained from (3.3.7) by the replacement $\mathcal{Q}\to Q$. We will have more to say about the relationship between these two forms of the ordinary Second Law in the following section.

3.3.2 A Refined Second Law from Monotonicity of Relative Entropy

Given any pair of probability distributions $p(x, y)$, $q(x, y)$ on multiple variables, we have

$$D(p(x,y)\|q(x,y)) \geq D\left(\int dy\, p(x,y)\middle\|\int dy\, q(x,y)\right). \tag{3.3.8}$$

This property is known as the monotonicity of relative entropy. To build intuition, it is useful to first consider a more general property of the relative entropy:

$$D(p\|q) \geq D\left(Wp\|Wq\right) \quad \forall W, \tag{3.3.9}$$

where W is a probability-conserving (i.e., stochastic) operator. This result follows straightforwardly from the definition of relative entropy and the convexity of the logarithm. In words, it means that performing any probability-conserving operation W on probability distributions p and q can only reduce their relative entropy.

In information theory, (3.3.9) is known as the Data Processing Lemma [157–159], since it states that processing a signal only decreases its information content. Marginalizing over a variable is one such way of processing (it is probability-conserving by the definition of p and q), so marginalization, in particular, cannot increase the relative information. Intuitively, (3.3.8) says that marginalizing over one variable decreases the amount of information lost when one approximates p with q.

Our single-time probability distributions $\rho_t(x)$ can be thought of as marginalized versions of the joint distribution $P_F(x, x')$, following (3.2.6). We can also define a new "cycled" distribution by marginalizing $P_R(x, x')$ over x' to obtain

$$\tilde{\rho}(x) \equiv \int dx'\ P_R(x,x') = \int dx' \rho_\tau(x') \pi_R(\overline{x'} \to \overline{x}). \tag{3.3.10}$$

This is the probability distribution we find at the conclusion of the reversed experiment, or, in other words, after running through a complete cycle of evolving forward, time-reversing the state, evolving with the reverse protocol, and then time-reversing once more. In the absence of environmental interaction, we expect the cycled distribution to match up with the initial distribution $\rho_0(x)$, since the evolution of an isolated system is completely deterministic.

Applying monotonicity to P_F and P_R by marginalizing over the final state x', we have

$$D(P_F\|P_R) \geq D(\rho_0\|\tilde{\rho}) \geq 0, \tag{3.3.11}$$

or simply, using the results of the previous subsection,

$$\Delta S + \langle \mathcal{Q} \rangle_F \geq D(\rho_0\|\tilde{\rho}) \geq 0. \tag{3.3.12}$$

This is a stronger form of the ordinary Second Law. It states that the change in entropy is bounded from below by an information-theoretic quantity that characterizes the difference between the initial distribution ρ_0 and a cycled distribution $\tilde{\rho}$ that has been evolved forward and backward in time.

In the context of a numerical simulation, it is easier to calculate $D(\rho_0\|\tilde{\rho})$ than $D(P_F\|P_R)$, since the former only depends on knowing the probability distribution of the system at two specified points in time. $D(\rho_0\|\tilde{\rho})$ can readily be calculated by evolving the distribution according to the forward and reverse protocols. This is in contrast with $D(P_F\|P_R)$, the computation of which requires knowledge of joint probability distributions. Obtaining the joint distributions is more difficult, because one must know how each microstate at the given initial time relates to the microstates of the future time. This bound therefore provides an easily-calculable contraint on the full behavior of the system. We note that similar approaches involving coarse-graining have been previously explored in the literature [155].

Monotonicity of the relative entropy also allows us to succinctly state the relationship between the path-space and endpoint-space formulations of the Second Law. Indeed, the relationship between the probabilities $P_F[x(t)]$ and $P_F(x, x')$ is

$$P_F(x, x') = \int_{x(0)=x}^{x(\tau)=x'} \mathcal{D}x(t)\, P_F[x(t)], \tag{3.3.13}$$

with a similar relationship between the reversed quantities. Monotonicity of relative entropy then implies that

$$D(P_F[x(t)]\|P_R[x(t)]) \geq D(P_F(x, x')\|P_R(x, x')). \tag{3.3.14}$$

Since the changes in entropy are the same, this inequality reduces to the relationship $\langle Q[x(t)]\rangle_F \geq \langle \mathcal{Q}(x \to x')\rangle_F$ between the expected heat transfer and the expected coarse-grained heat transfer, which can also be shown directly with a convexity argument. The point here is that the path-space and endpoint-space formulations of the ordinary Second Law (as well as the Bayesian Second Law in the following section) are not independent of each other. Endpoint-space is simply a coarse-grained version of path-space, and the monotonicity of relative entropy tells us how the Second Law behaves with respect to coarse-graining.

3.4 The Bayesian Second Law

Now we are ready to include Bayesian updates. It is an obvious extension of the discussion above to consider the relative entropy of the updated joint probabilities

$P_{F|m}$ and $P_{R|m}$, which is again non-negative:

$$D(P_{F|m}\|P_{R|m}) \geq 0. \tag{3.4.1}$$

This is the most compact form of the Bayesian Second Law (BSL).

3.4.1 Cross-Entropy Formulation of the BSL

It will be convenient to expand the definition of relative entropy in several different ways. First, we can unpack the relative entropy to facilitate comparison with the ordinary Second Law:

$$D(P_{F|m}\|P_{R|m}) = \int dx\, \rho_{0|m}(x) \log \rho_0(x) - \int dx'\, \rho_{\tau|m}(x') \log \rho_\tau(x') + \langle \mathcal{Q} \rangle_{F|m}. \tag{3.4.2}$$

Here we have used the expressions (3.2.13) and (3.2.18) for the joint distributions, as well as the identity (3.2.19). We have also extracted the generalized heat term,

$$\langle \mathcal{Q} \rangle_{F|m} \equiv \int dx dx'\; P_{F|m}(x, x') \log \frac{\pi_F(x \to x')}{\pi_R(\overline{x'} \to \overline{x})}, \tag{3.4.3}$$

which is the expected transfer of generalized heat out of the system during the forward experiment given the final measurement outcome. This is an experimentally measurable quantity in thermodynamic setups: the heat transfer is measured during each trial of the experiment, and $\langle \mathcal{Q} \rangle_{F|m}$ is the average over the subset of trials for which the measurement outcome was m. The remaining two terms are not identifiable with a change in entropy, but we have a couple of options for interpreting them.

The form of (3.4.2) naturally suggests use of the cross entropy between two distributions, defined as

$$H(p, q) = -\int dx\, p(x) \log q(x). \tag{3.4.4}$$

(Note that this is not the joint entropy, defined for a joint probability distribution $p(x, y)$ as $-\int dxdy\, p(x, y) \log p(x, y)$.) Using this definition, the relative entropy between the updated joint distributions (3.4.2) may be rewritten in the form,

$$D(P_{F|m}\|P_{R|m}) = H(\rho_{\tau|m}, \rho_\tau) - H(\rho_{0|m}, \rho_0) + \langle \mathcal{Q} \rangle_{F|m}. \tag{3.4.5}$$

The Bayesian Second Law is then

$$\Delta H(\rho_m, \rho) + \langle \mathcal{Q} \rangle_{F|m} \geq 0. \tag{3.4.6}$$

Here, Δ is the difference in the values of a quantity evaluated at the final time τ and the initial time 0.

To get some intuition for interpreting this form of the BSL, it is useful to recall the information-theoretic meaning of the entropy and cross entropy. Given a probability distribution $p(x)$ over the set of microstates x in a phase space Γ, we can define the self-information associated with each state,

$$I_p(x) = \log \frac{1}{p(x)}. \tag{3.4.7}$$

This quantity is also referred to as the Shannon information, "surprisal", or in the context of stochastic thermodynamics, the stochastic Shannon entropy [141, 160]. The self-information measures the information we would gain by learning the identity of the specific microstate x. If x is highly probable, it's not that surprising to find the system in that state, and we don't learn that much by identifying it; if it's improbable we have learned a great deal. From this perspective, the entropy $S(p) = \int dx\, p(x) I_p(x)$ is the expectation value, with respect to $p(x)$, of the self-information associated with $p(x)$ itself. It is how much we are likely to learn, on average, by finding out the actual microstate of the system. In a distribution that is highly peaked in some region, the microstate is most likely to be in that region, and we don't learn much by finding it out; such a distribution has a correspondingly low entropy. In a more uniform distribution, we always learn something by finding out the specific microstate, and the distribution has a correspondingly higher entropy.

In contrast, the cross entropy $H(p, q) = \int dx\, p(x) I_q(x)$ is the expectation value with respect to $p(x)$ of the self-information associated with $q(x)$. Typically $p(x)$ is thought of as the "true" or "correct" distribution, and $q(x)$ as the "assumed" or "wrong" distribution. We believe that the probability distribution is given by $q(x)$, when it is actually given by $p(x)$. The cross entropy is therefore a measure of how likely we are to be surprised (and therefore learn something) if we were to be told the actual microstate of the system, given that we might not be using the correct probability distribution. The cross entropy is large when the two distributions are peaked, but in different places; that maximizes the chance of having a large actual probability $p(x)$ for a state with a large self-information $I_q(x)$. When the two distributions differ, we are faced with two distinct sources of uncertainty about the true state of the system: the fact that there can be uncertainty in the true distribution, and the fact that we are working with an assumed distribution rather than the true one. Mathematically, this is reflected in the cross entropy being equal to the entropy of the true distribution plus the relative entropy:

$$H(p, q) = S(p) + D(p\|q). \tag{3.4.8}$$

The cross entropy is always greater than the entropy of the true distribution (by positivity of relative entropy), and reduces to the ordinary entropy when the two distributions

are the same.

The Bayesian Second Law, then, is the statement that the cross entropy of the updated ("true") distribution with respect to the original ("wrong") distribution, plus the generalized heat flow, is larger when evaluated at the end of the experiment than at the beginning. In other words, for zero heat transfer, the expected amount of information an observer using the original distribution function would learn by being told the true microstate of the system, conditioned on an observation at the final time, is larger at the final time than at the initial one.

We note that the quantity $H(\rho_{t|m}, \rho_t)$ only has operational meaning once a measurement has occurred, since performing the Bayesian update to take the measurement into account requires knowledge of the actual measurement outcome. The BSL is a statement about how much an experimenter who knows the measurement outcome would expect someone who didn't know the outcome to learn by being told the microstate of the system. Therefore, there is not any sense in which one can interpret an increase of $H(\rho_{t|m}, \rho_t)$ with increasing t as an increase in a dynamical quantity. This is in contrast with the dynamical interpretation of the monotonic increase in entropy over time in the ordinary Second Law. It is, in fact, the case that $H(\rho_{t|m}, \rho_t)$ *does* increase with increasing t for zero heat transfer, but this increase can only be calculated retroactively once the measurement has actually been made. Of course, in the case of a trivial measurement that tells us nothing about the system, the BSL manifestly reduces to the ordinary Second Law, since $H(\rho, \rho) = S(\rho)$.

3.4.2 Alternate Formulations of the BSL

Another natural quantity to extract is the total change in entropy after the two-step process of time evolution and Bayesian updating, which we will call ΔS_m:

$$\Delta S_m \equiv S(\rho_{\tau|m}) - S(\rho_0). \tag{3.4.9}$$

This is the actual change in the entropy over the course of the experiment in the mind of the experimenter, who initially believes the distribution is ρ_0 (before the experiment begins), and ultimately believes it to be $\rho_{\tau|m}$. In terms of this change in entropy, we have

$$D(P_{F|m} \| P_{R|m}) = \Delta S_m + \langle \mathcal{Q} \rangle_{F|m} + D(\rho_{\tau|m} \| \rho_\tau) + \int dx \; (\rho_{0|m}(x) - \rho_0(x)) \log \rho_0(x). \tag{3.4.10}$$

The second to last term, $D(\rho_{\tau|m} \| \rho_\tau)$, is the relative entropy of the posterior distribution at time τ with respect to the prior distribution; it can be thought of as the amount of information one gains about the final probability distribution due to the measurement outcome. This is a natural quantity in Bayesian analysis, called simply the *information*

gain [161]; maximizing its expected value (and hence the expected information learned from a measurement) is the goal of Bayesian experimental design [162]. Because it measures information gained, it tends to be largest when the measurement outcome m was an unlikely one from the point of view of ρ_τ. The final term exactly vanishes in the special case where the initial probability distribution is constant on its domain, which is an important special case we will consider in more detail below.

Using (3.4.10), the positivity of relative entropy is equivalent to

$$\Delta S_m + \langle \mathcal{Q} \rangle_{F|m} \geq -D(\rho_{\tau|m} \| \rho_\tau) + \int dx \; (\rho_0(x) - \rho_{0|m}(x)) \log \rho_0(x). \tag{3.4.11}$$

The left-hand side of this inequality is similar to that of the ordinary Second Law, except that the result of the measurement is accounted for. In the event of an unlikely measurement we would intuitively expect that it should be allowed to be negative. Accordingly, on the right-hand side we find that it is bounded from below by a quantity that can take on negative values. Indeed, the more unlikely the measurement is, the greater $D(\rho_{\tau|m} \| \rho_\tau)$ is, and the more the entropy is allowed to decrease.

Finally, we can expand the relative entropy in terms of $S(\rho_{0|m})$ instead of $S(\rho_0)$. That is, we define the change in entropy between the initial and final updated distributions,

$$\Delta S(\rho_m) \equiv S(\rho_{\tau|m}) - S(\rho_{0|m}). \tag{3.4.12}$$

(Note the distinction between $\Delta S(\rho_m)$ here and ΔS_m in (3.4.9)). This is the answer to the question: "Given the final measurement, how much has the entropy of the system changed?" Then (3.4.11) is equivalent to

$$\Delta S(\rho_m) + \langle \mathcal{Q} \rangle_{F|m} \geq D(\rho_{0|m} \| \rho_0) - D(\rho_{\tau|m} \| \rho_\tau). \tag{3.4.13}$$

This change of entropy can be contrasted with $S(\rho_{\tau|m}) - S(\rho_0)$, which is a statement about the change in the experimenter's knowledge of the system before and after the measurement is performed.

The right hand side of (3.4.13) has the interesting property that it is always less than or equal to zero. This can be shown by taking the difference of the relative entropies and expressing it in the form

$$D(\rho_{0|m} \| \rho_0) - D(\rho_{\tau|m} \| \rho_\tau) = \int dx dx' \frac{\rho_0(x) \pi_F(x \to x') P(m|x')}{P(m)} \log \frac{\pi_F(x \to m)}{P(m|x')}. \tag{3.4.14}$$

We have defined $\pi_F(x \to m) \equiv \int dx' \pi_F(x \to x') P(m|x')$ for convenience. It is only possible to write the difference in this form because the initial and final distributions are related by evolution (3.2.14). Using the concavity of the logarithm, it can then be shown that this quantity is non-positive.

One final point of interest in regards to (3.4.11) and (3.4.13) is their average with respect to measurement outcomes. The inequality is predicated on a specific measurement outcome, m; averaging with respect to the probability of obtaining a given measurement, we find

$$\langle \Delta S_m \rangle + \langle \mathcal{Q} \rangle_F \geq -I(X_\tau; M), \tag{3.4.15}$$

$$\langle \Delta S(\rho_m) \rangle + \langle \mathcal{Q} \rangle_F \geq I(X_0; M) - I(X_\tau; M), \tag{3.4.16}$$

where $I(X_t; M)$ is the mutual information between the microstate of the system at time t and the measurement outcome observed at time τ. Here the mutual information can be expressed as the relative entropy of a joint probability distribution to the product of its marginal distributions, $I(X; M) = D(\rho(x, m) \| \rho(x)\rho(m))$.

Inequalities similar to (3.4.15) and (3.4.16) can be found in the existing literature for nonequilibrium feedback-control [139, 160, 163–166]. A careful accounting of mutual informations and entropy production in feedback controlled systems has been performed in [160]. While we have not explicitly included feedback-control in our formalism, we see that (3.4.15) corresponds to the bounds on the entropy production of a system under measurement and feedback-control when the initial state of the system is uncorrelated with that of the measuring device. This equivalence is unsurprising as this is essentially the setup we are considering presently. One may also show that both (3.4.15) and (3.4.16) are equivalent to (3.3.7) using standard identities involving the mutual information and conditional entropy.

Unlike the averaged inequalities (3.4.16), (3.4.15), and other similar inequalities in the literature [139, 141, 147, 160], the forms (3.4.11) and (3.4.13) of the Bayesian Second Law hold independently for each possible measurement outcome. These enable us to make contact with the change in thermodynamic properties of the system for specific measurement outcomes. In particular, (3.4.11) and (3.4.13) allow us to study a system over only those measurement outcomes which appear to naively violate the Second Law.

3.4.3 A Refined BSL from Monotonicity of Relative Entropy

So far we have rewritten the relative entropy of the forward and reverse distributions (3.4.1) in various ways, but there is a refined version of the BSL that we can formulate using monotonicity of relative entropy, analogous to the refined version of the ordinary Second Law we derived in Section 3.3.2. Following the definition of the cycled distribution $\tilde{\rho}$ in (3.3.10), we can define an updated cycled distribution by marginalizing the updated reverse distribution over initial states,

$$\tilde{\rho}_m(x) \equiv \int dx' \; P_{R|m}(x, x') = \int dx' \rho_{\tau|m}(x') \pi_R(\overline{x'} \to \overline{x}). \tag{3.4.17}$$

The monotonicity of relative entropy then implies that

$$D(P_{F|m}\|P_{R|m}) \geq D(\rho_{0|m}\|\tilde{\rho}_m). \tag{3.4.18}$$

This is the refined Bayesian Second Law of Thermodynamics in its most compact form, analogous to the refined Second Law (3.3.12).

Expanding the definitions as above, the refined BSL can be written as

$$\Delta H(\rho_m, \rho) + \langle \mathcal{Q} \rangle_{F|m} \geq D(\rho_{0|m}\|\tilde{\rho}_m), \tag{3.4.19}$$

or equivalently as

$$\Delta S_m + \langle \mathcal{Q} \rangle_{F|m} \geq D(\rho_{0|m}\|\tilde{\rho}_m) - D(\rho_{\tau|m}\|\rho_\tau) + \int dx\ (\rho_0(x) - \rho_{0|m}(x)) \log \rho_0(x). \tag{3.4.20}$$

From the form of (3.4.19), we see that the change in the cross entropy obeys a tighter bound than simple positivity, as long as the cycled distribution deviates from the original distribution (which it will if the evolution is irreversible).

In a similar manner to (3.4.11), we can average (3.4.20) over all possible measurement outcomes to arrive at a stronger form of (3.4.15) which accounts for the average irreversibility of the experiment.

Other versions of the Second Law can be obtained from the relative entropy by inserting different combinations of $P_{F|m}$, $P_{R|m}$, P_F, and P_R. We have chosen to highlight $D(P_F\|P_R)$ and $D(P_{F|m}\|P_{R|m})$ because these are the combinations which we can expect to vanish in the case of perfect reversibility, and thus characterize the time-asymmetry of the dynamics. Other possibilities, like $D(P_{F|m}\|P_R)$, are always nonzero as long as information is gained from the measurement.

3.5 Bayesian Integral Fluctuation Theorems

The inequalities derived in Section 3.4 allow us to make statements about the average values of various thermodynamic variables during the non-equilibrium evolution of the system. However, such expressions are fundamentally limited in the amount of information they provide for small systems where fluctuations are significant. To describe the role of fluctuations, a wide literature of fluctuation theorems has been developed [39, 41, 137, 138, 141]. In this section, we will derive Bayesian analogs of such fluctuation theorems.

Recall the simple identity (3.2.19):

$$\frac{P_R(x,x')}{P_F(x,x')} = \frac{P_{R|m}(x,x')}{P_{F|m}(x,x')} = \frac{\rho_\tau(x')}{\rho_0(x)} e^{-\mathcal{Q}(x\to x')}, \tag{3.5.1}$$

which we have made use of in previous sections. We may rewrite (3.5.1) in the following form,

$$\frac{P_R(x,x')}{P_F(x,x')} = e^{\log \rho_\tau(x') - \log \rho_0(x) - \mathcal{Q}(x \to x')}, \tag{3.5.2}$$

We note that the first pair of terms in the exponential consitute the negative of the change in stochastic entropy during the evolution of the system along a forward trajectory [141]. As such, we see that (3.5.2) is the Crooks fluctuation theorem for entropy production [41]. However, making use of the relation in (3.5.1), we also find

$$\frac{P_{R|m}(x,x')}{P_{F|m}(x,x')} = e^{\log \rho_\tau(x') - \log \rho_0(x) - \mathcal{Q}(x \to x')}. \tag{3.5.3}$$

This demonstrates the rather surprising fact that the same fluctuation theorem which holds for the unupdated joint distributions also holds for the updated joint distributions. While (3.5.3) appears to have the form of the Crooks fluctuation theorem, it is important to note that the stochastic entropy production in the exponential corresponds to unupdated trajectories. This is generally different from what would be considered the stochastic entropy production along the same trajectory after performing a Bayesian update.

We can obtain integral fluctuation theorems from (3.5.2) and (3.5.3) by computing the expectation value of the ratio with respect to P_F (or $P_{F|m}$). Naively, one would multiply by P_F and find $P_F P_R / P_F = P_R$, but we need to keep track of the domain of integration: we are only interested in points where $P_F \neq 0$ ($P_{F|m} \neq 0$) when computing an average with respect to P_F ($P_{F|m}$). So we have, for instance,

$$\left\langle \frac{P_R}{P_F} \right\rangle_F = \int_{P_F \neq 0} dx dx' \; P_R(x,x'). \tag{3.5.4}$$

This integral will be equal to one unless there is a set of zero P_F-measure with nonzero P_R-measure. On such a set, the ratio P_R/P_F diverges. Generically this will include all points where $\rho_0(x)$ vanishes, unless $\mathcal{Q}$ happens to diverge for some choices of x' (*e.g.*, one reason for P_R to vanish is that certain transitions are strictly irreversible). Note that if $\rho_0(x)$ is nowhere zero and $\mathcal{Q}$ does not ever diverge (as in most physically relevant situations), then this integral is equal to one. This is true no matter how small $\rho_0(x)$ is or how large $\mathcal{Q}$, as long as they are nonzero and finite everywhere, respectively. For this reason, (3.5.4) generically is equal to one.

The same reasoning holds for the updated probabilities:

$$\left\langle \frac{P_{R|m}}{P_{F|m}} \right\rangle_{F|m} = \int_{P_{F|m} \neq 0} dx dx' \; P_{R|m}(x,x'). \tag{3.5.5}$$

Since the ratio $P_{R|m}/P_{F|m}$ is identical to the ratio P_R/P_F, the condition for this integral to equal one is the same as the previous integral, which means it is generically so.

To summarize, we have constants a, b_m such that

$$\left\langle \frac{P_R}{P_F} \right\rangle_F = \left\langle e^{\log \rho_\tau(x') - \log \rho_0(x) - \beta \mathcal{Q}(x \to x')} \right\rangle_F = a \leq 1, \tag{3.5.6}$$

$$\left\langle \frac{P_{R|m}}{P_{F|m}} \right\rangle_{F|m} = \left\langle e^{\log \rho_\tau(x') - \log \rho_0(x) - \beta \mathcal{Q}(x \to x')} \right\rangle_{F|m} = b_m \leq 1. \tag{3.5.7}$$

By perturbing the initial state by an arbitrarily small amount, we can make P_R/P_F finite everywhere (excluding divergences in $\mathcal{Q}$), and so $a \neq 1$ and $b_m \neq 1$ are in some sense unstable. We can use Jensen's inequality on each of these to extract a Second Law:

$$D(P_F \| P_R) \geq -\log a \geq 0 \tag{3.5.8}$$

$$D(P_{F|m} \| P_{R|m}) \geq -\log b_m \geq 0. \tag{3.5.9}$$

Thus, these integral fluctuation theorems contain within them the results from the positivity of relative entropy, (3.3.7) and (3.4.1).

