Title
Sharp Switching in Tunnel Transistors and Physics-based Machines for Optimization

Author
Vadlamani, Sri Krishna

Publication Date
2021

Peer reviewed|Thesis/dissertation

Sharp Switching in Tunnel Transistors and Physics-based Machines for Optimization

by

Sri Krishna Vadlamani

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eli Yablonovitch, Chair
Professor David Limmer
Professor Alistair Sinclair

Spring 2021

Sharp Switching in Tunnel Transistors and Physics-based Machines for Optimization

Abstract

Sharp Switching in Tunnel Transistors and Physics-based Machines for Optimization

by

Sri Krishna Vadlamani

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Eli Yablonovitch, Chair

In this thesis, I report my work on two different projects in the field of energy-efficient computation:

Part 1 (Device Physics): Tunnel Field-Effect Transistors (tFETs) are one of the candidate devices being studied as energy-efficient alternatives to the present-day MOSFETs. In these devices, the preferred switching mechanism is the alignment (ON) or misalignment (OFF) of two energy levels or band edges. Unfortunately, energy levels are never perfectly sharp. When a quantum dot interacts with a wire, its energy level is broadened. Its actual spectral shape controls the current/voltage response of such transistor switches, from on (aligned) to off (misaligned). The most common model of spectral line shape is the Lorentzian, which falls off as reciprocal energy offset squared. Unfortunately, this is too slow a turnoff, algebraically, to be useful as a transistor switch. Electronic switches generally demand an ON/OFF ratio of at least a million. Steep exponentially falling spectral tails would be needed for rapid off-state switching. This requires a new electronic feature, not previously recognized: narrowband, heavy-effective mass, quantum wire electrical contacts, to the tunneling quantum states.

Part 2 (Systems Physics): Optimization is built into the fundamentals of physics. For example, physics has the principle of least action, the principle of minimum power dissipation, also called minimum entropy generation, and the adiabatic principle, which, in its quantum form, is called quantum annealing. Machines built on these principles can solve the mathematical problem of optimization, even when constraints are included. Further, these machines become digital in the same sense that a flip–flop is digital when binary constraints are included. A wide variety of machines have had recent success at approximately optimizing the Ising magnetic energy. We demonstrate that almost all those machines perform optimization according to the principle of minimum power dissipation as put forth by Onsager. Moreover, we show that this optimization is equivalent to Lagrange multiplier optimization for constrained problems.

To my wonderful Mummy, Daddy, and Madhav

Contents

List of Figures

Acknowledgments

I am greatly indebted to several people who played crucial roles in sculpting my career over the past few years.

I am deeply grateful to my advisor, Prof. Eli Yablonovitch, for teaching me that a good researcher is brave and is willing to rethink or challenge 'accepted' wisdom. His refreshingly original perspectives on various problems taught me the importance of thinking out of the box. He taught me how to identify and get to the core of any scientific question without beating around the bush. Through our Bell Labs-style 'Journal Club' every week, I learnt how to read scientific literature critically. His incessant enthusiasm for new concepts and ideas is remarkable and is a splendid source of inspiration.

Next, I would like to thank Prof. David Limmer with whom I collaborated quite closely. We had many interesting discussions and I am grateful for the various things I learned from him. I am grateful to Prof. Alistair Sinclair for serving on my thesis committee and giving me important advice at times of need. I also wish to thank Prof. Sayeef Salahuddin for his very kind and generous help at important points and Prof. Ali Javey for serving on my qual committee. I was fortunate to have taken/followed classes offered by a galaxy of excellent teachers that includes, in chronological order, Profs. Eli Yablonovitch, Ali Javey, Laura Waller, Ming Wu, Jonathan Shewchuk, Sayeef Salahuddin, Vivek Subramanian, Sanjam Garg, Jaijeet Roychowdhury, Alistair Sinclair, Sergey Levine, and many others.

I am very grateful to my current and ex-groupmates Sean Hooten, Zunaid Omair, Andy Michaels, Patrick Xiao, Luis Pazos-Outon, and Gregg Scranton for the wonderful time we spent together. I had many illuminating physics discussions with all of them, and I will always remember the fun we had together. I immensely enjoyed working with Patrick on the Ising solver problem and we made great progress through a truly joint effort. I would also like to particularly thank Patrick, Andy, and Sean for their support and encouragement when the going got tough.

I would like to thank members of the Wu, Chang-Hasnain, and Kante groups, Therese George, Shirley Salanio, and the students and staff in the Center for Energy Efficient Electronics Science (E^3S). I gratefully acknowledge funding for my projects from the National Science Foundation (NSF) through the E^3S Center, and the Office of Naval Research (ONR).

Next, I would like to thank my wonderful flatmates of five years and fellow EECS graduate students, Soham Phade and Raaz Dwivedi, for the fun times we shared and the brotherhood we developed. I am also thankful to Kush Bhatia and Neha Gupta for the fun get-togethers we frequently had. I am thankful to my friends across the Bay, Vikranth Reddy, Saideep Reddy, and others.

Finally, I would like to profusely thank my awesome mother, father, and brother. Their constant love, support, and encouragement went a long way in shaping me. I wish to particularly thank my mother for the exceptional amount of time and energy she invested in me and my education. Her immense positive strength, moral support, and encouragement were indispensable in my story so far and helped me scale height after height. Thank you very much!

Chapter 1

Introduction

1.1 Part 1: MOSFETs, subthreshold slope, tunnel transistors

The fundamental building blocks of electronic computing systems today are electronic devices called transistors. The transistor was invented at the Bell Telephone Laboratories in 1947 by Shockley, Bardeen, and Brattain; they were later awarded the Nobel Prize for this world-changing innovation.

Transistors are three-terminal electronic devices that perform two very important functions—**amplification** in analog circuits, and **switching** in digital logic circuits. The most popular transistor today, by far, is the Metal Oxide Semiconductor Field Effect Transistor (MOSFET). The MOSFET switch has truly revolutionized computation; its intrinsically low-power operation and the exponential miniaturization of the device effected by researchers and engineers in the past few decades have given us access to tremendous computational power today in the form of smartphones, laptops, and supercomputers.

The MOSFET has three terminals, the source, the drain, and the gate. When a voltage is applied between the source and the drain terminals, current flows between them through an intermediary channel. The switching mechanism in a MOSFET switch is based on the fact that the conductance of the channel can be modulated using the gate terminal. In the conventional NMOS transistor depicted in Fig. 1.1, a high gate voltage makes the channel highly conducting and causes current to flow between the source and the drain. This is the ON state of the switch. A low gate voltage, on the other hand, turns the transistor OFF because it greatly reduces the channel conductance. The ratio between the ON and OFF state currents is called the ON-OFF ratio of the device. We now look at the switching mechanism in more detail.

The transition of the MOSFET from the ON state to the OFF state is composed of two stages. As the gate voltage V_{gs} is reduced from a high starting value, the source-drain current I_{ds} falls off slowly initially, in the 'above-threshold' region, and then decays rapidly in the 'subthreshold' regime. The subthreshold decay of I_{ds} is in fact exponential with a