The relation (3.5.6) is related to the Jarzynski equality, or nonequilibrium work relation [39]

$$\left\langle e^{-W} \right\rangle = e^{-\Delta F}. \tag{3.5.10}$$

Here, W is the work performed between initial and final states and ΔF is the free-energy difference between the equilibrium states corresponding to the initial and final conditions. For this equality we do not require that the system actually attains an equilibrium distribution at the end of the experiment, even though the result is phrased in terms of the equilibrium free energy. Our (3.5.6), by contrast, refers to the actual initial and final distributions, whatever they may be, and does not invoke equilibrium quantities. But it is straightforward to show that (3.5.6) implies (3.5.10) if we assume microscopic reversibility of the underlying dynamics. Consider an initial distribution $\rho_0(x)$ that is in equilibrium. The distribution $\rho_\tau(x')$ will in general not be in equilibrium, but we can imagine extending our protocol from τ to τ', keeping fixed all of the control parameters, until an equilibrium distribution $\rho_{\tau'}(x'')$ is reached. The work done on the system can be defined in terms of the energy of a microstate $E(x)$ and the heat flow (3.2.20) through

$$E(x'') - E(x) = Q[x(t)] + W[x(t)], \tag{3.5.11}$$

while the free energy F can be defined for an equilibrium distribution through

$$\rho_{\text{eq}}(x) = \frac{1}{Z} e^{F - E(x)}. \tag{3.5.12}$$

For the extension period from τ to τ' (when the control parameters are fixed), microscopic reversibility implies that the heat flow satisfies

$$Q[x_{\tau\to\tau'}(t)] \equiv \log \frac{\pi_F[x_{\tau\to\tau'}(t)]}{\pi_R[\overline{x_{\tau\to\tau'}}(\tau'-t)]} = \log \frac{\rho_{\tau'}(x'')}{\rho_\tau(x')} = E(x'') - E(x'), \tag{3.5.13}$$

consistent with our intuition that no work is performed during this period. We can consider a version of (3.5.6) that is expressed in terms of specific paths (rather than simply endpoints) by replacing $\mathcal{Q} \to Q$ and including an average over these paths in the expectation value. Then applying this to the entire process from equilibrium at $t = 0$ to a different equilibrium at $t = \tau'$, we have

$$\left\langle e^{\log \rho_{\tau'}(x'') - \log \rho_0(x) - Q(x\to x'')} \right\rangle_F = \left\langle e^{\Delta F - \Delta E - Q(x\to x'')} \right\rangle_F = e^{\Delta F} \left\langle e^{-W} \right\rangle_F = a. \tag{3.5.14}$$

This recovers the Jarzynski equality (3.5.10), in the generic case when $a = 1$.

We see that (3.5.7) provides a generalized integral fluctuation theorem which includes Bayesian updating. Importantly, (3.5.7) holds independently for each possible measurement outcome. A similar type of integral fluctuation theorem has been proven in the context of two-time measurements of a quantum system without Bayesian updating [167–169]. Taken together, (3.5.6) and (3.5.7) demonstrate that for any experiment, two different classes of integral fluctuation theorems hold. If we partition a large set of experimental trials based on measurement outcomes, each subset obeys its own integral fluctuation theorem, (3.5.7). However, if we consider all experimental trials together, (3.5.6) holds. This leads us to the relation

$$\int dm\ P(m) b_m = a. \tag{3.5.15}$$

Finally, we note that there are also integral fluctuation theorems corresponding to the monotonicity inequalities. Consider

$$\left\langle \frac{P_R}{P_F} \frac{\rho_0}{\tilde{\rho}} \right\rangle_F = \int_{P_F \neq 0} dx dx'\ \frac{P_R(x,x')}{\int dy'\ P_R(x,y')} \rho_0(x) \leq 1. \tag{3.5.16}$$

Applying Jensen's inequality reproduces the monotonicity result:

$$D(P_F \| P_R) \geq D(\rho_0 \| \tilde{\rho}). \tag{3.5.17}$$

The refined Bayesian Second Law follows similarly from the integral fluctuation theorem,

$$\left\langle \frac{P_{R|m}}{P_{F|m}} \frac{\rho_{0|m}}{\tilde{\rho}_m} \right\rangle_{F|m} = c_m \leq 1. \tag{3.5.18}$$

This may be re-expressed as:

$$\left\langle e^{\log \rho_\tau(x') - \log \rho_0(x) - \beta \mathcal{Q}(x \to x') + \log[\rho_{0|m}(x)] - \log[\tilde{\rho}_m(x)]} \right\rangle_{F|m} = c_m. \tag{3.5.19}$$

We see that (3.5.19) extends (3.5.7) by also including the information loss from the stochastic evolution of the system. As was the case for (3.5.7), (3.5.19) holds independently for each possible measurement outcome.

3.6 Applications

As a way to build some intuition for the Bayesian point of view we have been discussing, we will go through a few simple examples and special cases.

3.6.1 Special Cases

Perfect Complete Measurement. If a measurement does not yield any new information, then the updated probabilities are identical to the prior probabilities and the Bayesian Second Law reduces to the ordinary Second Law. On the other hand, consider a measuring device that is able to tell us with certainty what the exact microstate of the system is at the time of measurement. The outcome m of the experiment is then a single point in phase space. If we employ such a device, we have the following simplified expressions:

$$\rho_{0|m}(x) = \frac{\rho_0(x)\pi_F(x \to m)}{\rho_\tau(m)}, \tag{3.6.1}$$

$$\rho_{\tau|m}(x') = \delta(x' - m), \tag{3.6.2}$$

$$\pi_{F|m}(x \to x') = \delta(x' - m)\theta(\pi_F(x \to m)), \tag{3.6.3}$$

$$\tilde{\rho}_m(x) = \pi_R(\overline{m} \to \overline{x}). \tag{3.6.4}$$

Using these simplifications, we find

$$D(P_{F|m} \| P_{R|m}) = D(\rho_{0|m} \| \tilde{\rho}_m), \tag{3.6.5}$$

so the refined Bayesian Second Law is always saturated. This is because marginalization of the joint distribution over the final endpoint results in no loss of information: we are still conditioning on the measurement outcome m, which tells us the final endpoint.

The Boltzmann Second Law of Thermodynamics. In the Boltzmann formulation of the Second Law, phase space is partitioned into a set of macrostates. Each microstate is assigned to a macrostate; the entropy of a microstate x is defined as the entropy of its associated macrostate $\Sigma(x)$, which is the logarithm of the macrostate's

phase space volume $|\Sigma|$. We can reproduce this formulation as a special case of the Bayesian measurement formalism: the measuring device determines which macrostate the microstate belongs to with absolute certainty. If the measurement outcome m indicates that the system is in some particular macrostate (but doesn't include any additional information), we have

$$P(m|x) = \mathbb{1}_m(x) \equiv \begin{cases} 1 \text{ if } x \in m, \\ 0 \text{ if } x \notin m. \end{cases} \tag{3.6.6}$$

We also choose our initial distribution to be uniform over an initial macrostate Σ_0:

$$\rho_0(x) = \frac{1}{|\Sigma_0|}\mathbb{1}_{\Sigma_0}(x). \tag{3.6.7}$$

Then we have the identities

$$\langle -\log \rho_0(x)\rangle_{F|m} = \log|\Sigma_0| = S(\rho_0), \tag{3.6.8}$$

$$\langle -\log \rho_\tau(x)\rangle_{F|m} = -\int dx\ \rho_{\tau|m}(x)\log\rho_{\tau|m}(x) + \int dx\ \rho_{\tau|m}(x)\log\frac{\rho_{\tau|m}(x)}{\rho_\tau(x)} \tag{3.6.9}$$

$$= S(\rho_{\tau|m}) + D(\rho_{\tau|m}\|\rho_\tau). \tag{3.6.10}$$

Then the refined Bayesian Second Law (3.4.20) simplifies to

$$\Delta S_m + \langle \mathcal{Q}\rangle_{F|m} \geq D(\rho_{0|m}\|\tilde{\rho}_m) - D(\rho_{\tau|m}\|\rho_\tau). \tag{3.6.11}$$

The left-hand side of this inequality is not quite the same as in the Boltzmann formulation, because $S(\rho_{\tau|m})$ is not the entropy associated with any of the previously established macrostates. But we do have the inequality $S(\rho_{\tau|m}) \leq \log|m|$, which *is* the entropy of the final macrostate. So the left-hand side of (3.6.11) can be replaced by the usual left-hand side of the Boltzmann Second Law while maintaining the inequality.[3]

The right-hand side of the Boltzmann Second Law is zero, while in (3.6.11) we have the difference of two positive terms. The Boltzmann Second Law can be violated by rare fluctuations, and here we are able to characterize such fluctuations by the fact that they render the right-hand side of our inequality negative. We can also give an explicit formula for the term $D(\rho_{\tau|m}\|\rho_\tau)$ that comes in with a minus sign:

$$D(\rho_{\tau|m}\|\rho_\tau) = -\log\int_m dx'\ \rho_\tau(x') = -\log P(m) = I_m, \tag{3.6.12}$$

[3]And, as we have discussed previously, the coarse-grained $\mathcal{Q}$ can be replaced by the path-space Q as well.

where I_m is the self-information associated with the measurement outcome m. When the observed measurement is very surprising, the entropy change has the opportunity to become negative. This gives quantitative meaning to the idea that we gain information when we observe rare fluctuations to lower-entropy macrostates. In particular, the entropy change may be negative if the information gain from the measurement is greater than the information loss due to irreversible dynamics.

3.6.2 Diffusion of a Gaussian in n Dimensions.

As our final analytic example, we consider a dynamical model that can be solved analytically. Let the configuration space be $\mathbb{R}^n$, and suppose the time evolution of the probability density is diffusive. That is,

$$\rho_\tau(x') = \int d^n x \frac{1}{(2\pi D\tau)^{n/2}} e^{-\frac{|x-x'|^2}{2D\tau}} \rho_0(x). \tag{3.6.13}$$

Then we can identify the transition function with the heat kernel:

$$\pi_F(x \to x') = \frac{1}{(2\pi D\tau)^{n/2}} e^{-\frac{|x-x'|^2}{2D\tau}}. \tag{3.6.14}$$

We will assume for simplicity that the diffusion is unaffected by time reversal, so that $\pi_F = \pi_R \equiv \pi$, and that the states x are also unaffected by time reversal. (Alternatively, we can assume that time-reversal is some sort of reflection in x. The distributions we consider will be spherically symmetric, and hence invariant under such reflections.) Note that since $\pi(x \to x') = \pi(x' \to x)$, this implies $\mathcal{Q} = 0$. We will analyze the system without including measurement, again for simplicity, and we will also assume that the initial density profile is Gaussian with initial width σ. Diffusion causes the Gaussian to spread:

$$\rho_\tau(x) = \frac{1}{(2\pi(\sigma + D\tau))^{n/2}} e^{-\frac{x^2}{2(\sigma+D\tau)}}. \tag{3.6.15}$$

We can also calculate the entropy as a function of time:

$$S(\tau) = \int d^n x \frac{1}{(2\pi(\sigma + D\tau))^{n/2}} e^{-\frac{x^2}{2(\sigma+D\tau)}} \left[\frac{x^2}{2(\sigma + D\tau)} + \frac{n}{2} \log(2\pi(\sigma + D\tau)) \right] \tag{3.6.16}$$

$$= \frac{n}{2} \log(\sigma + D\tau) + \frac{n}{2} \log 2\pi e. \tag{3.6.17}$$

Therefore, we have $\Delta S = \frac{n}{2} \log(1 + \frac{D\tau}{\sigma})$. The relative entropy $D(\rho_0 \| \tilde{\rho})$ is also easy to calculate, since in this case $\tilde{\rho} = \rho_{2\tau}$:

$$D(\rho_0 \| \tilde{\rho}) = \frac{n}{2} \left[\log\left(1 + \frac{2D\tau}{\sigma}\right) - \frac{2D\tau}{\sigma + 2D\tau} \right]. \tag{3.6.18}$$

The refined Second Law from monotonicity of the relative entropy says that $\Delta S \geq D(\rho_0 \| \tilde{\rho})$. Let us see how strong this is compared to $\Delta S \geq 0$. For small τ, we have $D(\rho_0 \| \tilde{\rho}) \approx n(D\tau/\sigma)^2$, as compared to $\Delta S \approx nD\tau/2\sigma$. So the bound from monotonicity is subleading in τ, so perhaps not so important. For large τ, though, we have $D(\rho_0 \| \tilde{\rho}) \approx \frac{n}{2}\left[\log \frac{D\tau}{\sigma} - \log \frac{e}{2}\right]$, as compared to $\Delta S \approx \frac{n}{2} \log \frac{D\tau}{\sigma}$. Now the bound is fairly tight, with the relative entropy matching the leading behavior of ΔS.

3.6.3 Randomly Driven Harmonic Oscillator

As a slightly more detailed – and potentially experimentally realizable – example to which we can apply the Bayesian Second Law, we consider the harmonic oscillator. Imagine a single, massive particle confined to a one-dimensional harmonic potential, with spring constant and potential minimum treated as time-dependent control parameters, coupled to a heat bath which generates dissipative and fluctuating forces. Such a system may be described by the Fokker-Planck equation,

$$\frac{\partial \rho(x,p,t)}{\partial t} = \frac{2}{\tau_*}\rho(x,p,t) + \left(k(t)\left[x - z\left(t\right)\right] + \frac{2}{\tau_*}p\right)\frac{\partial \rho(x,p,t)}{\partial p} - \frac{p}{M}\frac{\partial \rho(x,p,t)}{\partial x} + \frac{2M}{\beta\tau_*}\frac{\partial^2 \rho(x,p,t)}{\partial p^2}. \tag{3.6.19}$$

Here we have defined τ_* to be the dissipation time-scale, $k(t)$ to be the spring constant, $z(t)$ to be the location of the potential's minimum, M to be the mass of the oscillator, and β to be the inverse temperature of the heat bath. For simplicity, we choose to work in units natural for this system by taking $\beta = 1$, $M = 1$, and $k(t = 0) = 1$. We also choose $\tau_* = 1$, so that we are in the interesting regime where the dissipation and oscillation time scales are comparable.

We assume that the experimenter is only capable of measuring the position of the particle and not its momentum. For a microstate with position x, we assume that $P(m|x)$ is given by a Gaussian distribution in m centered at x with standard deviation $\sigma = 0.2$. This means that the experimenter is likely to find a measured value m within a range ± 0.2 of the true position x. This measuring device is therefore quite sensitive when compared to the typical size of thermal fluctuations, which is of order unity.

There is no analytical solution to (3.6.19) in the regime of interest, so the system must be modeled numerically. This can be done by discretizing phase space on a lattice and using finite-difference methods to evolve the discrete probability distribution. We have performed this process using the finite element solver package FiPy [170] for the Python programming language. To elucidate different aspects of the BSL, we consider three different simulated experiments. The phase space evolution of these experiments

	Figure 3.3	Figure 3.4	Figure 3.5
$S(\rho_0)$	2.84	0.31	0.31
$S(\rho_\tau)$	2.91	2.93	2.96
ΔS	0.07	2.61	2.65
$\langle \mathcal{Q} \rangle_F$	−0.04	5.99	7.99
$\Delta S + \langle \mathcal{Q} \rangle_F$	0.02	8.61	10.64
$D(\rho_0 \| \tilde{\rho})$	0.01	7.68	10.64
$S(\rho_{0\|m})$	2.47	−0.43	0.31
$S(\rho_{\tau\|m})$	1.23	1.12	1.23
ΔS_m	−1.61	0.81	0.92
$D(\rho_{0\|m} \| \rho_0)$	1.01	0.70	< 0.01
$D(\rho_{\tau\|m} \| \rho_\tau)$	2.71	1.37	1.24
$H(\rho_{0\|m}, \rho_0)$	3.48	0.26	0.31
$H(\rho_{\tau\|m}, \rho_\tau)$	3.94	2.49	2.47
ΔH	0.46	2.23	2.16
$\langle \mathcal{Q} \rangle_{F\|m}$	−0.40	6.14	8.47
$\Delta H + \langle \mathcal{Q} \rangle_{F\|m}$	0.06	8.36	10.64
$D(\rho_{0\|m} \| \tilde{\rho}_m)$	0.04	7.65	10.63
LHS of Eqn 3.4.20	−2.01	6.94	9.39
RHS of Eqn 3.4.20	−2.03	6.235	9.39
$\left\lvert \frac{\text{LHS}-\text{RHS}}{\text{LHS}} \right\rvert$	< 0.01	0.10	< 0.01
$\left\langle \frac{P_R}{P_F} \right\rangle_F$	1.00	1.00	1.00
$\left\langle \frac{P_R}{P_F} \frac{\rho_0}{\tilde{\rho}} \right\rangle_F$	1.00	1.00	1.00
$\left\langle \frac{P_{R\|m}}{P_{F\|m}} \right\rangle_{F\|m}$	1.00	1.00	1.00
$\left\langle \frac{P_{R\|m}}{P_{F\|m}} \frac{\rho_{0\|m}}{\tilde{\rho}_m} \right\rangle_{F\|m}$	1.00	1.00	1.00

Table 3.2. List of thermodynamic properties calculated for three numerically simulated experiments.

is shown in Figures 3.3 - 3.5, found in Appendix 3.A, while the thermodynamic quantities calculated are tabulated in Table 3.2. The source code which was used to carry out these simulations and animations of the evolution are also available.[4]

We first consider the simple experiment shown in Figure 3.3. The system begins in thermal equilibrium, Figure 3.3a. The experiment is carried out under a "trivial"

[4]See: http://preposterousuniverse.com/science/BSL/

protocol, where the experimenter fixes $k(t) = 1$ and $z(t) = 0$. Under this protocol, the system is allowed to evolve from $t = 0$ to $t = 1$ before a measurement is performed. As seen in Figure 3.3b, the thermal distribution is nearly unchanged by this evolution. (Due to finite-size effects, the thermal distribution is not perfectly stationary). At the end of the experiment, a measurement of the position is made and we assume that the unlikely fluctuation $m = 2$ is observed. The experimenter can then use this information to perform a Bayesian update on both the initial and final distributions as shown in Figures 3.3d and 3.3e. To evaluate the irreversibility of this experiment, the experimenter must also examine the time-reversed process. The updated cycled distribution which results from evolving under the time-reversed protocol is shown in Figure 3.3f.

While this experiment and its protocol are fairly simple, they illustrate several key features of the Bayesian Second Law. Before the final measurement is performed, the experimenter would state that $\Delta S = 0.07$. After performing the measurement, this becomes $\Delta S_m = -1.61$ with a heat transfer of $\langle \mathcal{Q} \rangle_{F|m} = -0.40$. Naively using these updated quantities in (3.3.7) leads to an apparent violation of the usual Second Law of Thermodynamics. However, this is remedied when one properly takes into account the information gained as a result of the measurement. A more careful analysis then shows $\Delta H = 0.46$ and $D(\rho_{0|m}|\tilde{\rho}_m) = 0.04$. As such, we see that (3.4.19) is satisfied and that the inequality is very tight.

We will now consider the same (trivial) protocol with a different initial distribution. The experimenter knows the initial position of the oscillator and the magnitude, but not the direction, of its initial momentum with a high degree of certainty. As such, there are two regions of phase space the experimenter believes the system could be in. The initial distribution is shown in Figure 3.4a. The system is then allowed to evolve until $t = 0.5$ as shown in Figure 3.4b. At the end of the experiment, the position of the oscillator is measured to be $m = 2$. The impact of this measurement can be seen in Figures 3.4d and 3.4e.

Due to the outcome of the measurement, the experimenter is nearly certain that the oscillator had positive initial momentum. One therefore expects this information gain to be roughly one bit and this is confirmed by $D(\rho_{0|m}\|\rho_0) = 0.70 \approx \log 2$. Despite this sizable information gain for the initial distribution, we note that the information gain for the final distribution is even greater with $D(\rho_{\tau|m}\|\rho_\tau) = 1.37$. This is expected because, regardless of the measurement outcome, the experimenter will always gain at least as much information about the final distribution than the initial when performing a measurement. Evaluating the remaining terms, see Table 3.2, we once again find that the BSL is satisfied.

Lastly, consider an experiment that starts with the same initial state but uses a

non-trivial protocol where the potential is "dragged". The experimenter keeps $k(t) = 1$ fixed but varies $z(t)$. For times between $t = 0$ and $t = 1$, the experimenter rapidly drags the system according to $z(t \leq 1) = 2t$. After this rapid dragging motion, the experimenter keeps $z(t > 1) = 2$ and allows the system to approach equilibrium until a measurement performed at $t = 5$. Importantly, this gives the system a significant amount of time to reach its new equilibrium distribution before the measurement is performed. The experimenter then measures the oscillator's position and finds it to be centered in the new potential ($m = 2$). The evolution of this system is shown in Figure 3.5.

Due to the change in protocol, the experimenter gains an appreciable amount of information about the final distribution of the system, but negligible information about the initial distribution. Specifically, we find that $D(\rho_{\tau|m}\|\rho_\tau) = 1.24$, while $D(\rho_{0|m}\|\rho_0) < 0.01$. This is because the system is given time to fully thermalize before the measurement, so any information about the initial state is lost by the time the measurement is performed. Also of interest is the difference between the forward and reverse protocol. As shown in Figures 3.5a and 3.5b, the forward protocol results in most distributions reaching the new thermal equilibrium. However, the same is not true of the reverse protocol: the distributions in Figures 3.5c and 3.5f are not near equilibrium. This is due to the asymmetry between the forward and reverse protocols.

We also calculated the quantities appearing in the Bayesian integral fluctuation theorems derived in Section 3.5; they appear in Table 3.2. We find that for all three experimental protocols considered, these are well defined and equal to unity.

3.7 Discussion

We have shown how to include explicit Bayesian updates due to measurement outcomes into the evolution of probability distributions obeying stochastic equations of motion, and derived extensions of the Second Law of Thermodynamics that incorporate such updates. Our main result is the Bayesian Second Law, which can be written in various equivalent forms (3.4.1), (3.4.6), (3.4.11), (3.4.13):

$$D(P_{F|m}\|P_{R|m}) \geq 0, \tag{3.7.1}$$

$$\Delta H(\rho_m, \rho) + \langle \mathcal{Q} \rangle_{F|m} \geq 0, \tag{3.7.2}$$

$$\Delta S_m + \langle \mathcal{Q} \rangle_{F|m} \geq -D(\rho_{\tau|m}\|\rho_\tau) + \int dx\ (\rho_0(x) - \rho_{0|m}(x)) \log \rho_0(x), \tag{3.7.3}$$

$$\Delta S(\rho_m) + \langle \mathcal{Q} \rangle_{F|m} \geq D(\rho_{0|m}\|\rho_0) - D(\rho_{\tau|m}\|\rho_\tau). \tag{3.7.4}$$

We also used monotonicity of the relative entropy to derive refined versions of the ordinary Second Law and the BSL, (3.3.12) and (3.4.19):

$$\Delta S + \langle \mathcal{Q} \rangle_F \geq D(\rho_0 \| \tilde{\rho}) \geq 0, \tag{3.7.5}$$

$$\Delta H(\rho_m, \rho) + \langle \mathcal{Q} \rangle_{F|m} \geq D(\rho_{0|m} \| \tilde{\rho}_m) \geq 0. \tag{3.7.6}$$

Finally, we applied similar reasoning to obtain Bayesian integral fluctuation theorems, such as (3.5.7):

$$\left\langle \frac{P_{R|m}}{P_{F|m}} \right\rangle_{F|m} = b_m \leq 1. \tag{3.7.7}$$

In the remainder of this section we briefly discuss some implications of these results.

Downward fluctuations in entropy. As mentioned in the Introduction, there is a tension between a Gibbs/Shannon information-theoretic understanding of entropy and the informal idea that there are rare fluctuations in which entropy decreases. The latter phenomenon is readily accommodated by a Boltzmannian definition of entropy using coarse-graining into microstates, but it is often more convenient to work with distribution functions $\rho(x)$ on phase space, in terms of which the entropy of a system with zero heat flow will either increase or remain constant.

The BSL resolves this tension. The post-measurement entropy of the updated distribution $\rho_{\tau|m}$ can be less than the original starting entropy ρ_0, as the right-hand side of (3.7.3) can be negative. On the rare occasions when that happens, there is still a lower bound on their difference. From the information-theoretic perspective, downward fluctuations in entropy at zero heat flow are necessarily associated with measurements.

This perspective is also clear from the refined Bayesian version of the Boltzmann Second Law (3.6.11), in which the right-hand side can be of either sign. We can see that downward fluctuations in entropy at zero heat flow occur when the amount of information gained by the experimenter exceeds the amount of information lost due to irreversible dynamics.

The usefulness of the BSL is not restricted to situations in which literal observers are making measurements of the system. We might be interested in fluctuating biological or nanoscale systems in which a particular process of interest necessarily involves a downward fluctuation in entropy. In such cases, even if there are no observers around to witness the fluctuation, we may still be interested in conditioning on histories in which such fluctuations occur, and asking questions about the evolution of entropy along the way. The BSL can be of use whenever we care about evolution conditioned on certain measurement outcomes.

The Bayesian arrow of time. Shalizi [171] has previously considered the evolution of conservative systems with Bayesian updates. For a closed, reversible system,

the Shannon entropy remains constant over time, as the distribution evolves in accordance with Liouville's Theorem. If we occasionally observe the system and use Bayes's rule to update the distribution, our measurements will typically cause the entropy to decrease, because conditioning reduces entropy when averaged over measurement outcomes, $\langle S(\rho_m)\rangle_m \leq S(\rho)$. At face value, one might wonder about an apparent conflict between this fact and the traditional understanding of the arrow of time, which is based on entropy increasing over time. This should be a minor effect in realistic situations, where systems are typically open and ordinary entropy increase is likely to swamp any decrease due to conditioning, but it seems like a puzzling matter of principle.

Our analysis suggests a different way of addressing such situations: upon making a measurement, we can update not only the current distribution function, but the distribution function at all previous times as well. As indicated by (3.7.4), the entropy of the updated distribution can decrease even at zero heat transfer. We have identified, however, a different quantity; the cross entropy $H(\rho_m, \rho)$ of the updated distribution with respect to the unupdated one, which has the desired property of never decreasing (3.7.2). For a closed system, both the updated entropy and the cross entropy will remain constant; for open systems the cross entropy will increase. It is possible to learn about a system by making measurements, but we will always know as much or more about systems in the past than we do about them in the present.

Statistical physics of self-replication. The application of statistical mechanics to the physics of self-replicating biological systems by England [140] was one of the inspirations for this work. England considers the evolution of a system from an initial macrostate, **I**, to a final macrostate, **II**, and finds an inequality which bounds from below the sum of the heat production and change in entropy by a quantity related to the transition probabilities between the two macrostates. This inequality, however, does not explicitly make use of a Bayesian update based on the observation of the system's final macrostate: as we have seen previously, the inclusion of Bayesian updates can significantly change one's interpretation of the entropy production.

In seeking to interpret England's inequality within our framework, we consider the form of the BSL in an experiment where the initial distribution has support only on the initial macrostate, and the measurement at the conclusion determines the final macrostate. This is a slight generalization of the Boltzmann setup considered in Section 3.6.1 above. We then have the option to consider the difference between the entropy of the updated final distribution and the entropy of either the updated or unupdated initial distribution.

First, making use of the unupdated initial state, it can be shown that

$$S(\rho_{\tau|\mathbf{II}}) - S(\rho_0) + \langle \mathcal{Q} \rangle_{F|\mathbf{II}} \geq -\log \frac{\pi(\mathbf{II} \to \mathbf{I})}{\pi(\mathbf{I} \to \mathbf{II})} + S(\rho_{0|\mathbf{II}}) - S(\rho_0). \tag{3.7.8}$$

This inequality is similar in spirit to England's. When $S(\rho_{0|\mathbf{II}}) \geq S(\rho_0)$, England's inequality immediately follows. Alternatively, using the updated initial state, we find

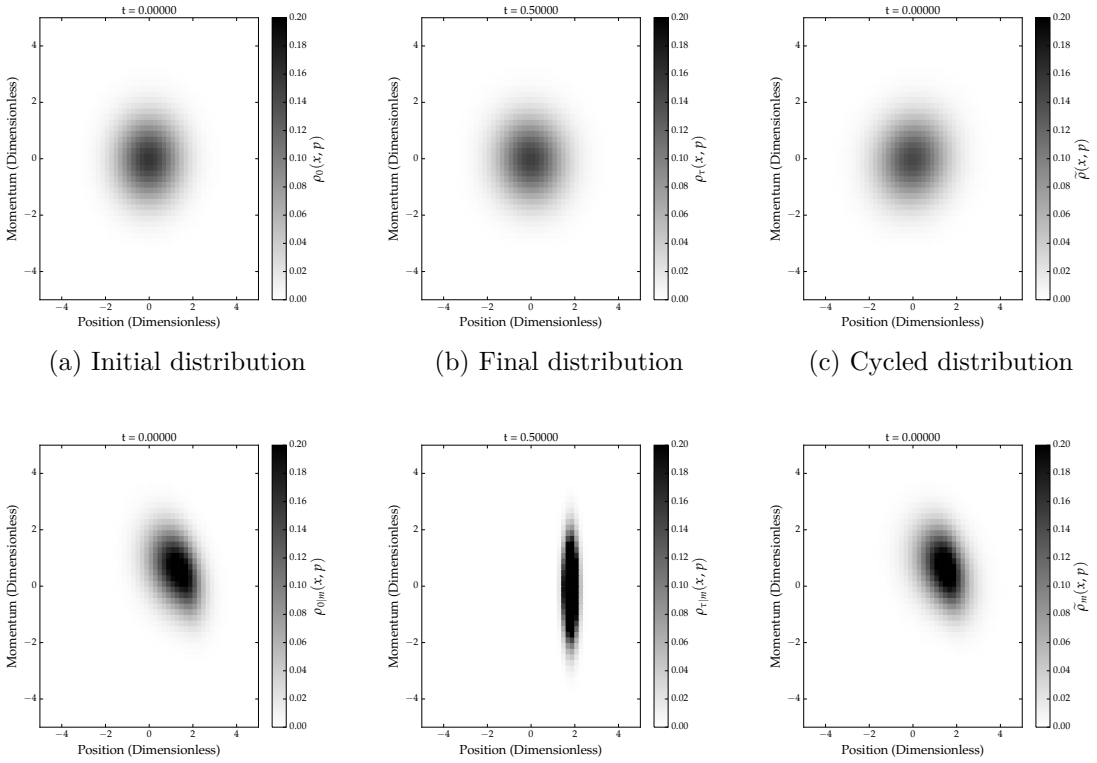
$$S(\rho_{\tau|\mathbf{II}}) - S(\rho_{0|\mathbf{II}}) + \langle \mathcal{Q} \rangle_{F|\mathbf{II}} \geq D(\rho_{0|\mathbf{II}} \| \tilde{\rho}_{\mathbf{II}}) + D(\rho_{0|\mathbf{II}} \| \rho_0) - D(\rho_{\tau|\mathbf{II}} \| \rho_\tau) \geq -\log \frac{\pi(\mathbf{II} \to \mathbf{I})}{\pi(\mathbf{I} \to \mathbf{II})}. \tag{3.7.9}$$

This differs from England's result only in that the entropy of the initial state has been replaced by the entropy of the updated initial state. Making this adjustment to England's inequality, we recover his bound from the bound given by the BSL. (We thank Timothy Maxwell for proving this relation).

Future directions. In this paper we have concentrated on incorporating Bayesian updates into the basic formalism of statistical mechanics, but a number of generalizations and applications present themselves as directions for future research. Potential examples include optimization of work-extraction (so-called "Maxwell's demon" experiments) and cooling in nanoscale systems, as well as possible applications to biological systems. It would be interesting to experimentally test the refined versions of the ordinary and Bayesian Second Laws, to quantify how close the inequalities are to being saturated. We are currently working to extend the BSL to quantum systems.

3.A Oscillator Evolution

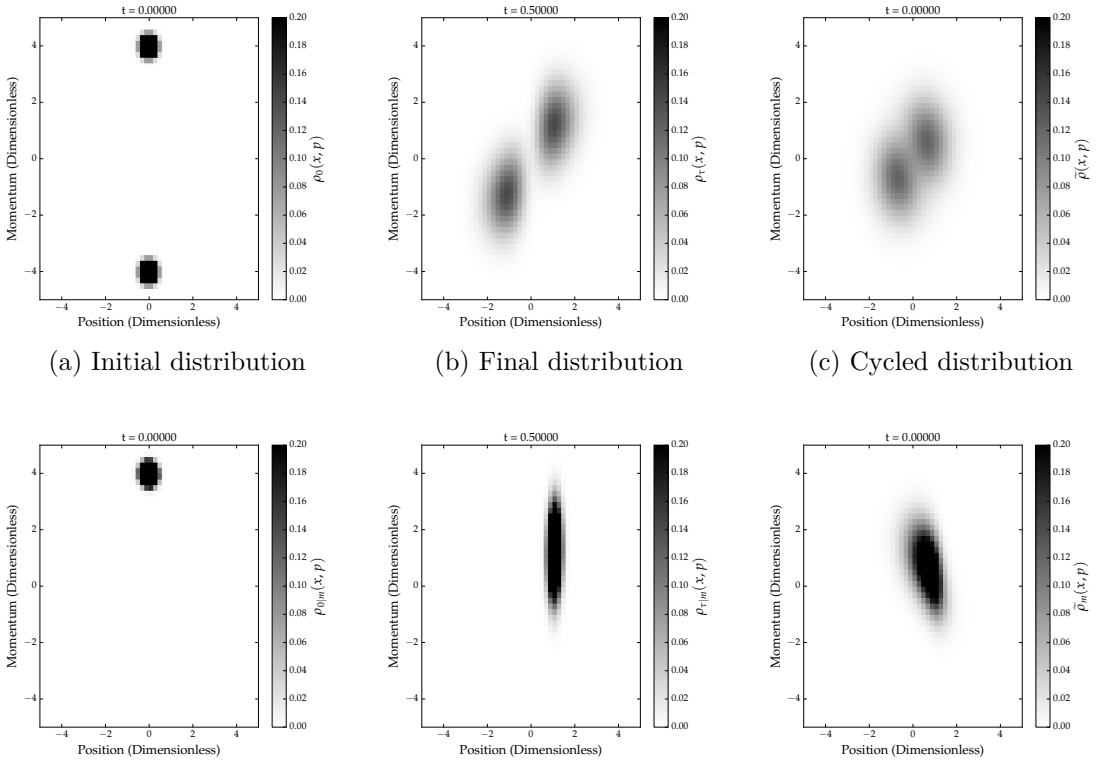
Here we show plots of the distribution functions for the three numerical harmonic-oscillator experiments discussed in Section 3.6.3.



(a) Initial distribution (b) Final distribution (c) Cycled distribution

(d) Updated initial distribution (e) Updated final distribution (f) Updated cycled distribution

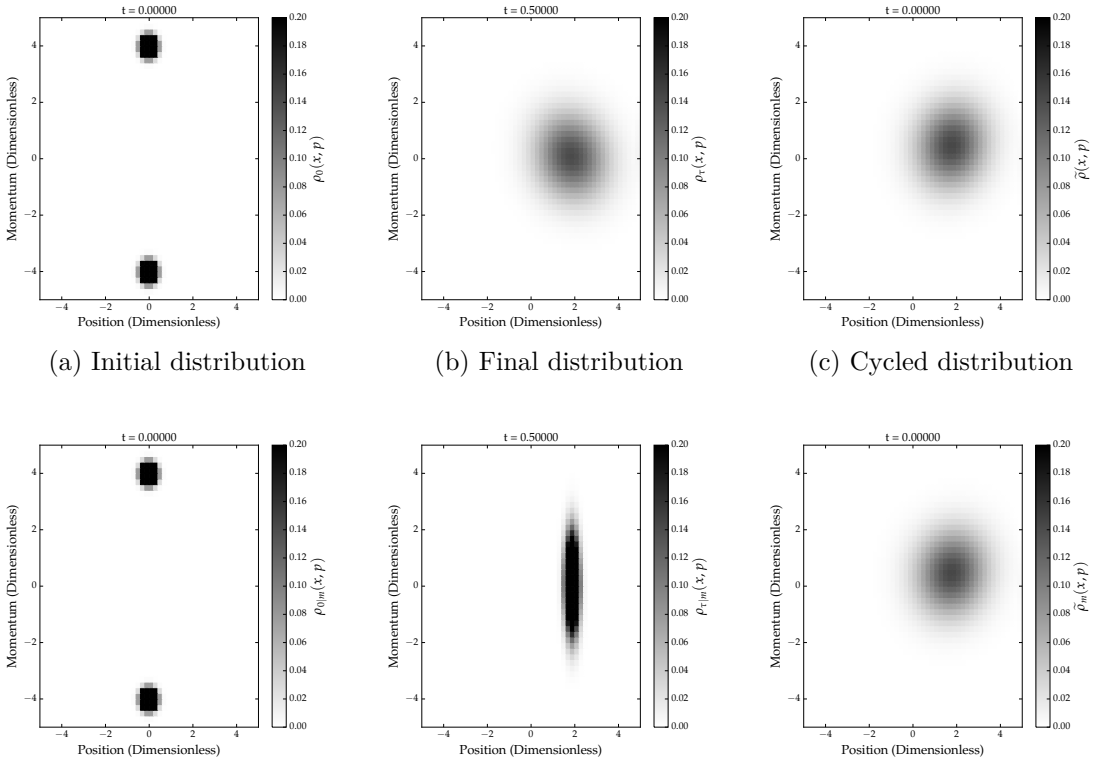
Figure 3.3. Evolution of a damped harmonic oscillator coupled to a heat bath in initial thermal equilibrium under a trivial protocol. Units are chosen such that $M = 1$, $k(t = 0) = 1$, and $\beta = 1$. Each graph shows the phase space probability distribution with respect to position and momentum at different points in the experiment.



(a) Initial distribution (b) Final distribution (c) Cycled distribution

(d) Updated initial distribution (e) Updated final distribution (f) Updated cycled distribution

Figure 3.4. Evolution of a damped harmonic oscillator coupled to a heat bath with known position and magnitude of momentum under a trivial protocol. Units are chosen such that $M = 1$, $k(t = 0) = 1$, and $\beta = 1$. Each graph shows the phase space probability distribution with respect to position and momentum at different points in the experiment.



(a) Initial distribution (b) Final distribution (c) Cycled distribution

(d) Updated initial distribution (e) Updated final distribution (f) Updated cycled distribution

Figure 3.5. Evolution of a damped harmonic oscillator coupled to a heat bath in initial thermal equilibrium under a "dragging" protocol. Units are chosen such that $M = 1$, $k(t = 0) = 1$, and $\beta = 1$. Each graph shows the phase space probability distribution with respect to position and momentum at different points in the experiment.

Jarzynski Equality for Driven Quantum Field Theories

The fluctuation theorems, and in particular the Jarzynski equality, are the most important pillars of modern non-equilibrium statistical mechanics. We extend the quantum Jarzynski equality together with the Two-Time Measurement Formalism to their ultimate range of validity – to quantum field theories. To this end, we focus on a time-dependent version of scalar phi-four. We find closed form expressions for the resulting work distribution function, and we find that they are proper physical observables of the quantum field theory. Also, we show explicitly that the Jarzynski equality and Crooks fluctuation theorems hold at one-loop order independent of the renormalization scale. As a numerical case study, we compute the work distributions for an infinitely smooth protocol in the ultra-relativistic regime. In this case, it is found that work done through processes with pair creation is the dominant contribution.

4.1 Introduction

In physics there are two kinds of theories to describe motion: microscopic theories whose range of validity is determined by a length scale and the amount of kinetic energy, such as classical mechanics or quantum mechanics; and phenomenological theories, such as thermodynamics, which are valid as long as external observables remain close to some equilibrium value.

Over the last two centuries, microscopic theories have undergone a rapid development from classical mechanics over special relativity and quantum mechanics to quantum field theory. While quantum field theories were originally developed for particle physics and cosmology, this apporach has been shown to be powerful in the description of condensed matter systems. Examples include: quasiparticle excitations in graphene, cavity quantum electrodynamics, topological insulators, and many more [7–9, 172].

In contrast, the development of thermodynamics has been rather stagnant – until only two decades ago when the first fluctuation theorems were discovered [43, 44, 173]. Conventional thermodynamics can only fully describe infinitely slow, equilibrium processes. About all real, finite-time processes the second law of thermodynamics only asserts that some amount of entropy is dissipated into the environment, which can be expressed with the average, irreversible entropy production as $\langle \Sigma \rangle \geq 0$ [174]. The (detailed) fluctuation theorem makes this statement more precise by expressing that negative fluctuations of the entropy production are exponentially unlikely [41, 43, 44],

$$\mathcal{P}(-\Sigma) = \exp(-\Sigma)\, \mathcal{P}(\Sigma)\,. \tag{4.1.1}$$

The most prominent (integral) fluctuation theorem [175] is the Jarzynski equality [39], which holds for all systems initially prepared in equilibrium and undergoing isothermal processes,

$$\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta F)\,, \tag{4.1.2}$$

where β is the inverse temperature, W is the thermodynamic work, and ΔF is the free energy difference between the instantaneous equilibrium states at the initial and final times. In its original inception the Jarzynski equality (4.1.2) was formulated for classical systems with Hamiltonian [39] and Langevin dynamics [40]. Thus, W is essentially a notion from classical mechanics, where work is given by a force along a trajectory. The advent of modern fluctuation theorems for classical systems [39–41, 43, 44, 173, 176] has spurred the development of a new field, which has been dubbed stochastic thermodynamics [42, 138, 141, 177–179].

In stochastic thermodynamics one focuses on the fluctuating properties of the central quantities, such as work and heat, which are defined for single realizations of

processes operating far from equilibrium. Thus, in the study of nanoscale systems out of thermal equilibrium it is natural to ask in what regimes quantum effects become significant and how fluctuation theorems apply to quantum systems [45–58]. Nevertheless, it took another decade before it was clearly stated that in quantum mechanical systems W is not a quantum observable in the usual sense [59]. This means that there is no hermitian operator, whose eigenvalues are the classically observable values of W. This is the case because thermodynamic work is a path dependent quantity – a non-exact differential. Hence, thermodynamic work is given by a time-ordered correlation function [48, 49, 59].

To gain more insight into the underlying statistics of quantum work the Two-Time Measurement Formalism [60, 61] has proven powerful: in this formulation, a quantum system is prepared in contact with a heat bath of inverse temperature β. The system is then decoupled from the environment and a projective measurement onto the initial energy eigenbasis is performed. Then, the system is let to evolve before another projective measurement of the energy is performed. As the system is isolated, the work performed on the system is identical to the change in energy. Despite its success, this formalism has several limitations [62], including the lack of thermodynamic accounting for the measurement process [63] and its inapplicability to coherently controlled quantum systems [64]. Nevertheless, it is important to remark that in complete analogy to how classical mechanics is contained in quantum mechanics (in the appropriate limits) the Two-Time Measurement Formalism produces work distribution functions which correspond to those of classical systems in semiclassical approximations [180–184].

To date another decade has gone by, and quantum stochastic thermodynamics is still incomplete. How to describe thermodynamic work and entropy production in open quantum systems is still hotly debated [152, 185–194], and with a few exceptions [195–204] most of the literature is restricted to standard Schrödinger quantum mechanics.

The purpose of the present analysis is to significantly broaden the scope of stochastic thermodynamics, and take the next, important step – extending quantum stochastic thermodynamics to interacting quantum field theories. Conventional thermodynamics is a phenomenological theory that has no knowledge of the underlying microscopic dynamics. In small systems, the dynamics are governed by fluctuations, and in particular heat and work become fluctuating quantities. The magnitude and characteristics of these fluctuations, however, are determined by the underlying dynamics and it crucially matters whether one studies a classical, a quantum mechanical, or a quantum field theoretic model. Therefore, "stochastic thermodynamics of quantum field theories" studies the fluctuations of the standard thermodynamic quantities as arising from a quantum field theory in contrast to classical, thermal noise.

In the following, we demonstrate that the Two-Time Measurement Formalism can

be systematically used to investigate the work distribution functions of a restricted class of quantum field theories, focusing on a time-dependent version of $\lambda\phi^4$. Closed form expressions for these work distributions are found at leading order, including loop corrections, through the use of a new diagrammatic technique and a mapping between finite-time transition amplitudes and infinite-time scattering amplitudes. It is found that to the perturbative order considered, the work distribution function does not run with the renormalization scale indicating that the distribution is an observable of the quantum field theory. We verify that the quantum Jarzynski and Crooks fluctuation theorems hold exactly and are independent of the renormalization scale. Due to the form of the work distributions, it is straightforward to show that the fluctuation theorems hold if one removes the loop corrections (as would be the case for a classical field theory), and also in the non-relativistic limit.

These results demonstrate that quantum fluctuation theorems and stochastic thermodynamics can be extended to include quantum field theories, our most fundamental theory of nature. Thus, our results open the door for future application of fluctuation theorems to the study of problems at the forefronts of physics – in condensed matter physics, particle physics, and cosmology.

This paper is organized as follows: In Sec. 4.2 we review the Two-Time Measurement Formalism and the quantum Jarzynski equality. We define a restricted class of quantum field theories in Sec. 4.3 for which the work distribution function can be calculated. The energy projection operators for a generic real scalar field theory are calculated in Sec. 4.4 and a method for calculating finite-time transition amplitudes from infinite-time scattering amplitudes is introduced. The mathematical details of this relationship between finite-time and infinite-time amplitudes are detailed in Appendix 4.A. In Sec. 4.5 we specialize to a time-dependent version of $\lambda\phi^4$ and discuss its renormalization. Then, Sec. 4.6 discusses how closed form expressions for the work distribution function can be calculated at leading order using a graph theoretic technique. The details of the derivation can be found in Appendix 4.B while the closed form expressions for the work distribution function are in Appendix 4.C. We discuss the analytic properties of the work distribution function in Sec. 4.7 and analytically verify both the Crooks fluctuation theorem and quantum Jarzynski equality at leading order for time-dependent $\lambda\phi^4$. In Sec. 4.8 we numerically evaluate the work distribution function for a relativistic bath and a particular driving protocol, and verify the fluctuation theorems. Interestingly, we find that the dominant process in the work distribution function is particle pair-production through a loop diagram, an effect only found in a quantum field theory. We conclude in Sec. 4.9 with a few remarks.

4.2 Preliminaries: Two-Time Measurement Formalism

We begin by reviewing the Two-Time Measurement Formalism to establish notions and notation [49]: A quantum system is initially, at $t = t_1$, in thermal equilibrium with a classical heat bath of inverse temperature β [1]. At $t = t_1 + 0^+$ the system is disconnected from the heat bath and the energy of the system is projectively measured to be E_1. The system then evolves according to a time dependent protocol until time $t = t_2$. At this time, the energy of the system is measured to be E_2.

Let $\hat{H}(t)$ be the Hamiltonian at time t and let $U(t_2, t_1)$ be the time evolution operator from t_1 to t_2, and $\hat{\Pi}_E$ is the energy projection operator onto the (potentially degenerate) subspace of eigenstates with energy E. This projection operator is time-dependent due to the time-dependent Hamiltonian, but for compactness of notation, this dependence will be implicit.

As the system starts in equilibrium, the initial state of the system is given by the thermal density matrix

$$\hat{\rho}_0 = \frac{\exp\left(-\beta \hat{H}(t_1)\right)}{\mathrm{tr}\left\{\exp\left(-\beta \hat{H}(t_1)\right)\right\}}. \tag{4.2.1}$$

The probability of measuring energy E_1 at time t_1 is then given by

$$P(E_1) = \mathrm{tr}\left\{\hat{\Pi}_{E_1} \hat{\rho}_0\right\}, \tag{4.2.2}$$

with the normalized post-measurement state

$$\hat{\rho}_{E_1} = \frac{\hat{\Pi}_{E_1} \hat{\rho}_0 \hat{\Pi}_{E_1}}{\mathrm{tr}\left\{\hat{\Pi}_{E_1} \hat{\rho}_0 \hat{\Pi}_{E_1}\right\}}. \tag{4.2.3}$$

After being projected into the E_1 energy subspace, the system is evolved according to a time-dependent protocol. The conditional probability of measuring energy E_2 is

$$P(E_2|\, E_1) = \mathrm{tr}\left\{\hat{\Pi}_{E_2} U(t_2, t_1) \hat{\rho}_{E_1} U(t_1, t_2)\right\}. \tag{4.2.4}$$

Importantly, the system is isolated from, or at least very weakly coupled to, the heat bath during its evolution. As such, the work performed by the experimenter on the system can be identified with the change in system energy, $W \equiv E_2 - E_1$. One may then define the work distribution function

$$\mathcal{P}(W) = \sum\!\!\!\!\!\!\int_{E_1, E_2} \delta\left(W - E_2 + E_1\right) P(E_1, E_2). \tag{4.2.5}$$

[1] The choice of an initial Gibbs state is not generic, however it allows one to make contact with classical thermodynamic quantities.

Using the definition of the joint probability distribution and Eqs. (4.2.1)-(4.2.5),

$$\mathcal{P}(W) = \sum\!\!\!\!\!\!\int_{E_1,E_2} \delta\left(W - E_2 + E_1\right) \frac{\mathrm{tr}\left\{\hat{\Pi}_{E_1}\hat{\rho}_0\right\}}{\mathrm{tr}\left\{\hat{\Pi}_{E_1}\hat{\Pi}_{E_1}\hat{\rho}_0\right\}} \mathrm{tr}\left\{\hat{\Pi}_{E_2}U(t_2,t_1)\hat{\Pi}_{E_1}\hat{\rho}_0\hat{\Pi}_{E_1}U(t_1,t_2)\right\}. \tag{4.2.6}$$

This expression differs from what has previously been shown in the literature due to the presence of the ratio of traces of the projection operators. This is because in previous works the quantum system of interest was assumed to have a discrete energy eigenspectrum. As a consequence, the energy projection operator can be thought of as an idempotent matrix, *i.e.* $\hat{\Pi}_{E_1}\hat{\Pi}_{E_1} = \hat{\Pi}_{E_1}$, and thus this additional term is trivial. However, for systems with a continuum of states the projection operator involves a delta-function which is not idempotent and has non-zero mass dimension. As such, this additional term is essential for proper normalization of the work distribution when one considers a quantum system with a continuum of states.

If the time evolution of system is at least unital [2], the quantum Jarzynski equality [59–61, 168, 206–208] follows from (4.2.5),

$$\int dW\, \mathcal{P}\left(W\right) \exp\left(-\beta W\right) = \exp\left(-\beta \Delta F\right). \tag{4.2.7}$$

In this expression, ΔF is the change in free energy from the instantaneous equilibrium distribution at time t_1 to time t_2.

4.3 Restricted Field Theories

The work distribution function (4.2.6) and corresponding quantum Jarzynski equality (4.2.7) are natural objects to consider in the context of non-equilibrium statistical physics. The work distribution function fully classifies all fluctuations involving energy transfer and the quantum Jarzynski equality strongly constrains the form of these fluctuations [209]. However, $\mathcal{P}(W)$, (4.2.6), is not phrased in a natural manner for studying a quantum field theory. The work distribution function requires one to know the energy projection operators, $\hat{\Pi}_E$, for the Hamiltonian at the initial and final times. For a generic quantum field theory, the calculation of these operators may prove intractable. Furthermore, (4.2.6) is a fundamentally finite-time object as one is performing energy projection measurements at times t_1 and t_2. Usually, quantum field theory is applied to infinite-time scattering processes as is commonly done in particle physics [5]. This

[2] A unital map is a completely positive map which preserves the identity. More simply, any superposition of unitary quantum maps is a unital map. [205]

approximation is valid in the context of particle physics because observations are made on timescales significantly greater than the characteristic timescale of particle dynamics. However, non-equilibrium work distributions are of greatest interest when these timescales are comparable.

Given the difficulties associated with the general case, we will restrict the class of quantum field theories and driving protocols considered. Working in the rest frame of the experimenter and heat bath, we will assume that the system is governed by a Hamiltonian of the form $H(t) = H_0 + H_I(t)$ where H_0 is the Hamiltonian for a free field theory. The interacting part of the Hamiltonian is assumed to be sufficiently smooth and have the general form

$$H_I(t) = \begin{cases} H_I(t), & \text{for } t \in (t_1, t_2) \\ 0, & \text{otherwise} \end{cases} . \tag{4.3.1}$$

It should be noted that these restrictions disallow gauge theories where the matter fields have fixed gauge charges. This is because even in the absence of a classical background field, charged particles self-interact and interact with each other through the exchange of gauge bosons.

Imposing these requirements, it follows that the energy projection operators needed at the beginning and end of the experiment are just those for a free field theory. Furthermore, as will be shown in Sec. 4.4 and Appendix 4.A, it will be possible to map the finite-time transition probability onto an infinite-time process because the theory is free at the initial and final times.

These assumptions are essential for our approach in finding the work distribution function. However, we will make an additional set of assumptions for both simplicity and definiteness. For the remainder of this paper, we will restrict ourselves to theories of a single real scalar field, ϕ, with non-zero mass, m. Such theories are described by the Lagrangian

$$\mathcal{L} = -\frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 + \Omega_0 + \mathcal{L}_{\text{int}}, \tag{4.3.2}$$

where the constant Ω_0 is included to cancel the zero-point energy. Note that we have chosen to work in units where $\hbar = c = 1$ and are using the Minkowski metric $\eta_{\mu\nu} = \text{diag}\,(-1, +1, +1, +1)$.

Despite their simplicity, such field theories (4.3.2) have applications across a wide variety of energy scales [4–6]: from phonons [210, 211], the Ginzburg-Landau theory of superconductivity [212], Landau's theory of second order phase transitions [213], and critical phenomena more generally [214] to the study of spontaneous symmetry breaking [215, 216], the Higgs mechanism [13, 217], and inflationary cosmology [218].

4.4 Projection Operators And Finite-Time Transitions

Given the form of the interaction (4.3.1), the energy projection operators are the free theory projection operators. Note that the free Hamiltonian commutes with the number operator. Hence, we can express energy projection operators as a sum over projections with definite energy, E, and particle number n. They can be written as

$$\hat{\Pi}_{E,n} = \int \widetilde{d^3k}_1 \ldots \widetilde{d^3k}_n \, \delta\left(E - \omega_1 - \ldots - \omega_n\right) \frac{1}{n!} \left|k_1, \ldots, k_n\right\rangle \left\langle k_1, \ldots, k_n\right| , \tag{4.4.1}$$

where $\omega_j = (m^2 + k_j^2)^{(1/2)}$ is the energy of the jth particle and $\widetilde{d^3k}_j = d^3k_j/\left(2\pi\right)^3 2\omega_j$ is the Lorentz invariant measure [4]. Summing over energetically degenerate subspaces we can further write $\hat{\Pi}_E = \sum_n \hat{\Pi}_{E,n}$. Even though the field theory has a mass gap, this general form holds for all energy projection operators, including the ground state projection with $E = 0$.

Returning to the work distribution function (4.2.6) and making use of these definitions for the energy projection operators, we obtain

$$\begin{aligned} \mathcal{P}(W) = \sum_{n_1,n_2} \int \prod_i^{n_1} \prod_j^{n_2} \widetilde{d^3k}_i \, \widetilde{d^3k'}_j \, \delta\left(W + \sum_{l=1}^{n_1} \omega_l - \sum_{l=1}^{n_2} \omega'_l\right) \\ \times \left| \frac{\left\langle k'_1, \ldots, k'_{n_2}\right| U(t_2, t_1) \left|k_1, \ldots, k_{n_1}\right\rangle}{\sqrt{n_2! n_1!}} \right|^2 \frac{\exp\left(-\beta \sum_{l=1}^{n_1} \omega_l\right)}{\mathrm{tr}\left\{\exp\left(-\beta \hat{H}_0\right)\right\}} . \end{aligned} \tag{4.4.2}$$

The distribution (4.4.2) is normalized by the free energy of the free field theory, $\mathrm{tr}\{\exp\left(-\beta \hat{H}_0\right)\} = \exp\left(-\beta F_0\right)$. The momenta of the incoming and outgoing particles are integrated over in a Lorentz invariant manner and thus the integration measure is frame independent. Furthermore, each incoming particle is associated with a Boltzmann weight $\exp\left(-\beta\omega\right)$. The single delta-function ensures conservation of energy. Lastly, the quantity $\left\langle k'_1, \ldots, k'_{n_2}\right| U(t_2, t_1) \left|k_1, \ldots, k_{n_1}\right\rangle$ is the finite-time transition amplitude for the time-dependent system.

To make use of the machinery of quantum field theory, it will be necessary to rewrite this finite-time amplitude in terms of an infinite-time scattering process. The mathematical details are in Appendix 4.A, but a high-level description and the intuition for the mapping are provided here.

Due to the restrictions placed on the form of the interaction Hamiltonian (4.3.1), the quantum field theory is free at the initial and final times. One can imagine extending the finite-time experiment outside of the interval $[t_1, t_2]$ by assuming the Hamiltonian

remains non-interacting before and after the projective energy measurements. As the Hamiltonian is time independent for $t \leq t_1$ and $t \geq t_2$, no additional work is performed and the work distribution function is identical to the finite-time process. Furthermore, as the projective measurements place the system in an energy eigenstate of the free theory at the initial and final times, these states can be evolved arbitrarily far into the past or future, respectively, in the Schrödinger picture at the cost of an overall, yet irrelevant, phase. Thus, we map the finite-time transition amplitude onto an infinite-time scatttering process, and we find

$$\begin{aligned}\left|\left\langle k'_1,\ldots,k'_{n_2}\right| U(t_2,t_1)\left|k_1,\ldots,k_{n_1}\right\rangle\right| = \Big| \int d^3x'_1 d^3x_1 \ldots \\ \times \exp\left(-ik'_1x'_1\right)\exp\left(ik_1x_1\right)\ldots \overleftrightarrow{\partial}_{0_{x'_1}} \overleftrightarrow{\partial}_{0_{x_1}} \ldots \\ \times_I \langle\Omega| T\left[U_I(\infty,-\infty)\phi_I(x'_1)\ldots\phi_I(x_1)\ldots\right]|\Omega\rangle_I \Big| .\end{aligned} \tag{4.4.3}$$

In this expression, the subscript I is used to indicate operators in the Interaction picture. The state $|\Omega\rangle_I$ is defined as the vacuum state of the free theory, *i.e.* $\hat{H}_0|\Omega\rangle_I = 0$. We also have $f\overleftrightarrow{\partial}_\mu g \equiv f\left(\partial_\mu g\right) - \left(\partial_\mu f\right) g$, see Ref. [4].

4.5 Renormalization of time-dependent theories

For non-trivial work to be performed on the system, the interaction Hamiltonian (4.3.1) must be time-dependent. This time-dependence breaks Lorentz invariance by singling out a preferred frame, the experimenter's frame. Thus, quantities such as energy and time are always measured with respect to this frame. This differs significantly from the usual approach to quantum field theory where Lorentz invariance essential [5]. As such, significant care must be taken in the definition and renormalization of the quantum field theory.

Formulation Generally, we may choose any time-dependent interaction in (4.3.2), however, we will focus on a time-dependent variant of $\lambda\phi^4$, and we have,

$$\mathcal{L}_{\text{int}} = -\frac{1}{4!}\lambda(t)\phi^4. \tag{4.5.1}$$

The time-independent $\lambda\phi^4$ is a renormalizable field theory [4–6], which can be shown rigorously through Dyson-Weinberg power counting arguments [219, 220]. Being renormalizable, the theory only requires a finite number of counterterms to cancel divergences due to loop corrections and is valid at all energy scales, up to considerations of strong coupling. However, these power counting arguments rely on the Lorentz invariance of

the field theory's Lagrangian density. As Lorentz invariance is broken in (4.5.1), it is not clear that this theory can be renormalized with a finite number of counterterms.

A mathematically equivalent, but more intuitive approach, is to rewrite this field theory as a non-renormalizable effective field theory with a classical source. This is done by promoting λ to a classical, non-dynamical, scalar field $\chi_{\rm cl}$ with mass M. This mass scale is assumed to be much greater than any other energy scale in the system and sets the cut-off scale for this effective field theory. As a book-keeping mechanism, it will be convenient to introduce a dimensionless parameter $g = 1$ to keep track of the perturbative expansion as the theory no longer has an explicit coupling constant. The Lagrangian density then becomes

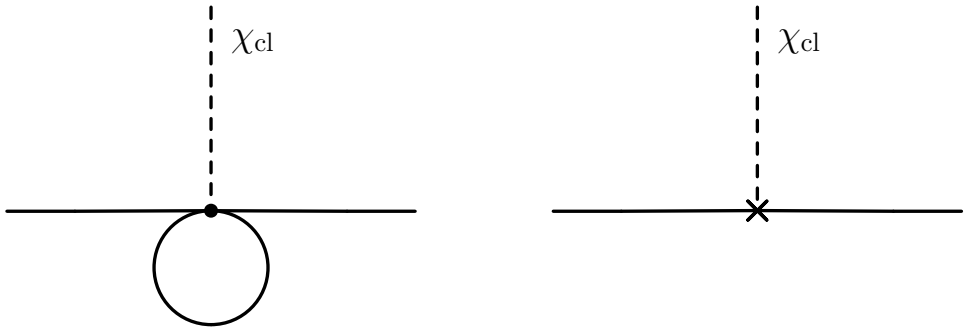
$$\mathcal{L} = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + \Omega_0 - \frac{g}{4!M}\chi_{\rm cl}\phi^4. \tag{4.5.2}$$

Additional interaction terms induced by the breaking of Lorentz invariance are suppressed by increasing powers of $\frac{g}{M}$. Equation (4.5.2) can be thought of as the leading order expression of (4.5.1) as an effective field theory. The interaction term in this theory has mass-dimension five and thus this theory is non-renormalizable [5]. Being non-renormalizable, an infinite set of counterterms is required to cancel divergences and the theory may only be applied at energy scales up to its cutoff, M. For present purposes, the counterterms of interest may be expressed as

$$\mathcal{L}_{\rm ctr} = -\sum_{j,k} c_{j,k}\frac{g^j}{M^j}\chi_{\rm cl}^j\phi^k. \tag{4.5.3}$$

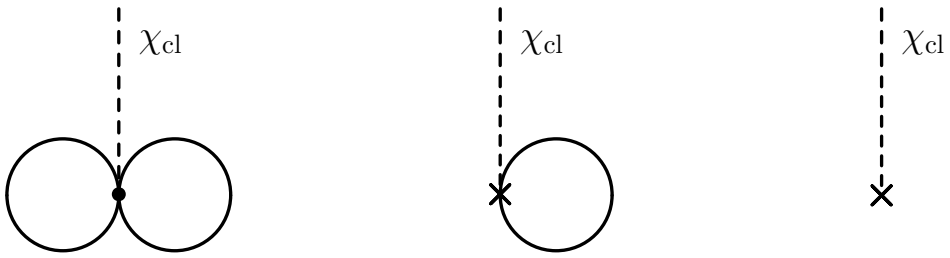
One key advantage of the effective field theory (4.5.2) over Eq. (4.5.1) is that the classical field $\chi_{\rm cl}$ can be thought of as a work reservoir [174, 221]. This reservoir sources all interactions and the $\chi_{\rm cl}$ field carries this energy into or out of the system. In the present case, this leads to more intuitive Feynman diagrams where energy is conserved at every vertex as opposed to the theory described in Eq. (4.5.1) where vertices only include ϕ, and hence do not conserve energy. Note, however, that the two approaches are mathematically fully equivalent and we may freely switch between them by identifying $g/M\,\chi_{\rm cl}(t) = \lambda(t)$.

Renormalization We will be working to leading order in the perturbative parameter $g = 1$. However, even at this order, we must consider loop diagrams which are formally divergent. We will use dimensional regularization in $d = 4-\epsilon$ dimensions to parameterize the divergences and work within the framework of the $\overline{MS}$ renormalization scheme at an energy scale μ to systematically assign values to the counterterms, see *e.g.* Refs. [4–6]. At leading order, there are only two divergent diagrams we will need to consider.



(a) Loop correction to the propagator. (b) Counterterm for the propagator.

Figure 4.1. Leading order corrections to the propagator of the scalar field ϕ. The interactions are sourced by insertions of the classical, non-dynamical field $\chi_{\rm cl}$.



(a) Two loop vacuum diagram. (b) Counterterm loop vacuum diagram. (c) Counterterm vacuum diagram.

Figure 4.2. Vacuum energy contributions of $\chi_{\rm cl}$. The interactions are sourced by insertions of the classical, nondynamical field $\chi_{\rm cl}$.

The first is the loop correction to the ϕ propagator, Fig. 4.1. The second is the vacuum energy diagram sourced by the classical field $\chi_{\rm cl}$, Fig. 4.2. In these diagrams, the scalar field ϕ is denoted by a solid line while the classical background field $\chi_{\rm cl}$ is represented by a dotted line.

We first consider the corrections to the ϕ propagator shown in Fig. 4.1. At leading order, the propagator is modified by a loop correction, Fig. 4.1a, whose divergent part is canceled by a counterterm, Fig. 4.1b. The relevant counterterm from the Lagrangian

(4.5.3) is $c_{1,2}\frac{g}{M}\chi_{\rm cl}\phi^2$. Working in $d = 4-\epsilon$ dimensions, the $\overline{MS}$ renormalization scheme requires us to fix

$$c_{1,2} = \frac{1}{2}\left(\frac{m}{4\pi}\right)^2 \frac{1}{\epsilon}. \tag{4.5.4}$$

In the limit of $\chi_{\rm cl}$ being a time independent background, the remaining finite part of the loop-diagram matches with that of regular $\lambda\phi^4$ theory.

The diagrams in Fig. 4.2 are used to calculate the change in vacuum energy of the ϕ field due to the classical background field $\chi_{\rm cl}$. The two loop diagram in Fig. 4.2a is the vacuum bubble induced by the $\chi_{\rm cl}$ background. Figure 4.2b includes the contribution of the counterterm fixed by the loop corrections to the propagator while Fig. 4.2c corresponds to the contribution of the $c_{1,0}\frac{g}{M}\chi_{\rm cl}$ counterterm. As the counterterm in Fig. 4.2b is already fixed by the propagator, the $\overline{MS}$ scheme requires the choice

$$c_{1,0} = \frac{1}{2}\left(\frac{m}{4\pi}\right)^4 \frac{1}{\epsilon^2}. \tag{4.5.5}$$

The remaining finite part of the diagrams in Fig. 4.2 is given by

$$\ldots = -\frac{i}{8}\left(\frac{m}{4\pi}\right)^4 \int d^4z\, \frac{g}{M}\chi_{\rm cl}(z). \tag{4.5.6}$$

The expression (4.5.6) involves the integral over all space and time of the background field $\chi_{\rm cl}$. This quantity will be formally infinite unless the system is restricted to a large but finite spatial volume V. As $\chi_{\rm cl}$ is spatially uniform in the experimenter's frame, it then follows that

$$\ldots = -\frac{i}{8}\left(\frac{m}{4\pi}\right)^4 V \int dt\, \frac{g}{M}\chi_{\rm cl}(t). \tag{4.5.7}$$

Disconnected Vacuum Diagrams In the standard framework of quantum field theory, one assumes that all interactions are switched on and off adiabatically in the distant past and future. Consequently, for a field theory with a mass gap, one can use the adiabatic theorem to show that the disconnected vacuum diagrams only contribute an irrelevant overall phase to any scattering amplitude [222, 223]. However, non-equilibrium evolution requires that the interaction parameters are varied non-adiabatically. As such, one must include the disconnected vacuum diagrams in the calculation of scattering amplitudes.

Consider an n-point correlation function of the form $\langle\Omega|\, T\left[U_I(\infty,-\infty)\phi_I(x_1)\ldots\right]|\Omega\rangle$. For a contributing Feynman diagram, we will call any part of the diagram that can be traced to an external field source $\phi_I(x_i)$ connected. The contributions of all such connected diagrams will be denoted by $\langle\Omega|\, T\left[U_I(\infty,-\infty)\phi_I(x_1)\ldots\right]|\Omega\rangle_C$. Any other component of the diagram will be considered a disconnected vacuum subdiagram. Note

that all such vacuum diagrams necessarily involve the background field $\chi_{\rm cl}$ as it sources all interactions. The contribution of the set of disconnected vacuum diagrams is given by $\langle\Omega|\, U_I(\infty,-\infty)\,|\Omega\rangle$.

The n-point correlation function factorizes into the product of the connected n-point diagrams, $\langle\Omega|\, T\left[U_I(\infty,-\infty)\phi_I(x_1)\ldots\right]|\Omega\rangle_C$, and the vacuum diagrams, $\langle\Omega|\, U_I(\infty,-\infty)\,|\Omega\rangle$. The vacuum diagram contribution $\langle\Omega|\, U_I(\infty,-\infty)\,|\Omega\rangle$ has the property that it can be expressed as the exponential of the sum of all unique vacuum diagrams. This is due to the fact that if multiple copies of the same vacuum subdiagram are present in a Feynman diagram, one must divide by the number of possible rearrangements of these identical diagrams. Thus,

$$\begin{aligned}\langle\Omega|\, T\left[U_I(\infty,-\infty)\phi_I(x)\ldots\right]|\Omega\rangle = \;& \exp\Big(\sum \text{Vacuum Diagrams}\Big) \\ & \times \langle\Omega|\, T\left[U_I(\infty,-\infty)\phi_I(x)\ldots\right]|\Omega\rangle_{\rm C}\,.\end{aligned} \tag{4.5.8}$$

As was shown in Eq. (4.5.7), the leading order vacuum diagram is purely imaginary. Therefore, the disconnected vacuum diagrams only contribute an overall phase at leading order, and thus one only needs to consider diagrams connected to the field sources.

From a thermodynamic perspective, the failure of disconnected vacuum diagrams to contribute to the work distribution function is expected. Disconnected vacuum diagrams by definition cannot involve field sources of ϕ and thus cannot involve the transfer of energy into or out of the system.

4.6 Work in Quantum Field Theories

Trivial and Non-Trivial Scattering As seen in Sec. 4.4 with details provided in Appendix 4.A, the finite-time transition probability can be calculated from an infinite-time scattering process. From (4.4.3) it can be shown that this requires the evaluation of an n-point correlation function in the Interaction picture. This can naturally be done by perturbatively expanding the time evolution operator in terms of the Dyson series and subsequently applying Wick's theorem to evaluate the resulting free-field correlation functions. At leading order in perturbation theory we have

$$U_I\left(\infty,-\infty\right) = \mathcal{T}_{>}\left[\exp\left(-i\int_{-\infty}^{\infty} H_{\rm int}(t)\,dt\right)\right] \approx \mathbb{1} - i\int_{-\infty}^{\infty} dt\; H_{\rm int}(t). \tag{4.6.1}$$

In this expression $H_{\rm int}(t)$ is the interaction Hamiltonian in the Interaction picture. Explicitly, it is given by

$$H_{\rm int} = \int d^3x\left[\,\frac{g}{4!M}\chi_{\rm cl}\left(x\right)\phi_I^4\left(x\right) + c_{1,0}\frac{g}{M}\chi_{\rm cl}\left(x\right) + c_{1,2}\frac{g}{M}\chi_{\rm cl}\left(x\right)\phi_I^2\left(x\right)\right]. \tag{4.6.2}$$

Using Eq. (4.6.1) in the scattering amplitude (4.4.3) schematically yields an expression of the form $|\langle \text{out}|\ \text{in}\rangle|^2 + |\langle \text{out}|\ H\ |\text{in}\rangle|^2$. This is the sum of two terms with distinct physical origins: The first term, $|\langle \text{out}|\ \text{in}\rangle|^2$, is the scattering amplitude for the trivial process where the perturbation does not enter and no work is performed on the system. As no work is performed, this will contribute a delta-function to the work distribution. The second term, $|\langle \text{out}|\ H\ |\text{in}\rangle|^2$, involves non-trivial scattering through the time-dependent perturbation. The combined probability distributions of these two processes, however, will not integrate to unity. This is a consequence of working at finite order in the Dyson series, (4.6.1). The approximation violates unitarity, which generally has to be imposed by hand, see for instance Ref. [224]. That unitarity does not hold can be seen as a consequence of the optical theorem [4, 5]. The loop corrections to the propagator that enter at order g^2 were not calculated, but are needed for unitarity to hold if we consider scattering processes of order g.

In the present case, however, it is possible to sidestep this issue since the Jarzynski equality holds separately for the non-trivial component of the work distribution, $\rho(W)$. The full work distribution has the general form $\mathcal{P}(W) = a\,\delta(W) + \rho(W)$ where $a = 1 - \int dW\,\rho(W)$ is a positive constant chosen to impose unitarity. This mirrors the expected contribution from the neglected loop diagrams required for unitarity by the optical theorem. Given the restrictions on the interaction Hamiltonian imposed by (4.3.1), the system starts and ends as a free field theory and thus $\Delta F = 0$. From the Jarzynski equality, $1 = \int dW\,\mathcal{P}\,(W) \exp\left(-\beta W\right)$, we can write,

$$\begin{aligned} 1 - a &= \int dW\,\rho(W)\ \exp\left(-\beta W\right) \\ \Rightarrow \int dW\,\rho(W) &= \int dW\,\rho(W)\ \exp\left(-\beta W\right) \end{aligned} \tag{4.6.3}$$

In conclusion, the Jarzynski equality (4.1.2) holds for the normalized, non-trivial part of the work distribution. Therefore, as the trivial component of the scattering process only contributes a delta-function to the work distribution and does not impact the Jarzynski equality, it suffices to consider the non-trivial part of $\mathcal{P}(W)$.

Calculational Approach Several complications arise in the treatment of the non-trivial scattering term. The interaction Hamiltonian (4.6.2) is composed of terms which involve the scattering of at most four incoming or outgoing particles. As the general work distribution function, (4.4.2), involves any number of incoming or outgoing particles, the Feynman diagrams which describe these processes will be composed of several disconnected subdiagrams. One must sum over all possible permutations of these subdiagrams before squaring the resulting amplitude. This is in stark contrast to the usual procedure in quantum field theory where one is only interested in fully connected dia-

grams and their permutations. To further complicate the matter, even once one has the square of the amplitude of all permutations, one still must integrate over all momenta and sum over all possible particle numbers as proscribed in (4.4.2). Carrying out this procedure in generality proves a formidable challenge to a direct application of existing field theoretic techniques.

In this work, we instead pioneer a graph theoretic approach which allows us to classify the products of Feynman diagrams in such a manner that the infinite sums over particle number can be carried out exactly. This leads to closed form expressions for the leading order work distribution where only a few kinematic integrals must be performed. The details of this procedure are in Appendix 4.B, but a brief description is provided here. While $|\langle \text{out}| H |\text{in}\rangle|^2$ can be thought of as the square of the sum of all permutated diagrams, it will be more helpful to think of it in terms of the sum over the crossterms of two permutated diagrams. The incoming and outgoing field sources of each diagram are labeled by integers up to n_1 and n_2, respectively. One proceeds to "glue" the two diagrams together by identifying the corresponding field sources in each diagram. The resulting "glued" diagram can then be classified in terms of its graph topology, specifically the topology of the connected subgraph(s) that contain insertions of the interaction Hamiltonian. Rephrased in this language, the combinatorics of the sum over permutations and subsequent sum over particle number becomes tractable.

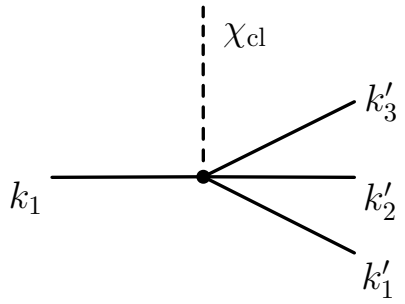
Ultimately, one finds that the work distribution function is naturally written as the sum of five distributions: the work distribution for when the particle number is unchanged, the distributions for when the particle number increases, or decreases by two, and the distributions for when the particle number increases, or decreases by four. These are denoted by the distributions $\rho_{n\to n}(W)$, $\rho_{n\to n\pm 2}(W)$, and $\rho_{n\to n\pm 4}(W)$, respectively. It should be stressed that the subscript n in these distributions does not correspond to a specific particle number as the particle number has been summed over. Closed form, unnormalized, expressions for these five distributions are given in Appendix 4.C.

4.7 Analytic Properties

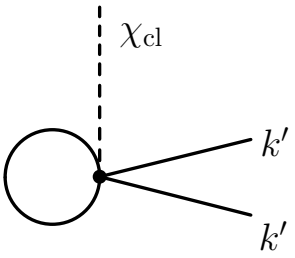
Form of the Work Distributions As an example for the five contributions to $\mathcal{P}(W)$, we discuss $\rho_{n\to n+2}(W)$ in detail as it illustrates all key properties. This distribution may be decomposed into two components:

$$\rho_{n\to n+2}(W) = \rho^{\text{tree}}_{n\to n+2}(W) + \rho^{\text{loop}}_{n\to n+2}(W), \tag{4.7.1}$$

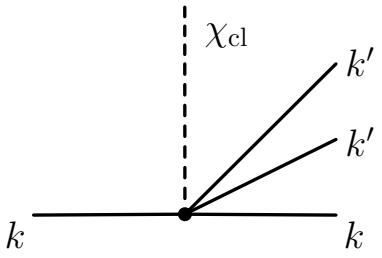
where $\rho^{\text{tree}}_{n\to n+2}(W)$ is the distribution of work arising from tree-level processes and $\rho^{\text{loop}}_{n\to n+2}(W)$ originates from diagrams involving a loop. In a loose sense, $\rho^{\text{tree}}_{n\to n+2}(W)$



(a) Tree diagram contribution to the work distribution.



(b) Loop diagram contribution to the work distribution.



(c) Tree-level diagram which interferes with the loop diagram.

Figure 4.3. Diagrams which contribute to the work distribution function $\rho_{n\to n+2}(W)$.

can be thought of as "classical" contributions to the work distribution as tree-level diagrams satisfy the classical equations of motion. The distribution $\rho^{\rm loop}_{n\to n+2}(W)$ corresponds to processes which violate the classical equations of motion and are purely a result of second quantization. Only the distributions $\rho_{n\to n+2}(W)$ and $\rho_{n\to n-2}(W)$ have contributions from loop diagrams at this order.

We begin with the tree-level contribution. While both the incoming and outgoing states will potentially involve many particles, the relevant subdiagram generated by the interaction Hamiltonian is shown in Fig. 4.3a. This is simply the tree-level process where one particle becomes three. The distribution of resulting work done on the

system is given by [3],

$$\begin{aligned}\rho^{\text{tree}}_{n\to n+2}(W) =& \frac{V}{3!}\left|\int dt\,\lambda(t)\exp(iWt)\right|^2 \int \widetilde{d^3k_1}\,\widetilde{d^3k_1'}\,\widetilde{d^3k_2'}\,\widetilde{d^3k_3'}\,\delta(W+\omega_1-\omega_1'-\omega_2'-\omega_3') \\ &\times (2\pi)^3\,\delta^3(k_1-k_1'-k_2'-k_3')\,\left(\frac{1}{\exp(\beta\omega_1)-1}\right) \\ &\times\left(1+\frac{1}{\exp(\beta\omega_1')-1}\right)\left(1+\frac{1}{\exp(\beta\omega_2')-1}\right)\left(1+\frac{1}{\exp(\beta\omega_3')-1}\right).\end{aligned} \tag{4.7.2}$$

While this expression appears rather involved, each factor has a clear physical interpretation. The combinatorial factor of 3! accounts for the symmetry of the three identical outgoing particles. The probability of doing a particular amount of work scales with volume of the system; the implications of this will be discussed shortly. The magnitude squared of the Fourier transform of the time-dependent coupling is the spectral density and can be thought of as a measure of how much the system is being driven in energy (frequency) space.

Finally, we have a kinematic integral which is a function of the work performed. The integration measure is the Lorentz invariant momentum measure $\widetilde{d^3k}$ for each incoming and outgoing particle. The two sets of delta-functions impose conservation of energy and momentum including the contributions of the time-dependent background. The incoming particle is associated with the Bose-Einstein statistics factor, $1/(\exp(\beta\omega)-1)$. This is the density of states for a thermal system of bosons which should be expected because the system was initially prepared in thermal state. The outgoing particles, however, are associated with the unusual factor $1+1/(\exp(\beta\omega)-1)$. This is the appropriate density of states because the original occupancy number for a given energy level is just $1/(\exp(\beta\omega)-1)$, but due to the scattering process the occupancy of this level must go up by one.

These observations can be generalized to a set of rules for constructing any of the tree-level work distributions. One associates each incoming particle with the density of states $1/(\exp(\beta\omega)-1)$ and each outgoing particle with $1+1/(\exp(\beta\omega)-1)$. One then integrates over these kinematic factors in a Lorentz invariant manner and includes delta-functions for conserving energy and momentum. This is multiplied by the spectral density of the driving protocol and a factor of the volume. Appropriate symmetry factors for the incoming and outgoing particles are then included. In principle, one could arrive at these rules from a simple thermodynamic treatment of the density of states and subsequent use of classical field theory. We stress that this is not the approach that

[3] For convenience, we have chosen to work in terms of $\lambda(t)$ rather than the mathematically equivalent $g/M\,\chi_{\text{cl}}(t)$.

we used and that these expressions for the work distributions were derived by summing an infinite collection of Feynman diagrams in a fully quantum treatment.

As mentioned earlier, the work distribution function is proportional to the volume of the system. This leads to restrictions on the applicability of the work distributions in Appendix 4.C to systems with large volume. As explained in Sec. 4.6, unitary is not manifest at finite order in the Dyson series. This was sidestepped by noting that the Jarzynski equality still held for just the non-trivial component of the work distribution alone. However, it was assumed that the total work distribution function could be expressed as $\mathcal{P}(W) = a\,\delta(W) + \rho(W)$ where $a = 1 - \int dW\,\rho(W)$ is a positive constant and $\rho(W)$ is the non-trivial part of the work distribution. Since $\rho(W)$ is proportional to the volume, a will become negative for large systems. At this point, our leading order approximation is no longer valid. Therefore, the range of validity of the present treatment is $\int dW\,\rho(W) < 1$. It may be possible, however, to extend the range of validity by working to higher orders in perturbation theory.

We now turn our attention to the component of the work distribution function which arises from loop diagrams,

$$\rho^{\text{loop}}_{n\to n+2}(W) = \frac{V}{2}\left|\int dt\,\lambda(t)\exp(iWt)\right|^2\left(1+\frac{1}{\exp(\beta W/2)-1}\right)^2\frac{1}{W}\left(\int \widetilde{d^3k}\,\delta(W-2\omega)\right)$$
$$\times\left(\int \widetilde{d^3k}\,\frac{1}{\exp(\beta\omega)-1}+\frac{1}{2}\left(\frac{m}{4\pi}\right)^2\left[1+\log\left(\frac{\mu^2}{m^2}\right)\right]\right)^2. \tag{4.7.3}$$

In this expression, μ is the renormalization scale in the $\overline{MS}$ scheme, see *e.g.* Refs. [4–6]. Once again, we see that the work distribution is proportional to the volume and spectral density of the driving. In this process, there are two outgoing particles, each carrying half of the work put into the system which is reflected in $(1 + 1/(\exp(\beta W/2) - 1))^2$. The next two terms in (4.7.3) are a measure of the phase space available to the outgoing particles. It should be noted that because the particles have non-zero mass, $W \geq 2m$, one does not need to worry about the singular behavior of $1/W$.

The final term in (4.7.3) results from the interference of two Feynman diagrams. The first diagram, shown in Fig. 4.3b, is the one-loop process by which two particles can be created. This one-loop diagram, however, interferes with the tree-level diagram given in Fig. 4.3c. This tree-level process involves the production of two particles where the initial particle is merely a spectator and experiences no change in energy. The renormalization parameter μ then controls the relative size of the contribution from each diagram.

Loop diagrams do not exist in classical field theory and are the hallmark of second quantization. In a classical field theory Eq. (4.7.3) would vanish and thus it may be

thought of as the change in the work distribution function due to second quantization. Note, however, that $\mathcal{P}(W)$ may be dominated by these contributions, as we will see in Sec. 4.8.

As the work distribution (4.7.3) explicitly depends on the renormalization scale μ, this raises the question whether the work distribution is an observable quantity in quantum field theory. To be a physical, observable quantity, the work distribution should be independent of the renormalization scale and remain invariant under renormalization group flow. Beyond the explicit dependence on μ, the coupling constant and mass have implicit dependence on μ due to renormalization. Using the β-function for this theory, it can be shown that the running of the work distribution enters at $\mathcal{O}(\lambda^3)$. This is a higher order effect and may be modified by terms beyond leading order. To leading order the work distribution does not depend on the renormalization scale and therefore we conclude that the work distribution is, indeed, a physical observable. This can also be seen as a consequence of the form of Eq. (4.4.2). The work distribution function can be constructed from knowledge of the scattering amplitudes, which are observables of the field theory. Therefore, the work distribution must also be an observable.

Fluctuation Theorems We now investigate how fluctuation theorems manifest in a quantum field theory. Throughout this section, we will always refer to the work distribution functions which are normalized such that the total work distribution integrates to unity. To make this normalization clear, we will utilize P instead of ρ.

We first consider the Crooks fluctuation theorem [41, 49]. Assuming no change in free energy, the Crooks fluctuation theorem states that the probability distribution for a forward process, $P_{A\to B}(W)$, is related to the distribution for the reversed process, $P_{B\to A}(-W)$, through

$$\frac{P_{B\to A}(-W)}{P_{A\to B}(W)} = \exp(-\beta W). \tag{4.7.4}$$

Accordingly, for the time-dependent field theory we have

$$\frac{P_{n\to n-4}(-W)}{P_{n\to n+4}(W)} = \frac{P_{n\to n-2}(-W)}{P_{n\to n+2}(W)} = \frac{P_{n\to n}(-W)}{P_{n\to n}(W)} = \exp(-\beta W). \tag{4.7.5}$$

Using the explicit form of the work distribution functions from Appendix 4.C, we verify (4.7.5) analytically. Again, this holds independently of the renormalization scale, μ. The key property of the work distribution functions which allows for a proof of Eq. (4.7.5) is that each incoming state is associated with the factor $1/(\exp(\beta\omega)-1)$ while each outgoing state is associated with $1+1/(\exp(\beta\omega)-1) = \exp(\beta\omega)/(\exp(\beta\omega)-1)$. The latter is nothing else but an expression of local detailed balance.

More surprisingly, it can be shown that, independent of renormalization scale,

$$\frac{P^{\text{tree}}_{n\to n-2}(-W)}{P^{\text{tree}}_{n\to n+2}(W)} = \frac{P^{\text{loop}}_{n\to n-2}(-W)}{P^{\text{loop}}_{n\to n+2}(W)} = \exp\left(-\beta W\right). \tag{4.7.6}$$

Thus, the Crooks fluctuation theorem holds both with, and without, the contribution from loop diagrams. Without loop diagrams, one simply has a classical field theory and the the validity of the fluctuation theorem is well established for classical systems. Moving to a quantum field theory, loop diagrams must be included, but the Crooks fluctuation theorem still holds! This requires that order by order loop corrections must enter in a pairwise manner such that the fluctuation theorem holds at every order. That the Crooks fluctuation theorem still holds once loop corrections are introduced is a consequence of the fact that the theorem only depends on unitarity and local detailed balance.

As the Crooks fluctuation theorem has been verified for our time-dependent field theory, the Jarzynski equality immediately follows as a consequence. This can quickly be shown through

$$\int dW\, \mathcal{P}(W)\, \exp\left(-\beta W\right) = \int dW\, \mathcal{P}(-W) = 1. \tag{4.7.7}$$

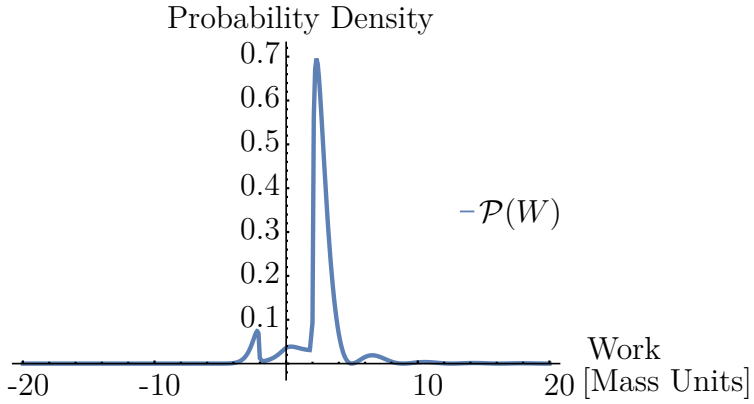
As was true for the Crooks fluctuation theorem, the Jarzynski equality holds independent of the renormalization scale and will hold with or without the loop contributions.

In conclusion, we have analytically verified that the Crooks fluctuation theorem and Jarzynski equality hold independent of renormalization scale for a time-dependent quantum field theory at leading order. However, the quantum Jarzynski equality made no assumptions of perturbativity and only required the mild assumption of unital dynamics in Eqs. (4.2.6) and (4.2.7). Thus, while not verified analytically, these fluctuation theorems should hold to any order perturbatively and may even hold non-perturbatively.

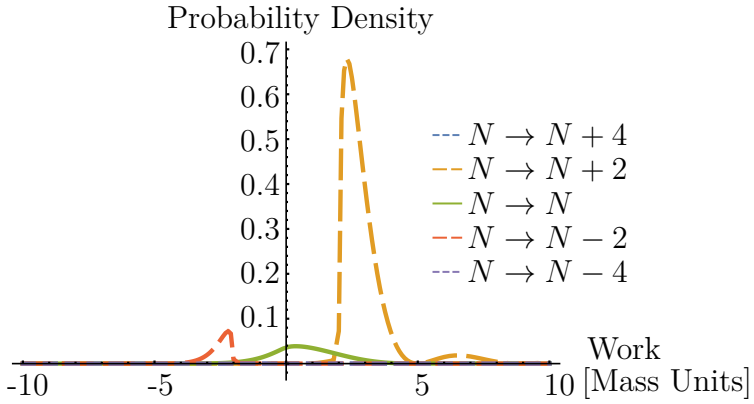
4.8 Example: numerical case study

We conclude the analysis with a numerical case study. Throughout this section, we will work in units such that $m = 1$. To accentuate the contributions of particle creation and annihilation we will work with a relativistic bath and driving protocol. We will assume that the bath has inverse temperature comparable to the particle mass, $\beta = 1$. As driving protocol, we will consider the infinitely smooth but non-analytic "bump" function,

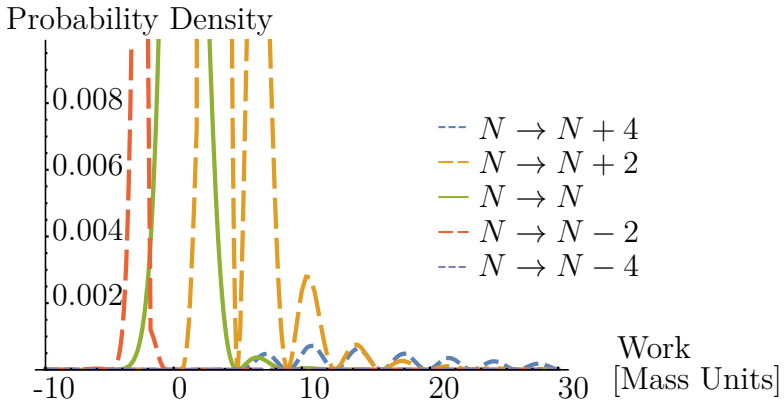
$$\lambda(t) = \begin{cases} \lambda_0 \, \exp\left(\frac{-t^2}{1-t^2}\right), & \text{for } |t| \leq 1 \\ 0, & \text{otherwise} \end{cases}. \tag{4.8.1}$$



(a) Normalized work distribution function.



(b) Work distribution functions for subprocesses.



(c) Zoom-in of Fig. 4.4b focusing on the tail of the work distributions.

Figure 4.4. Work distribution function and its decomposition into subprocesses for $\beta = m = 1$ with driving protocol specified in (4.8.1).

This function is chosen to avoid any potential issues with continuity of derivatives at the start and end of the protocol. The overall scale of the driving protocol, λ_0, will ultimately drop out when the work distribution function is normalized. We only require that λ_0 is sufficiently small that the theory is perturbative and our expressions for the work distribution are valid.

We numerically evaluate the work distribution functions of Appendix 4.C and subsequently normalize the combined $\mathcal{P}(W)$. This yields the total work distribution of Fig. 4.4a and the work distributions for the various subprocesses shown in Fig. 4.4b. Both Fig. 4.4a and Fig. 4.4b show the characteristic "exponential asymmetry" which is indicative of the Jarzynski equality. It is found that both the quantum Jarzynski equality and Crooks fluctuation theorem hold to within numerical precision. The dominant contribution to the work distribution function is from $P_{n\to n+2}(W)$ with over 80% of trials resulting in particle pair production. Surprisingly, within $P_{n\to n+2}(W)$, over 95% of the distribution is from the loop diagram contributions, $P^{\text{loop}}_{n\to n+2}(W)$. Thus, for this protocol and bath, the majority of the work distribution comes from loop diagrams which pair produce particles, an effect which only exists in a quantum field theory. Not including these loop diagrams would produce a markedly different $\mathcal{P}(W)$.

Figure 4.4c gives a zoomed in view of the tail of the work distribution functions. It can immediately be seen that all of the work distribution functions experience the same type of oscillatory behavior. This is a result of the spectral density of the time-dependent coupling vanishing at these energies (frequencies) and is not kinematic in origin. Figure 4.4c also shows that $P_{n\to n+4}(W)$ is dominant over $P_{n\to n+2}(W)$ but only at large values of work ($W \gtrsim 20$ mass units).

4.9 Concluding remarks

Two decades ago, the Jarzynski equality was formulated for classical systems. It required another decade to formalize the concept of work in the quantum regime. However, quantum mechanics is not our most complete description of nature. This is quantum field theory, and another decade later quantum fluctuation theorems have now been extended to their ultimate limit.

For the sake of accessibility and specificity, we considered a time-dependent variant of $\lambda\phi^4$, for which we found closed form expressions for the work distribution functions. While this is only one particular quantum field theory, these distribution functions demonstrate a variety of features which should be anticipated in other field theories. It should be emphasized that our work is more general than a mere case study, since the developed methodology can be directly applied to other field theories.

In the present case, it was shown that the work distribution functions do not run with the renormalization scale through the perturbative order considered, implying that these distributions are physical observables of the field theory. It was also found that both the Crooks fluctuation theorem and Jarzynski equality hold independent of the renormalization scale, both with and without loop corrections. Remarkably, the contribution of loop diagrams to the work distribution function can be substantial and thus essential in the proper description of work fluctuations of quantum field theories in the relativistic regime.

Until now, particle pair-production and loop effects had not been incorporated in any study of quantum fluctuation theorems. These effects become dominant in the relativistic regime. Our results were presented in units of the particle mass which can obscure physical intuition for the system and protocol being considered. To get better insight, consider a hypothetical condensed matter system that is described by the time-dependent $\lambda\phi^4$ theory with an effective mass $m \sim 1\,\mathrm{eV}$. Our work distributions describe the behavior of such a system with an effective temperature of $T \sim 10^6\,\mathrm{K}$ with a driving time of $\Delta t = 10^{-15}\,\mathrm{s}$. For a particle with mass comparable to the electron, $m \sim 1\,\mathrm{MeV}$, this corresponds to temperatures of $T \sim 10^{12}\,\mathrm{K}$ and driving times of $\Delta t = 10^{-21}\,\mathrm{s}$; conditions relevant for the study of quark-gluon plasma. This is well beyond the original regime in which the fluctuation theorems were conceived and outside previous treatments in the literature.

It should be stressed that these results were directly calculated from in-out scattering amplitudes which are the natural building block for the quantum Jarzynski equality. In particular, we did not need to utilize the Schwinger-Keldysh in-in formalism at any point in the calculation. This was only possible because we were able to find a mapping between finite-time transition amplitudes and infinite-time amplitudes, and demonstrated that disconnected vacuum diagrams did not alter the scattering amplitude. Due to the form of the work distribution function a new diagrammatic technique had to be developed so that the infinite sum over particle number and sum over permutations of disconnected Feynman diagrams can be performed analytically. This technique relies on the topological properties of "glued Feynman diagrams" to classify permutations and enables the rephrasing of the sum over particle number in terms of a sum over graph theoretic properties of the glued diagrams.

In the present work, field theoretic calculations were carried out using in-out scattering amplitudes; however, future studies may wish to make use other approaches to non-equilibrium quantum field theory such as the Schwinger-Keldysh in-in formalism. The greatest strength of the Schwinger-Keldysh formalism is that it has been extensively studied, and the technical machinery for dealing with renormalization of time-dependent fields and unitarity is well understood, see *e.g.* Refs. [225, 226]. Its

shortcoming in the present context is that the Two-Time Measurement Formalism is not amenable to an in-in interpretation. The trace form of the work distribution (4.2.6), appears to be a natural fit for the Schwinger-Keldysh approach, however, the energy projection operators generally can not be expressed in terms of field valued operators. It still may be possible to address this with a method similar to what is used in determining the full counting statistics in the Schwinger-Keldysh formalism, but further research would be required in this direction. An alternative would be to rewrite the work distribution function in terms of scattering amplitudes as done in (4.4.2), but these scattering amplitudes are explicitly in-out objects. This would require one to find the correlation functions in the Schwinger-Keldysh formalism and then calculate the in-out scattering amplitudes from them. Given these limitations, a fully in-out approach was preferrable for our study despite the technical difficulties addressed in this paper. However, the Schwinger-Keldysh approach could be useful in future investigations, particularly if (4.2.6) can be recast in a form compatible with the in-in formalism.

While this work focused primarily on time-dependent $\lambda\phi^4$, these techniques should be applicable to any quantum field theory. Even with the restriction that the system begins and ends as a free field theory, this vastly expands the realm of applicability for quantum fluctuation theorems. One example of an interesting system which fits within this paradigm is a cyclic engine acting on a quantum field working medium. Previous investigations [227] calculated the average behavior of such an engine, however techniques outlined in this work should allow for a full treatement of the work fluctuations.

This opens new frontiers for the use of fluctuation theorems; from the relativistic charge carriers of graphene to the quark gluon plasma produced in heavy ion collidors to the evolution of the early universe, fluctuation theorems can provide insight into the short timescale behavior of nonequilibrium systems. While we may now begin applying quantum fluctuation theorems to the most extreme conditions found in nature, there is still more progress to be made. The most immediate challenge is to generalize the approach presented here to make it applicable to a wider variety of quantum field theories. This would enable the study of gauge fields and more interesting protocols. Twenty years after the advent of the Jarzynski equality, and ten years after its quantum equivalent, fluctuation theorems can finally be applied across the full range of energy and length scales understood in modern physics, but more work is still left to be done.

4.A Finite-time Amplitude to Infinite-time Amplitude Mapping

In calculating the work distribution function (4.4.2) for a quantum field theory, one must address the finite-time transition amplitude $\left\langle k'_1, \ldots, k'_{n_2}; t_2 \right| U(t_2, t_1) \left| k_1, \ldots, k_{n_1}; t_1 \right\rangle$. This finite-time amplitude must be rewritten in terms of an infinite-time scattering process so that the full machinery of quantum field theory can be used.

In general, the initial and final states of the system will be multiparticle states. We will specialize to the case of single particle states for notational simplicity; the generalization to multiparticle states is straightforward.

To distinguish operators in different quantum mechanical pictures, all states and operators in the Schrödinger picture will be denoted by a subscript S and those in the Interaction picture will have a subscript I. We start by defining the initial and final states in terms of the creation and annihilation operators as $|k; t_1\rangle_S = a_S^\dagger(k) \left|\Phi_{\text{in}}; t_1\right\rangle_S$ and $|k'; t_2\rangle_S = a_S^\dagger(k') \left|\Phi_{\text{out}}; t_2\right\rangle_S$. In these expressions, $\left|\Phi_{\text{in}}; t_1\right\rangle_S$ and $\left|\Phi_{\text{out}}; t_2\right\rangle_S$ are the incoming and out-going vacuum states. These states are defined such that $H_0 \left|\Phi_{\text{in}}; t_1\right\rangle_S = H_0 \left|\Phi_{\text{out}}; t_2\right\rangle_S = 0$. Using these definitions,

$$ {}_S\left\langle k'; t_2\right| U(t_2, t_1) \left|k; t_1\right\rangle_S = {}_S\left\langle \Phi_{\text{out}}; t_2\right| a_S(k') U(t_2, t_1) a_S^\dagger(k) \left|\Phi_{\text{in}}; t_1\right\rangle_S . \tag{4.A.1} $$

We now define a time $\tau \gg \max\left(\left|t_1\right|, \left|t_2\right|\right)$ with the intention of eventually taking the limit $\tau \to \infty$. As the system is assumed to be free at times t_1 and t_2, one may trivially extend the finite-time experiment by assuming that the system remains free outside of the interval $t \in (t_1, t_2)$. This cannot change the work distribution function as no work is performed keeping the Hamiltonian fixed. Then,

$$ {}_S\left\langle k'; t_2\right| U(t_2, t_1) \left|k; t_1\right\rangle_S = {}_S\left\langle \Phi_{\text{out}}; \tau\right| U(\tau, t_2) a_S(k') U(t_2, t_1) a_S^\dagger(k) U(t_1, -\tau) \left|\Phi_{\text{in}}; -\tau\right\rangle_S . \tag{4.A.2} $$

Note, as $H(t) = H_0$ for $t \notin (t_1, t_2)$, it is still true that $H_0 \left|\Phi_{\text{in}}; -\tau\right\rangle_S = H_0 \left|\Phi_{\text{out}}; \tau\right\rangle_S = 0$.

We now define some reference time $t_0 \notin (t_1, t_2)$ when the Schrödinger and Interaction pictures coincide. Passing to the Interaction picture,

$$ \begin{aligned} {}_S\left\langle k'; t_2\right| U(t_2, t_1) \left|k; t_1\right\rangle_S = {}_S\left\langle \Phi_{\text{out}}; \tau\right| U_0(\tau, t_0) U_I(\tau, t_2) a_I(k'; t_2) U_I(t_2, t_1) \\ \times a_I^\dagger(k; t_1) U_I(t_1, -\tau) U_0(t_0, -\tau) \left|\Phi_{\text{in}}; -\tau\right\rangle_S . \end{aligned} \tag{4.A.3} $$

In this expression, U_0 is the time evolution operator under the free Hamiltonian while U_I is the evolution operator in the Interaction picture. As $\left|\Phi_{\text{out}}; \tau\right\rangle_S$ and $\left|\Phi_{\text{in}}; -\tau\right\rangle_S$ are vacuum states of the free theory, ${}_S\left\langle \Phi_{\text{out}}; \tau\right| U_0(\tau, t_0) = {}_S\left\langle \Phi_{\text{out}}; \tau\right|$ and $U_0(t_0, -\tau) \left|\Phi_{\text{in}}; -\tau\right\rangle_S =$

$|\Phi_{\text{in}}; -\tau\rangle_S$. Furthermore, as both states are annihilated by the free Hamiltonian and the ground state is unique, they may differ by at most a phase from the ground state $|\Omega; t_0\rangle_I$. As the scattering amplitude will be squared in the final calculation, these phase factors are irrelevant. Rewriting the scattering amplitude as a time-ordered product,

$$|{}_S\langle k'; t_2| U(t_2, t_1) |k; t_1\rangle_S| = \left|{}_I\langle\Omega| T\left[U_I(\tau, -\tau) a_I(k'; t_2) a_I^\dagger(k; t_1)\right] |\Omega\rangle_I\right|. \tag{4.A.4}$$

Using the mode-expansion of the free scalar field and the definition of time-evolution for operators in the Interaction picture, it is straightforward to show,

$$\exp(i\omega t) a_I(k; t) = i \int d^3x \ \exp(-ikx) \overset{\leftrightarrow}{\partial_0} \phi_I(x), \tag{4.A.5}$$

$$\exp(-i\omega t) a_I^\dagger(k; t) = -i \int d^3x \ \exp(ikx) \overset{\leftrightarrow}{\partial_0} \phi_I(x). \tag{4.A.6}$$

In these relations, the operator $\overset{\leftrightarrow}{\partial_\mu}$ is defined such that $f \overset{\leftrightarrow}{\partial_\mu} g = f(\partial_\mu g) - (\partial_\mu f) g$, see Ref. [4].

Making use of (4.A.5) and (4.A.6), it is possible to rewrite Eq. (4.A.4) purely in terms of field operators in the Interaction picture. As noted before, we are only interested in the magnitude of Eq. (4.A.4) as any overall phase disappears in Eq. (4.4.2). Thus, up to an overall irrelevant phase, we find

$$\begin{aligned} |{}_S\langle k'; t_2| U(t_2, t_1) |k; t_1\rangle_S| = \Bigg| \int d^3x' d^3x \ \exp(-ik'x') \exp(ikx) \\ \times \overset{\leftrightarrow}{\partial_{0_{x'}}} \overset{\leftrightarrow}{\partial_{0_x}} {}_I\langle\Omega| T\left[U_I(\tau, -\tau) \phi_I(x') \phi_I(x)\right] |\Omega\rangle_I \Bigg|. \end{aligned} \tag{4.A.7}$$

In this expression, it is understood that the time components of the four-vectors x and x' are to be evaluated at t_1 and t_2 respectively.

We may now formally take the limit as $\tau \to \infty$. Generalizing to the case of multiparticle initial and final states,

$$\begin{aligned} \left|\left\langle k'_1, \ldots, k'_{n_2}\right| U(t_2, t_1) \left|k_1, \ldots, k_{n_1}\right\rangle\right| = \Bigg| \int d^3x'_1 d^3x_1 \ldots \\ \times \ \exp(-ik'_1 x'_1) \exp(ik_1 x_1) \ldots \overset{\leftrightarrow}{\partial_{0_{x'_1}}} \overset{\leftrightarrow}{\partial_{0_{x_1}}} \ldots \\ \times \ {}_I\langle\Omega| T\left[U_I(\infty, -\infty) \phi_I(x'_1) \ldots \phi_I(x_1) \ldots\right] |\Omega\rangle_I \Bigg|. \end{aligned} \tag{4.A.8}$$

4.B Diagrammatic Technique

As mentioned in Sec. 4.6, when calculating the work distribution function from the Dyson series one runs into technical difficulties. The interaction Hamiltonian (4.6.2) is composed of terms which involve at most four field sources while the general work distribution function (4.4.2) involves any number of incoming or outgoing particles. As a result, the Feynman diagrams which describe these processes will be composed of several disconnected subdiagrams. One must sum over all possible permutations of these subdiagrams before squaring the resulting amplitude; unlike the usual procedure in quantum field theory where one is only interested in fully connected diagrams and their permutations.

To motivate our procedure for calculating this sum, it will be necessary to introduce notation for describing the permutations of Feynman diagrams. From Eq. (4.4.2) it can be seen that the scattering amplitude will depend on the momenta of the incoming and outgoing particles and all of these momenta are integrated over in a Lorentz invariant manner. Let K denote the collection of all momenta and let dK denote the Lorentz invariant measure. We will let $\mathcal{D}$ correspond to a Feynman diagram of interest, such as the one shown in Fig. 4.5a. Let $\mathcal{S}$ be the set of all permutations of the Feynman diagram which do not interchange incoming for outgoing particles and let $\sigma \in \mathcal{S}$ be a particular permutation. Now define $f(W, K)$ to be all the terms that appear in Eq. (4.4.2) that are not the scattering amplitude, *i.e.* the energy conserving delta-function, Boltzmann factors, and normalization constant. Then, Eq. (4.4.2) can schematically be rewritten as

$$P(W) = \int dK\, f(W,K) \cdot \left| \sum_{\sigma\in\mathcal{S}} (\sigma \circ \mathcal{D})(K) \right|^2 \tag{4.B.1}$$

$$= \sum_{\sigma_1,\sigma_2\in\mathcal{S}} \int dK\, f(W,K) \cdot (\sigma_1 \circ \mathcal{D})^\dagger (K) \cdot (\sigma_2 \circ \mathcal{D})(K) \tag{4.B.2}$$

$$= \sum_{\sigma_1,\sigma_2\in\mathcal{S}} \int dK\, f(W,K) \cdot \mathcal{D}^\dagger(K) \cdot \left(\sigma_2 \circ \sigma_1^{-1} \circ \mathcal{D}\right)(K) \tag{4.B.3}$$

$$= |\mathcal{S}| \sum_{\sigma\in\mathcal{S}} \int dK\, f(W,K) \cdot \mathcal{D}^\dagger(K) \cdot (\sigma \circ \mathcal{D})(K)\,, \tag{4.B.4}$$

where $|\mathcal{S}|$ is the total number of permutations of the diagram $\mathcal{D}$. In moving from Eq. (4.B.1) to Eq. (4.B.2) we have rewritten the square of the sum as the sum over crossterms. In Eq. (4.B.3) we have chosen to relabel the momenta K such that the first diagram $\mathcal{D}^\dagger$ appears unpermuted. Lastly, in Eq. (4.B.4) we have identified the composition of permutations as a permutation and performed the sum over the redundant permutation. In these expressions, we have suppressed the sum over incoming

and outgoing particle number and have ignored potential complications arising from a process mediated by more than one type of Feynman diagram. Using Eq. (4.B.4), the work distribution function can be calculated by integrating over the momenta of the product of an "unpermuted" Feynman diagram and its possible permutations.

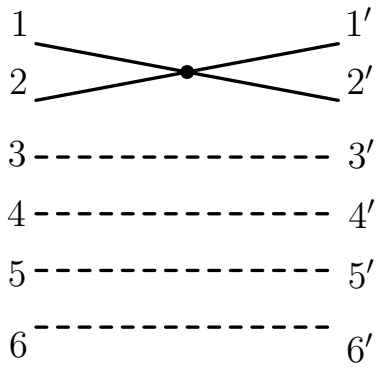
Since we are studying a variant of $\lambda\phi^4$, at leading order the particle number may either stay the same, change by two, or change by four. Diagrams with different numbers of incoming or outgoing particles do not interfere and thus can be considered separately, as mentioned in Sec. 4.6. For concreteness, we will now consider processes involving n particles where the particle number is unchanged.

For processes where the particle number is unchanged, the only Feynman diagrams that contribute at leading order are permutations of Fig. 4.5a. This particular diagram is drawn for $n = 6$ and for clarity the insertion of the background field χ_{cl} is not shown. We will choose this diagram to represent the "unpermuted" Feynman diagram $\mathcal{D}$. In principle, any other valid diagram could be chosen as the "unpermuted" reference, but Fig. 4.5a is chosen for convenience. Three possible permutations of this diagram are shown in Fig. 4.5. Note that these permutations only interchange incoming particles amongst themselves, or outgoing particles amongst themselves. It should also be noted that the exchange $1 \leftrightarrow 2$ is not considered a unique permutation because it leaves the overall diagram unchanged.

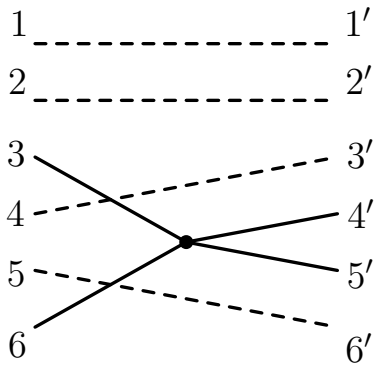
Before Eq. (4.B.4) may be utilized to calculate the work distribution function, one needs to define a scheme for enumerating possible permutations of Fig. 4.5a. It will now be demonstrated that the three permutations shown in Figs. 4.5b, 4.5c, and 4.5d define three classes of permutation which will uniquely catagorize any permutation of Fig. 4.5a.

Consider Eq. (4.B.4). One is interested in the product of two Feynman diagrams: $\mathcal{D}^\dagger(K)$, the conjugate of the unpermuted diagram, and $(\sigma \circ \mathcal{D})(K)$, a permuted Feynman diagram. In each Feynman diagram, the incoming and outgoing momenta are the same. It is only how these momenta are connected to one another through delta-functions and four-point functions which differs. As the momenta are identical, it will be helpful to define a "glued" Feynman diagram which is built from the two Feynman diagrams by treating the field sources for each diagram as identical. For example, in Figs. 4.5a and 4.5b, the momentum associated with particle 1 in each diagram is the same. As such, these diagrams can be connected by "gluing" the diagrams together at this point. Repeating this for each incoming and outgoing field source yields Fig. 4.6a. For each permuted diagram in Figure 4.5, the corresponding "glued diagram" is shown in Figure 4.6.

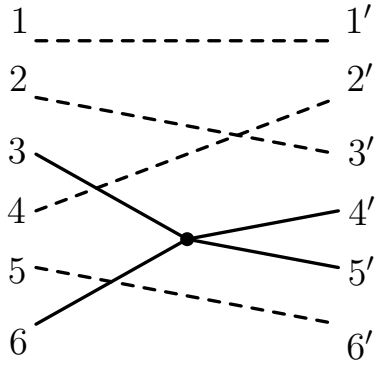
Deep properties of the permutations shown in Fig. 4.5 which are not immediately apparent in the Feynman diagrams are made manifest in the "glued diagrams" of



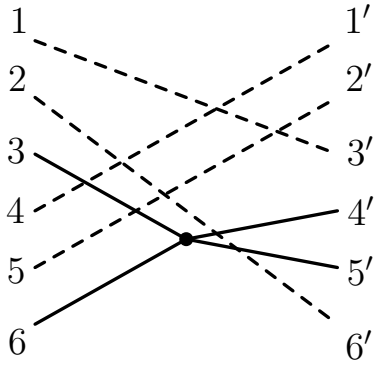
(a) The prototypical example for a Feynman diagram which contributes to $n \to n$ scattering at leading order.



(b) One type of permutation of the Feynman diagram in Fig. 4.5a.



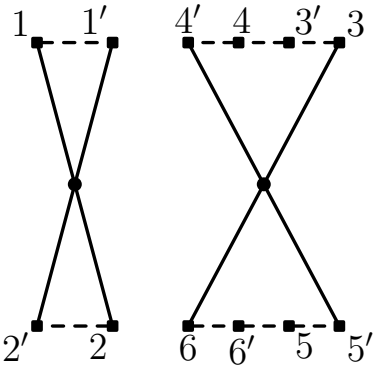
(c) A second possible permutation of the Feynman diagram in Fig. 4.5a.



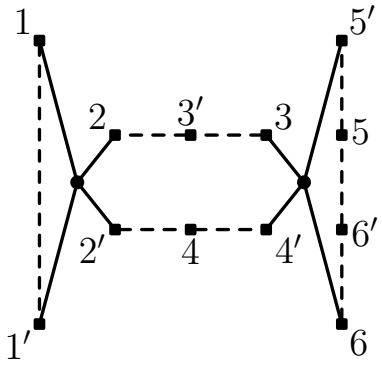
(d) A third permutation of the Feynman diagram in Fig. 4.5a.

Figure 4.5. An incomplete collection of possible Feynman diagrams which contribute to the non-trivial part of the work distribution for $n \to n$ scattering. For visual clarity, propogators which are not part of the four-point function are represented with dashed lines and insertions of the background field χ_{cl} are omitted.

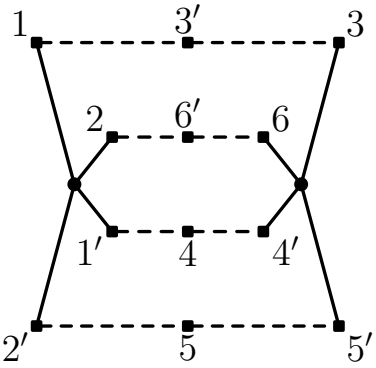
Fig. 4.6. For each type of permutation considered, the resulting graph topology in Fig. 4.6 is different. The permutation of Fig. 4.5b results in the glued diagram of Fig. 4.6a where the four-point interactions are in distinct subgraphs. This is in contrast to the permutations of Figs. 4.5c and 4.5d and their glued diagrams, Figs. 4.6b and 4.6c, which feature both four-point interactions in the same subgraph with a particular topology. In Fig. 4.6b, colloquially, two of the legs of each four-point interaction are



(a) Glued diagram resulting from the combination of Fig. 4.5a and Fig. 4.5b.



(b) Glued diagram resulting from the combination of Fig. 4.5a and Fig. 4.5c.



(c) Glued diagram resulting from the combination of Fig. 4.5a and Fig. 4.5d.

Figure 4.6. Glued diagrams generated by identifying the field sources of one Feynman diagram in Fig. 4.5 with the field sources of another and then connecting the diagrams. For visual clarity, propagators which are not part of the four-point function are represented with dashed lines and field sources are denoted by a small square. Insertions of the background field χ_{cl} are omitted.

connected to themselves resulting in "capped ends". Topologically, there exist closed cycles one can draw on the graph which pass through only one four-point function. In the case of Fig. 4.6c, each leg of one four-point interaction is connected to a leg of the other four-point interaction. This requires any closed cycle to pass through both four-point functions. The three topologies presented in Figs. 4.6a, 4.6b, and 4.6c are the only types possible for subgraphs constructed with exactly two four-point functions. As

such, we can use these three possible topologies to categorize all possible permutations of Fig. 4.5a.

While the topological approach of the "glued" diagrams has already proven useful in catagorizing permutations, its greatest value is in making the sum over particle number in Eq. (4.4.2) tractable. While left implicit in Eq. (4.B.4), the sum over particle number is troublesome because it requires one to first find a closed form expression for the $n \to n$ work distribution function in terms of the particle number n and then find a closed form expression for the infinite sum. This must be done in such a manner that the cancellation of the potentially divergent normalization factor $\mathrm{tr}\{\exp(-\beta\hat{H}_0)\} = \exp(-\beta F_0)$ is manifest. The glued diagram approach has the advantage of rephrasing the sum over particle number in terms of a sum over certain simple properties of the glued diagram.

The actual mathematical manipulations that go into the procedure are tedious and uninformative, but a high level description is provided here instead. Consider the Feynman diagram in Fig. 4.5d and the glued diagram Fig. 4.6c. For $n > 6$, Fig. 4.5d will include additional field sources and propagators. In the glued diagram, these propagators will either enter the subgraph containing the four-point functions, lengthening the paths in the graph but not changing the topology, or will create cycle graphs made entirely of propagators. With appropriate combinatorial factors, the sum over particle number can then be rephrased in terms of a sum over the length of paths in the subdiagram of the four-point functions, a sum over the number of disconnected cycles of propagators, and a sum over the length of each of these cycles. Importantly, the sum over the length of paths in the subgraph Fig. 4.6c is independent of the sums over the number and length of cycles of disconnected propagators. This sum over disconnected cycles of propagators is just the sum over all possible "trivial" scatterings where the four-point function never appears. Carrying out this sum ultimately yields the normalization factor $\mathrm{tr}\{\exp(-\beta\hat{H}_0)\}$. One may then evaluate the sum over path lengths in Fig. 4.6c by noting that each propagator is a delta-function and each incoming field source is associated with a Boltzmann weight $\exp(-\beta\omega)$. This gives a set of geometric series which can be summed into Bose-Einstein statistics factors. While not shown here, it can be demonstrated that, due to their subgraph topology, Figs. 4.6a and 4.6b are proportional to $\delta(W)$ and thus make trivial contributions to the work distribution function. While the exact diagrammatics differ, this scheme applies equally well to $n \to n \pm 2$ and $n \to n \pm 4$ processes.

4.C Work Distribution Functions

In the calculation of the work distribution function Eq. (4.4.2) for the time-dependent field theory (4.5.2), it was found that the distribution factored into five distinct parts. These correspond to realizations of the experiment where the particle number is unchanged; the particle number increases, or decreases by two; or the particle number increases, or decreases by four. These are denoted by the distributions $\rho_{n\to n}(W)$, $\rho_{n\to n\pm 2}(W)$, and $\rho_{n\to n\pm 4}(W)$, respectively. As explained in Sec. 4.6, these distributions are not normalized and this must be done by hand.

The work distribution functions for when the particle number is constant, or changes by four, are given by

$$\begin{aligned}\rho_{n\to n-4}(W) =& \frac{1}{4!}V\left|\int dt\,\lambda(t)\exp(iWt)\right|^2\int \widetilde{d^3k_1}\,\widetilde{d^3k_2}\,\widetilde{d^3k_3}\,\widetilde{d^3k_4}\\ &\times\delta(W+\omega_1+\omega_2+\omega_3+\omega_4)\,(2\pi)^3\,\delta^3(k_1+k_2+k_3+k_4)\\ &\times\left(\frac{1}{\exp(\beta\omega_1)-1}\right)\left(\frac{1}{\exp(\beta\omega_2)-1}\right)\\ &\times\left(\frac{1}{\exp(\beta\omega_3)-1}\right)\left(\frac{1}{\exp(\beta\omega_4)-1}\right),\end{aligned}\tag{4.C.1}$$

$$\begin{aligned}\rho_{n\to n}(W) =& \frac{1}{4}V\left|\int dt\,\lambda(t)\exp(iWt)\right|^2\int \widetilde{d^3k_1}\,\widetilde{d^3k_2}\,\widetilde{d^3k'_1}\,\widetilde{d^3k'_2}\\ &\times\delta(W+\omega_1+\omega_2-\omega'_1-\omega'_2)\,(2\pi)^3\,\delta^3(k_1+k_2-k'_1-k'_2)\\ &\times\left(\frac{1}{\exp(\beta\omega_1)-1}\right)\left(\frac{1}{\exp(\beta\omega_2)-1}\right)\\ &\times\left(1+\frac{1}{\exp(\beta\omega'_1)-1}\right)\left(1+\frac{1}{\exp(\beta\omega'_2)-1}\right),\end{aligned}\tag{4.C.2}$$

$$\begin{aligned}\rho_{n\to n+4}(W) =& \frac{1}{4!}V\left|\int dt\,\lambda(t)\exp(iWt)\right|^2\int \widetilde{d^3k_1}\,\widetilde{d^3k_2}\,\widetilde{d^3k_3}\,\widetilde{d^3k_4}\\ &\times\delta(W-\omega_1-\omega_2-\omega_3-\omega_4)\,(2\pi)^3\,\delta^3(k_1+k_2+k_3+k_4)\\ &\times\left(1+\frac{1}{\exp(\beta\omega_1)-1}\right)\left(1+\frac{1}{\exp(\beta\omega_2)-1}\right)\\ &\times\left(1+\frac{1}{\exp(\beta\omega_3)-1}\right)\left(1+\frac{1}{\exp(\beta\omega_4)-1}\right).\end{aligned}\tag{4.C.3}$$

In the calculation of these work distributions only tree-level diagrams enter. Thus, even though particle number is not conserved, these work distributions can be calculated from the classical equations of motion. This should be contrasted with the work distributions for when the particle number changes by two. Loop diagrams contribute

to these work distributions and their contributions can be separated out,

$$\rho_{n\to n\pm 2}\left(W\right) = \rho^{\text{tree}}_{n\to n\pm 2}\left(W\right) + \rho^{\text{loop}}_{n\to n\pm 2}\left(W\right). \tag{4.C.4}$$

The tree-level contributions to the work distributions are given by

$$\begin{aligned}
\rho^{\text{tree}}_{n\to n+2}\left(W\right) =& \frac{1}{3!}V\left|\int dt\,\lambda(t)\exp\left(iWt\right)\right|^2 \int \widetilde{d^3k_1}\,\widetilde{d^3k_1'}\,\widetilde{d^3k_2'}\,\widetilde{d^3k_3'} \\
&\times \delta\left(W+\omega_1-\omega_1'-\omega_2'-\omega_3'\right)\left(2\pi\right)^3\delta^3\left(k_1-k_1'-k_2'-k_3'\right) \\
&\times \left(\frac{1}{\exp\left(\beta\omega_1\right)-1}\right)\left(1+\frac{1}{\exp\left(\beta\omega_1'\right)-1}\right) \\
&\times \left(1+\frac{1}{\exp\left(\beta\omega_2'\right)-1}\right)\left(1+\frac{1}{\exp\left(\beta\omega_3'\right)-1}\right),
\end{aligned} \tag{4.C.5}$$

$$\begin{aligned}
\rho^{\text{tree}}_{n\to n-2}\left(W\right) =& \frac{1}{3!}V\left|\int dt\,\lambda(t)\exp\left(iWt\right)\right|^2 \int \widetilde{d^3k_1}\,\widetilde{d^3k_2}\,\widetilde{d^3k_3}\,\widetilde{d^3k_1'} \\
&\times \delta\left(W+\omega_1+\omega_2+\omega_3-\omega_1'\right)\left(2\pi\right)^3\delta^3\left(k_1+k_2+k_3-k_1'\right) \\
&\times \left(\frac{1}{\exp\left(\beta\omega_1\right)-1}\right)\left(\frac{1}{\exp\left(\beta\omega_2\right)-1}\right) \\
&\times \left(\frac{1}{\exp\left(\beta\omega_3\right)-1}\right)\left(1+\frac{1}{\exp\left(\beta\omega_1'\right)-1}\right).
\end{aligned} \tag{4.C.6}$$

These tree-level contributions follow the same pattern as the work distributions (4.C.1)-(4.C.3). The contributions which arise from the loop diagrams are

$$\begin{aligned}
\rho^{\text{loop}}_{n\to n+2}\left(W\right) =& \frac{1}{2}V\left|\int dt\,\lambda(t)\exp\left(iWt\right)\right|^2 \left(1+\frac{1}{\exp\left(\beta W/2\right)-1}\right)^2 \frac{1}{W}\left(\int \widetilde{d^3k}\,\delta\left(W-2\omega\right)\right) \\
&\times \left(\int \widetilde{d^3k}\,\frac{1}{\exp\left(\beta\omega\right)-1}+\frac{1}{2}\left(\frac{m}{4\pi}\right)^2\left[1+\log\left(\frac{\mu^2}{m^2}\right)\right]\right)^2,
\end{aligned} \tag{4.C.7}$$

$$\begin{aligned}
\rho^{\text{loop}}_{n\to n-2}\left(W\right) =& \frac{1}{2}V\left|\int dt\,\lambda(t)\exp\left(iWt\right)\right|^2 \left(\frac{1}{\exp\left(-\beta W/2\right)-1}\right)^2 \frac{1}{-W}\left(\int \widetilde{d^3k}\,\delta\left(W+2\omega\right)\right) \\
&\times \left(\int \widetilde{d^3k}\,\frac{1}{\exp\left(\beta\omega\right)-1}+\frac{1}{2}\left(\frac{m}{4\pi}\right)^2\left[1+\log\left(\frac{\mu^2}{m^2}\right)\right]\right)^2.
\end{aligned} \tag{4.C.8}$$

In these expressions, μ is the $\overline{MS}$ renormalization scale. These expressions partially include contributions from tree-level diagrams because the loop and tree diagrams interfere.

Bibliography

[1] A. Bartolotta and M. J. Ramsey-Musolf, *Coherent $\mu - e$ Conversion at Next-to-Leading Order*, arXiv:1710.0212.

[2] A. Bartolotta, S. M. Carroll, S. Leichenauer, and J. Pollack, *Bayesian second law of thermodynamics*, *Phys. Rev. E* **94** (Aug, 2016) 022102, [arXiv:1508.0242].

[3] A. Bartolotta and S. Deffner, *Jarzynski equality for driven quantum field theories*, *Phys. Rev. X* **8** (Feb, 2018) 011033, [arXiv:1710.0082].

[4] M. Srednicki, *Quantum Field Theory*. Cambridge Univ. Press, Cambridge, 2007.

[5] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory; 1995 ed.* Westview, Boulder, CO, 1995.

[6] A. Zee, *Quantum Field Theory in a Nutshell*. In a Nutshell. Princeton Univ. Press, Princeton, NJ, 2003.

[7] A. M. Tsvelik, *Quantum field theory in condensed matter physics*. Cambridge University Press, 2007.

[8] N. Nagaosa, *Quantum field theory in condensed matter physics*. Springer Science & Business Media, 2013.

[9] E. Fradkin, *Field theories of condensed matter physics*. Cambridge University Press, 2013.

[10] S. L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics* **22** (1961), no. 4 579 – 588.

[11] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[12] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.

[13] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509.

[14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587.

[15] **CMS** Collaboration, *Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc*, *Physics Letters B* **716** (2012), no. 1 30 – 61.

[16] **ATLAS** Collaboration, *Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc*, *Physics Letters B* **716** (2012), no. 1 1 – 29.

[17] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse, *Cavity control of a single-electron quantum cyclotron: Measuring the electron magnetic moment*, *Phys. Rev. A* **83** (May, 2011) 052122.

[18] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Tenth-order qed contribution to the electron g−2 and an improved value of the fine structure constant*, *Phys. Rev. Lett.* **109** (Sep, 2012) 111807.

[19] K. Hagiwara, A. Martin, D. Nomura, and T. Teubner, *Improved predictions for $g-2$ of the muon and $\alpha_{QED}(m_Z^2)$*, *Physics Letters B* **649** (2007), no. 2 173 – 179.

[20] **Muon g-2** Collaboration, J. Grange et al., *Muon (g-2) Technical Design Report*, arXiv:1501.0685.

[21] N. F. Ramsey, *Electric-dipole moments of particles*, *Annual Review of Nuclear and Particle Science* **32** (1982), no. 1 211–233, [https://doi.org/10.1146/annurev.ns.32.120182.001235].

[22] S. Dar, *The Neutron EDM in the SM: A Review*, hep-ph/0008248.

[23] J. M. Pendlebury et al., *Revised experimental upper limit on the electric dipole moment of the neutron*, *Phys. Rev. D* **92** (Nov, 2015) 092003.

[24] **MEG** Collaboration, A. M. Baldini et al., *Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment*, *Eur. Phys. J.* **C76** (2016), no. 8 434, [arXiv:1605.0508].

[25] **Mu2e** Collaboration, R. J. Abrams et al., *Mu2e Conceptual Design Report*, arXiv:1211.7019.

[26] **COMET** Collaboration, Y. G. Cui et al., *Conceptual design report for experimental search for lepton flavor violating $\mu - e$ conversion at sensitivity of 10^{-16} with a slow-extracted bunched proton beam (COMET)*, .

[27] R. H. Bernstein and P. S. Cooper, *Charged Lepton Flavor Violation: An Experimenter's Guide*, *Phys. Rept.* **532** (2013) 27–64, [arXiv:1307.5787].

[28] M. Lindner, M. Platscher, and F. S. Queiroz, *A Call for New Physics : The Muon Anomalous Magnetic Moment and Lepton Flavor Violation*, arXiv:1610.0658.

[29] **Super-Kamiokande** Collaboration, *Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ in a large water cherenkov detector*, *Phys. Rev. Lett.* **102** (Apr, 2009) 141801.

[30] F. D. Lodovico and H.-K. Collaboration, *The hyper-kamiokande experiment*, *Journal of Physics: Conference Series* **888** (2017), no. 1 012020.

[31] S. Carnot, *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance*, *Paris: Bachelier* (1824).

[32] J. W. Gibbs, *Elementary principles in statistical mechanics*. Charles Scribner's Sons, 1902.

[33] R. C. Tolman, *The principles of statistical mechanics*. Courier Corporation, 1938.

[34] J. Uffink, *Compendium of the foundations of classical statistical physics*, .

[35] L. Onsager, *Reciprocal relations in irreversible processes. i.*, *Phys. Rev.* **37** (Feb, 1931) 405–426.

[36] H. Nyquist, *Thermal agitation of electric charge in conductors*, *Phys. Rev.* **32** (Jul, 1928) 110–113.

[37] H. B. Callen and T. A. Welton, *Irreversibility and generalized noise*, *Phys. Rev.* **83** (Jul, 1951) 34–40.

[38] D. J. Evans and D. J. Searles, *Equilibrium microstates which generate second law violating steady states*, *Phys. Rev. E* **50** (Aug, 1994) 1645–1648.

[39] C. Jarzynski, *Nonequilibrium equality for free energy differences*, *Phys. Rev. Lett.* **78** (1997) 2690.

[40] C. Jarzynski, *Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach*, *Phys. Rev. E* **56** (1997), no. 5 5018.

[41] G. E. Crooks, *Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences*, *Phys. Rev. E* **60** (Sep, 1999) 2721–2726.

[42] C. Jarzynski, *Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale*, *Ann. Rev. Cond. Matt. Phys.* **2** (2011), no. 1 329–351.

[43] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, *Probability of second law violations in shearing steady states*, *Phys. Rev. Lett.* **71** (Oct, 1993) 2401–2404.

[44] D. J. Searles and D. J. Evans, *Fluctuation theorem for stochastic systems*, *Phys. Rev. E* **60** (Jul, 1999) 159–164.

[45] A. E. Allahverdyan and T. M. Nieuwenhuizen, *Fluctuations of work from quantum subensembles: The case against quantum work-fluctuation theorems*, *Phys. Rev. E* **71** (Jun, 2005) 066102.

[46] S. Mukamel, *Quantum extension of the jarzynski relation: Analogy with stochastic dephasing*, *Phys. Rev. Lett.* **90** (May, 2003) 170604.

[47] A. Engel and R. Nolte, *Jarzynski equation for a simple quantum system: Comparing two definitions of work*, *EPL (Europhysics Letters)* **79** (2007), no. 1 10003.

[48] P. Hänggi and P. Talkner, *The other QFT*, *Nat. Phys.* **11** (Feb, 2015) 108–110.

[49] M. Campisi, P. Hänggi, and P. Talkner, *Colloquium: Quantum fluctuation relations: Foundations and applications*, *Rev. Mod. Phys.* **83** (Jul, 2011) 771–791.

[50] M. Campisi, P. Hänggi, and P. Talkner, *Erratum: Colloquium: Quantum fluctuation relations: Foundations and applications*, *Rev. Mod. Phys.* **83** (Dec, 2011) 1653–1653.

[51] M. Esposito, U. Harbola, and S. Mukamel, *Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems*, *Rev. Mod. Phys.* **81** (Dec, 2009) 1665–1702.

[52] M. Esposito, U. Harbola, and S. Mukamel, *Erratum: Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems*, *Rev. Mod. Phys.* **86** (Sep, 2014) 1125–1125.

[53] M. Esposito and S. Mukamel, *Fluctuation theorems for quantum master equations*, *Phys. Rev. E* **73** (Apr, 2006) 046129.

[54] C. Jarzynski and D. K. Wójcik, *Classical and quantum fluctuation theorems for heat exchange*, *Phys. Rev. Lett.* **92** (Jun, 2004) 230602.

[55] T. Monnai, *Unified treatment of the quantum fluctuation theorem and the jarzynski equality in terms of microscopic reversibility*, *Phys. Rev. E* **72** (Aug, 2005) 027102.

[56] P. Talkner and P. Hänggi, *The Tasaki-Crooks quantum fluctuation theorem*, *J. Phys. A: Math. Theor.* **40** (2007), no. 26 F569.

[57] P. Talkner, P. Hänggi, and M. Morillo, *Microcanonical quantum fluctuation theorems*, *Phys. Rev. E* **77** (May, 2008) 051131.

[58] A. E. Allahverdyan, *Nonequilibrium quantum fluctuations of work*, *Phys. Rev. E* **90** (Sep, 2014) 032137.

[59] P. Talkner, E. Lutz, and P. Hänggi, *Fluctuation theorems: Work is not an observable*, *Phys. Rev. E* **75** (May, 2007) 050102.

[60] J. Kurchan, *A Quantum Fluctuation Theorem*, *arXiv:cond-mat/0007360* (July, 2000).

[61] H. Tasaki, *Jarzynski Relations for Quantum Systems and Some Applications*, *arXiv:cond-mat/0009244* (Sept., 2000).

[62] P. Talkner and P. Hänggi, *Aspects of quantum work*, *Phys. Rev. E* **93** (Feb, 2016) 022131.

[63] S. Deffner, J. P. Paz, and W. H. Zurek, *Quantum work and the thermodynamic cost of quantum measurements*, *Phys. Rev. E* **94** (Jul, 2016) 010103.

[64] M. Perarnau-Llobet, E. Bäumer, K. V. Hovhannisyan, M. Huber, and A. Acin, *No-go theorem for the characterization of work fluctuations in coherent quantum systems*, *Phys. Rev. Lett.* **118** (Feb, 2017) 070601.

[65] A. Czarnecki, X. Garcia i Tormo, and W. J. Marciano, *Muon decay in orbit: Spectrum of high-energy electrons*, *Phys. Rev. D* **84** (Jul, 2011) 013006.

[66] R. Szafron and A. Czarnecki, *High-energy electrons from the muon decay in orbit: Radiative corrections*, *Physics Letters B* **753** (2016) 61 – 64.

[67] P. Wintz, *Results of the SINDRUM-II experiment*, *Conf. Proc.* **C980420** (1998) 534–546.

[68] R. Kitano, M. Koike, and Y. Okada, *Detailed calculation of lepton flavor violating muon-electron conversion rate for various nuclei*, *Phys. Rev. D* **66** (Nov, 2002) 096002.

[69] R. Kitano, M. Koike, and Y. Okada, *Erratum: Detailed calculation of lepton flavor violating muon-electron conversion rate for various nuclei*, *Phys. Rev. D* **76** (Sep, 2007) 059902.

[70] V. Cirigliano, R. Kitano, Y. Okada, and P. Tuzon, *Model discriminating power of $\mu \to e$ conversion in nuclei*, *Phys. Rev. D* **80** (Jul, 2009) 013002.

[71] A. Crivellin, M. Hoferichter, and M. Procura, *Improved predictions for $\mu \to e$ conversion in nuclei and higgs-induced lepton flavor violation*, *Phys. Rev. D* **89** (May, 2014) 093024.

[72] A. Crivellin, M. Hoferichter, and M. Procura, *Accurate evaluation of hadronic uncertainties in spin-independent wimp-nucleon scattering: Disentangling two- and three-flavor effects*, *Phys. Rev. D* **89** (Mar, 2014) 054021.

[73] J. C. Pati and A. Salam, *Lepton number as the fourth "color"*, *Phys. Rev. D* **10** (Jul, 1974) 275–289.

[74] H. Georgi and S. L. Glashow, *Unity of all elementary-particle forces*, *Phys. Rev. Lett.* **32** (Feb, 1974) 438–441.

[75] E. Farhi and L. Susskind, *Technicolour*, *Physics Reports* **74** (1981), no. 3 277 – 321.

[76] M. J. Dugan, H. Georgi, and D. B. Kaplan, *Anatomy of a Composite Higgs Model*, *Nucl. Phys.* **B254** (1985) 299.

[77] L. Randall and R. Sundrum, *Large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (Oct, 1999) 3370–3373.

[78] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) 1–102, [arXiv:1106.0034].

[79] J. D. Bjorken and S. Weinberg, *Mechanism for nonconservation of muon number*, *Phys. Rev. Lett.* **38** (Mar, 1977) 622–625.

[80] W. Buchmüller, R. Rückl, and D. Wyler, *Leptoquarks in lepton-quark collisions*, *Physics Letters B* **191** (1987), no. 4 442 – 448.

[81] B. McWilliams and L.-F. Li, *Virtual Effects of Higgs Particles*, *Nucl. Phys.* **B179** (1981) 62–84.

[82] O. U. Shanker, *Flavor Violation, Scalar Particles and Leptoquarks*, *Nucl. Phys.* **B206** (1982) 253–272.

[83] I. Doršner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Košnik, *Physics of leptoquarks in precision experiments and at particle colliders*, arXiv:1603.0499.

[84] Y. Kuno and Y. Okada, *Muon decay and physics beyond the standard model*, *Rev. Mod. Phys.* **73** (Jan, 2001) 151–202.

[85] M. Shifman, A. Vainshtein, and V. Zakharov, *Remarks on higgs-boson interactions with nucleons*, *Physics Letters B* **78** (1978), no. 4 443 – 446.

[86] V. Bernard, N. Kaiser, and U.-G. Meissner, *Chiral dynamics in nucleons and nuclei*, *Int. J. Mod. Phys.* **E4** (1995) 193–346, [hep-ph/9501384].

[87] E. E. Jenkins and A. V. Manohar, *Baryon chiral perturbation theory using a heavy fermion Lagrangian*, *Phys. Lett.* **B255** (1991) 558–562.

[88] S. Weinberg, *Effective chiral lagrangians for nucleon-pion interactions and nuclear forces*, *Nuclear Physics B* **363** (1991), no. 1 3 – 18.

[89] P. F. Bedaque and U. van Kolck, *Effective field theory for few nucleon systems*, *Ann. Rev. Nucl. Part. Sci.* **52** (2002) 339–396, [nucl-th/0203055].

[90] V. Cirigliano, M. L. Graesser, and G. Ovanesyan, *Wimp-nucleus scattering in chiral effective theory*, *Journal of High Energy Physics* **2012** (2012), no. 10 1–22.

[91] M. Hoferichter, P. Klos, and A. Schwenk, *Chiral power counting of one- and two-body currents in direct detection of dark matter*, *Phys. Lett.* **B746** (2015) 410–416, [arXiv:1503.0481].

[92] M. Hoferichter, P. Klos, J. Menéndez, and A. Schwenk, *Analysis strategies for general spin-independent WIMP-nucleus scattering*, *Phys. Rev.* **D94** (2016), no. 6 063505, [arXiv:1605.0804].

[93] C. Körber, A. Nogga, and J. de Vries, *First-principle calculations of Dark Matter scattering off light nuclei*, *Phys. Rev.* **C96** (2017), no. 3 035805, [arXiv:1704.0115].

[94] U. L. van Kolck, *Soft Physics: Applications of Effective Chiral Lagrangians to Nuclear Physics and Quark Models*. PhD thesis, The University of Texas at Austin, Jan., 1993.

[95] U. van Kolck, *Isospin Violation in Low-energy Hadronic Physics*, pp. 444–448. Springer Vienna, Vienna, 1996.

[96] W. C. Haxton, *Parity Nonconservation in ^{18}F and Meson Exchange Contributions to the Axial Charge Operator*, *Phys. Rev. Lett.* **46** (1981) 698.

[97] W. C. Haxton, E. M. Henley, and M. J. Musolf, *Nucleon and nuclear anapole moments*, *Phys. Rev. Lett.* **63** (Aug, 1989) 949–952.

[98] M. J. Musolf and T. W. Donnelly, *Meson exchange currents and the strangeness radius of He-4*, *Phys. Lett.* **B318** (1993) 263–267, [nucl-th/9309009].

[99] B. A. Brown and W. A. Richter, *New 'USD' Hamiltonians for the sd shell*, *Phys. Rev.* **C74** (2006) 034315.

[100] T. W. Donnelly and I. Sick, *Elastic Magnetic Electron Scattering From Nuclei*, *Rev. Mod. Phys.* **56** (1984) 461–566.

[101] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, arXiv:1607.0029.

[102] P. M. Junnarkar and A. Walker-Loud, *Scalar strange content of the nucleon from lattice qcd*, *Phys. Rev. D* **87** (Jun, 2013) 114510.

[103] L. Alvarez-Ruso, T. Ledwig, M. J. V. Vacas, and J. Martin-Camalich, *The nucleon mass and pion-nucleon sigma term from a chiral analysis of $N_f = 2 + 1$ lattice QCD world data*, *Int. J. Mod. Phys. Conf. Ser.* **26** (2014) 1460089, [arXiv:1402.3991].

[104] S. Durr et al., *Lattice computation of the nucleon scalar quark contents at the physical point*, *Phys. Rev. Lett.* **116** (2016), no. 17 172001, [arXiv:1510.0801].

[105] **xQCD** Collaboration, Y.-B. Yang, A. Alexandru, T. Draper, J. Liang, and K.-F. Liu, *πN and strangeness sigma terms at the physical point with chiral fermions*, *Phys. Rev.* **D94** (2016), no. 5 054503, [arXiv:1511.0908].

[106] **ETM** Collaboration, A. Abdel-Rehim, C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, and A. Vaquero Aviles-Casco, *Direct Evaluation of the Quark Content of Nucleons from Lattice QCD at the Physical Point*, *Phys. Rev. Lett.* **116** (2016), no. 25 252001, [arXiv:1601.0162].

[107] **RQCD** Collaboration, G. S. Bali, S. Collins, D. Richtmann, A. Schäfer, W. Söldner, and A. Sternbeck, *Direct determinations of the nucleon and pion σ terms at nearly physical quark masses*, *Phys. Rev.* **D93** (2016), no. 9 094504, [arXiv:1603.0082].

[108] J. M. Alarcón, J. M. Camalich, and J. A. Oller, *Chiral representation of the πn scattering amplitude and the pion-nucleon sigma term*, *Phys. Rev. D* **85** (Mar, 2012) 051503.

[109] J. Ruiz de Elvira, M. Hoferichter, B. Kubis, and U.-G. Meissner, *Extracting the*

σ-term from low-energy pion-nucleon scattering, *J. Phys.* **G45** (2018), no. 2 024001, [arXiv:1706.0146].

[110] **Particle Data Group** Collaboration, C. Patrignani et al., *Review of Particle Physics*, *Chin. Phys.* **C40** (2016), no. 10 100001.

[111] **UCNA** Collaboration, J. Liu et al., *Determination of the axial-vector weak coupling constant with ultracold neutrons*, *Phys. Rev. Lett.* **105** (Oct, 2010) 181803.

[112] M. Rose, *Relativistic Electron Theory*. John Wiley, New York, 1961.

[113] R. R. Silbar and T. Goldman, *Solving the radial dirac equations: a numerical odyssey*, *European Journal of Physics* **32** (2011), no. 1 217.

[114] J. Negele and E. Vogt, *Advances in Nuclear Physics*. No. v. 19 in Advances in Nuclear Physics. Springer US, 1989.

[115] H. De Vries, C. De Jager, and C. De Vries, *Nuclear charge-density-distribution parameters from elastic electron scattering*, *Atomic data and nuclear data tables* **36** (1987), no. 3 495–536.

[116] C. Garcia-Recio, J. Nieves, and E. Oset, *Neutron distributions from pionic atoms*, *Nuclear Physics A* **547** (1992), no. 3 473 – 487.

[117] G. Ecker and M. Mojžiš, *Low-energy expansion of the pion-nucleon lagrangian*, *Physics Letters B* **365** (1996), no. 1âĂŞ4 312 – 318.

[118] G. Ecker and M. Mojzis, *Wave function renormalization in heavy baryon chiral perturbation theory*, *Phys. Lett.* **B410** (1997) 266–274, [hep-ph/9705216].

[119] V. Bernard, N. Kaiser, and U.-G. Meissner, *Aspects of chiral pion - nucleon physics*, *Nucl. Phys.* **A615** (1997) 483–500, [hep-ph/9611253].

[120] U.-G. Meissner, G. Muller, and S. Steininger, *Renormalization of the chiral pion - nucleon Lagrangian beyond next-to-leading order*, *Annals Phys.* **279** (2000) 1–64, [hep-ph/9809446].

[121] C. Perdrisat, V. Punjabi, and M. Vanderhaeghen, *Nucleon electromagnetic form factors*, *Progress in Particle and Nuclear Physics* **59** (2007), no. 2 694 – 764.

[122] Pate, Stephen and Trujillo, Dennis, *Strangeness vector and axial-vector form factors of the nucleon*, *EPJ Web of Conferences* **66** (2014) 06018.

[123] S. Durr, Z. Fodor, C. Hoelbling, S. Katz, S. Krieg, T. Kurth, L. Lellouch, T. Lippert, K. Szabo, and G. Vulvert, *Lattice {QCD} at the physical point: Light quark masses*, *Physics Letters B* **701** (2011), no. 2 265 – 268.

[124] S. Durr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, T. Kurth, L. Lellouch, T. Lippert, K. K. Szabo, and G. Vulvert, *Lattice QCD at the physical point: Simulation and analysis details*, *JHEP* **08** (2011) 148, [arXiv:1011.2711].

[125] **HPQCD** Collaboration, C. McNeile, C. T. H. Davies, E. Follana, K. Hornbostel, and G. P. Lepage, *High-precision c and b masses, and qcd coupling from current-current correlators in lattice and continuum qcd*, *Phys. Rev. D* **82** (Aug, 2010) 034512.

[126] **RBC and UKQCD** Collaboration, T. Blum et al., *Domain wall qcd with physical quark masses*, *Phys. Rev. D* **93** (Apr, 2016) 074505.

[127] A. Bazavov et al., *Staggered chiral perturbation theory in the two-flavor case and SU(2) analysis of the MILC data*, *PoS* **LATTICE2010** (2010) 083, [arXiv:1011.1792].

[128] **MILC** Collaboration, A. Bazavov et al., *MILC results for light pseudoscalars*, *PoS* **CD09** (2009) 007, [arXiv:0910.2966].

[129] **HPQCD and UKQCD** Collaboration, E. Follana, C. T. H. Davies, G. P. Lepage, and J. Shigemitsu, *High-precision determination of the π, k, d, and D_s decay constants from lattice qcd*, *Phys. Rev. Lett.* **100** (Feb, 2008) 062002.

[130] **MILC** Collaboration, A. Bazavov et al., *Results for light pseudoscalar mesons*, *PoS* **LATTICE2010** (2010) 074, [arXiv:1012.0868].

[131] **RBC and UKQCD** Collaboration, R. Arthur et al., *Domain wall qcd with near-physical pions*, *Phys. Rev. D* **87** (May, 2013) 094514.

[132] S. Borsányi, S. Dürr, Z. Fodor, S. Krieg, A. Schäfer, E. E. Scholz, and K. K. Szabó, *Su(2) chiral perturbation theory low-energy constants from 2+1 flavor staggered lattice simulations*, *Phys. Rev. D* **88** (Jul, 2013) 014513.

[133] **Budapest-Marseille-Wuppertal** Collaboration, S. Dürr et al., *Lattice qcd at the physical point meets su(2) chiral perturbation theory*, *Phys. Rev. D* **90** (Dec, 2014) 114504.

[134] **NPLQCD** Collaboration, S. R. Beane et al., *su(2) low-energy constants from mixed-action lattice qcd*, *Phys. Rev. D* **86** (Nov, 2012) 094509.

[135] E. J. Moniz, I. Sick, R. R. Whitney, J. R. Ficenec, R. D. Kephart, and W. P. Trower, *Nuclear fermi momenta from quasielastic electron scattering*, *Phys. Rev. Lett.* **26** (Feb, 1971) 445–448.

[136] T. Suzuki, D. F. Measday, and J. P. Roalsvig, *Total nuclear capture rates for negative muons*, *Phys. Rev. C* **35** (Jun, 1987) 2212–2224.

[137] G. E. Crooks, *Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems*, *Journal of Statistical Physics* **90** (1998), no. 5-6 1481–1487.

[138] D. J. Evans and D. J. Searles, *The fluctuation theorem*, *Adv. Phys.* **51** (2002), no. 7 1529–1585.

[139] T. Sagawa and M. Ueda, *Nonequilibrium thermodynamics of feedback control*, *Physical Review E* **85** (2012), no. 2 021104.

[140] J. L. England, *Statistical physics of self-replication*, *The Journal of chemical physics* **139** (2013), no. 12 121923.

[141] U. Seifert, *Stochastic thermodynamics, fluctuation theorems and molecular machines*, *Reports on Progress in Physics* **75** (2012), no. 12 126001.

[142] C. E. Shannon, *A mathematical theory of communication*, *Bell Syst.Tech.J.* **27** (1948) 379–423.

[143] H. D. Zeh, *The Physical Basis of the Direction of Time*. Springer Science & Business Media, 2001.

[144] R. T. Bayes, *An essay toward solving a problem in the doctrine of chances*, *Phil.Trans.Roy.Soc.Lond.* **53** (1764) 370–418.

[145] P. S. Marquis de Laplace, *Théorie Analytique des Probabilités*. V. Courcier, 1820.

[146] K. Maruyama, F. Nori, and V. Vedral, *Colloquium : The physics of maxwell's demon and information*, *Rev. Mod. Phys.* **81** (Jan, 2009) 1–23.

[147] J. M. Parrondo, J. M. Horowitz, and T. Sagawa, *Thermodynamics of information*, *Nature Physics* **11** (2015), no. 2 131–139.

[148] S. Lloyd, *Use of mutual information to decrease entropy: Implications for the second law of thermodynamics*, *Phys. Rev. A* **39** (May, 1989) 5378–5386.

[149] S. Kullback and R. A. Leibler, *On information and sufficiency*, *Ann. Math. Statist.* **22** (03, 1951) 79–86.

[150] U. Seifert, *Stochastic thermodynamics: principles and perspectives*, *The European Physical Journal B* **64** (2008), no. 3-4 423–431.

[151] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, *Path-integral analysis of fluctuation theorems for general langevin processes*, *Journal of Statistical Mechanics: Theory and Experiment* **2006** (2006), no. 08 P08001.

[152] S. Deffner, M. Brunner, and E. Lutz, *Quantum fluctuation theorems in the strong damping limit*, *EPL (Europhysics Letters)* **94** (2011), no. 3 30001.

[153] S. Rahav and C. Jarzynski, *Fluctuation relations and coarse-graining*, *Journal of Statistical Mechanics: Theory and Experiment* **2007** (2007), no. 09 P09012.

[154] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck, *Lower bounds on dissipation upon coarse graining*, *Phys. Rev. E* **78** (Jul, 2008) 011107.

[155] R. Kawai, J. M. R. Parrondo, and C. V. den Broeck, *Dissipation: The phase-space perspective*, *Phys. Rev. Lett.* **98** (Feb, 2007) 080602.

[156] C. Jarzynski, *Rare events and the convergence of exponentially averaged work values*, *Phys. Rev. E* **73** (Apr, 2006) 046105.

[157] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., Orlando, FL, USA, 1982.

[158] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.

[159] J. E. Cohen, Y. Derriennic, and G. Zbaganu, *Majorization, monotonicity of relative entropy, and stochastic matrices*, *Contemporary Mathematics* **149** (1993).

[160] T. Sagawa and M. Ueda, *Role of mutual information in entropy production under information exchanges*, *New Journal of Physics* **15** (2013), no. 12 125012.

[161] D. V. Lindley, *On a measure of the information provided by an experiment*, *Ann. Math. Statist.* **27** (12, 1956) 986–1005.

[162] K. Chaloner and I. Verdinelli, *Bayesian experimental design: A review*, *Statist. Sci.* **10** (08, 1995) 273–304.

[163] T. Sagawa and M. Ueda, *Second law of thermodynamics with discrete quantum feedback control*, *Phys. Rev. Lett.* **100** (Feb, 2008) 080403.

[164] T. Sagawa and M. Ueda, *Generalized jarzynski equality under nonequilibrium feedback control*, *Phys. Rev. Lett.* **104** (Mar, 2010) 090602.

[165] J. M. Horowitz and S. Vaikuntanathan, *Nonequilibrium detailed fluctuation theorem for repeated discrete feedback*, *Phys. Rev. E* **82** (Dec, 2010) 061120.

[166] M. Ponmurugan, *Generalized detailed fluctuation theorem under nonequilibrium feedback control*, *Phys. Rev. E* **82** (Sep, 2010) 031129.

[167] M. Campisi, P. Talkner, and P. Hänggi, *Influence of measurements on the statistics of work performed on a quantum system*, *Phys. Rev. E* **83** (Apr, 2011) 041114.

[168] D. Kafri and S. Deffner, *Holevo's bound from a general quantum fluctuation theorem*, *Phys. Rev. A* **86** (Oct, 2012) 044302.

[169] G. Watanabe, B. P. Venkatesh, P. Talkner, M. Campisi, and P. Hänggi, *Quantum fluctuation theorems and generalized measurements during the force protocol*, *Phys. Rev. E* **89** (Mar, 2014) 032114.

[170] J. E. Guyer, D. Wheeler, and J. A. Warren, *FiPy: Partial differential equations with Python*, *Computing in Science & Engineering* **11** (2009), no. 3 6–15.

[171] C. R. Shalizi, *The backwards arrow of time of the coherently bayesian statistical mechanic*, cond-mat/0410063.

[172] H. Walther, B. T. H. Varcoe, B.-G. Englert, and T. Becker, *Cavity quantum electrodynamics*, *Rep. Prog. Phys.* **69** (2006), no. 5 1325.

[173] G. Gallavotti and E. G. D. Cohen, *Dynamical ensembles in nonequilibrium statistical mechanics*, *Phys. Rev. Lett.* **74** (Apr, 1995) 2694–2697.

[174] H. B. Callen, *Thermodynamics and an introduction to thermostatistics*, .

[175] J. M. Ortiz de Zárate, *Interview with Michael E. Fisher*, *Europhys. News* **42** (2011), no. 1 14.

[176] U. Seifert, *Entropy production along a stochastic trajectory and an integral fluctuation theorem*, *Phys. Rev. Lett.* **95** (Jul, 2005) 040602.

[177] C. Van den Broeck, *Stochastic thermodynamics*, in *Selforganization by nonlinear irreversible processes*, pp. 57–61. Springer, 1986.

[178] E. M. Sevick, R. Prabhakar, S. R. Williams, and D. J. Searles, *Fluctuation theorems*, *Ann. Rev. Phys. Chem.* **59** (2008), no. 1 603–633. PMID: 18393680.

[179] S. Ciliberto, *Experiments in stochastic thermodynamics: Short history and perspectives*, *Phys. Rev. X* **7** (Jun, 2017) 021051.

[180] S. Deffner, O. Abah, and E. Lutz, *Quantum work statistics of linear and nonlinear parametric oscillators*, *Chem. Phys.* **375** (2010), no. 2 200.

[181] C. Jarzynski, H. T. Quan, and S. Rahav, *Quantum-classical correspondence principle for work distributions*, *Phys. Rev. X* **5** (Sep, 2015) 031038.

[182] L. Zhu, Z. Gong, B. Wu, and H. T. Quan, *Quantum-classical correspondence principle for work distributions in a chaotic system*, *Phys. Rev. E* **93** (Jun, 2016) 062108.

[183] Q. Wang and H. T. Quan, *Understanding quantum work in a quantum many-body system*, *Phys. Rev. E* **95** (Mar, 2017) 032113.

[184] I. García-Mata, A. J. Roncaglia, and D. A. Wisniacki, *Quantum-to-classical transition in the work distribution for chaotic systems*, *Phys. Rev. E* **95** (May, 2017) 050102.

[185] S. Deffner, *Quantum entropy production in phase space*, *EPL (Europhysics Letters)* **103** (2013), no. 3 30001.

[186] P. Talkner, M. Campisi, and P. Hänggi, *Fluctuation theorems in driven open quantum systems*, *J. Stat. Mech.: Theo. Exp.* **2009** (2009), no. 02 P02025.

[187] S. Suomela, P. Solinas, J. P. Pekola, J. Ankerhold, and T. Ala-Nissila, *Moments of work in the two-point measurement protocol for a driven open quantum system*, *Phys. Rev. B* **90** (Sep, 2014) 094304.

[188] M. Campisi, P. Talkner, and P. Hänggi, *Fluctuation theorem for arbitrary open quantum systems*, *Phys. Rev. Lett.* **102** (May, 2009) 210401.

[189] B. Leggio, A. Napoli, A. Messina, and H.-P. Breuer, *Entropy production and information fluctuations along quantum trajectories*, *Phys. Rev. A* **88** (Oct, 2013) 042111.

[190] B. Leggio, A. Napoli, H.-P. Breuer, and A. Messina, *Fluctuation theorems for non-Markovian quantum processes*, *Phys. Rev. E* **87** (Mar, 2013) 032113.

[191] M. C., *Quantum fluctuation relations for ensembles of wave functions*, *New J. Phys.* **15** (2013), no. 11 115008.

[192] S. Pigeon, L. Fusco, A. Xuereb, G. D. Chiara, and M. Paternostro, *Thermodynamics of trajectories and local fluctuation theorems for harmonic quantum networks*, *New J. Phys.* **18** (2016), no. 1 013009.

[193] J. P. Santos, G. T. Landi, and M. Paternostro, *Wigner entropy production rate*, *Phys. Rev. Lett.* **118** (Jun, 2017) 220601.

[194] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, *Quantum and information thermodynamics: A unifying framework based on repeated interactions*, *Phys. Rev. X* **7** (Apr, 2017) 021003.

[195] S. Deffner and A. Saxena, *Quantum work statistics of charged dirac particles in time-dependent fields*, *Phys. Rev. E* **92** (Sep, 2015) 032137.

[196] S. Deffner and A. Saxena, *Jarzynski equality in $\mathcal{PT}$-symmetric quantum mechanics*, *Phys. Rev. Lett.* **114** (Apr, 2015) 150601.

[197] A. E. Allahverdyan and S. G. Babajanyan, *Electromagnetic gauge-freedom and work*, *J. Phys. A: Math. Theor.* **49** (2016), no. 28 285001.

[198] B. Gardas, S. Deffner, and A. Saxena, *Non-hermitian quantum thermodynamics*, *Sci. Rep.* **6** (2016) 23408.

[199] S. Lin and Z. Song, *Non-hermitian heat engine with all-quantum-adiabatic-process cycle*, *J. Phys. A: Math. Theor.* **49** (2016), no. 47 475301.

[200] N. Liu, J. Goold, I. Fuentes, V. Vedral, K. Modi, and D. E. Bruschi, *Quantum thermodynamics for a model of an expanding universe*, *Classical and Quantum Gravity* **33** (2016), no. 3 035003.

[201] M. Caselle, G. Costagliola, A. Nada, M. Panero, and A. Toniato, *Jarzynski's theorem for lattice gauge theory*, *Phys. Rev. D* **94** (Aug, 2016) 034503.

[202] R. Auzzi, S. Elitzur, S. B. Gudnason, and E. Rabinovici, *On periodically driven ads/cft*, *Journal of High Energy Physics* **2013** (Nov, 2013) 16.

[203] J. Marino and S. Diehl, *Driven markovian quantum criticality*, *Phys. Rev. Lett.* **116** (Feb, 2016) 070407.

[204] J. Marino and S. Diehl, *Quantum dynamical field theory for nonequilibrium phase transitions in driven open systems*, *Phys. Rev. B* **94** (Aug, 2016) 085150.

[205] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information:*

10th Anniversary Edition. Cambridge University Press, New York, NY, USA, 10th ed., 2011.

[206] T. Albash, D. A. Lidar, M. Marvian, and P. Zanardi, *Fluctuation theorems for quantum processes*, *Phys. Rev. E* **88** (Sep, 2013) 032146.

[207] A. E. Rastegin, *Non-equilibrium equalities with unital quantum channels*, *J. Stat. Mech.: Theo. Exp.* **2013** (2013), no. 06 P06016.

[208] A. E. Rastegin and K. Życzkowski, *Jarzynski equality for quantum stochastic maps*, *Phys. Rev. E* **89** (Jan, 2014) 012127.

[209] L. Fusco, S. Pigeon, T. J. G. Apollaro, A. Xuereb, L. Mazzola, M. Campisi, A. Ferraro, M. Paternostro, and G. De Chiara, *Assessing the nonequilibrium thermodynamics in a quenched quantum many-body system via single projective measurements*, *Phys. Rev. X* **4** (Aug, 2014) 031029.

[210] G. D. Mahan, *Condensed Matter in a Nutshell*. In a Nutshell. Princeton University Press, 2011.

[211] H. Leutwyler, *Phonons as goldstone bosons*, *Helv. Phys. Acta* **70** (1997) 275–286.

[212] V. L. Ginzburg and L. D. Landau, *On the Theory of superconductivity*, *Zh. Eksp. Teor. Fiz.* **20** (1950) 1064–1082.

[213] L. Landau, *The theory of phase transitions*, *Nature* **138** (1936) 840–841.

[214] H. Kleinert and V. Schulte-Frohlinde, *Critical properties of phi4-theories*. World Scientific, 2001.

[215] J. Goldstone, *Field theories with Âń superconductor Âż solutions*, *Il Nuovo Cimento (1955-1965)* **19** (Jan, 1961) 154–164.

[216] J. Goldstone, A. Salam, and S. Weinberg, *Broken symmetries*, *Phys. Rev.* **127** (Aug, 1962) 965–970.

[217] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, *Phys. Rev.* **145** (May, 1966) 1156–1163.

[218] D. Baumann, *Inflation*, in *Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, pp. 523–686, 2011.

[219] F. J. Dyson, *The s matrix in quantum electrodynamics*, *Phys. Rev.* **75** (Jun, 1949) 1736–1755.

[220] S. Weinberg, *High-energy behavior in quantum field theory*, *Phys. Rev.* **118** (May, 1960) 838–849.

[221] S. Deffner and C. Jarzynski, *Information processing and the second law of*

thermodynamics: An inclusive, hamiltonian approach, *Phys. Rev. X* **3** (Oct, 2013) 041003.

[222] G. Nenciu and G. Rasche, *Adiabatic theorem and gell-mann-low formula*, *Helv. Phys. Acta* **62** (1989), no. 4 372–388.

[223] L. G. Molinari, *Another proof of gell-mann and lowâĂŹs theorem*, *J. Math. Phys.* **48** (2007), no. 5 052113.

[224] T. V. Acconcia and S. Deffner, *Quantum speed limits and the maximal rate of quantum learning*, *arXiv:1706.03826* (June, 2017).

[225] J. Berges, *Introduction to nonequilibrium quantum field theory*, *AIP Conference Proceedings* **739** (2004), no. 1 3–62, [https://aip.scitation.org/doi/pdf/10.1063/1.1843591].

[226] E. A. Calzetta and B.-L. B. Hu, *Nonequilibrium Quantum Field Theory*. July, 2008.

[227] D. E. Bruschi and I. Fuentes, *Thermodynamics of relativistic quantum fields: extracting energy from gravitational waves*, arXiv:1607.0129.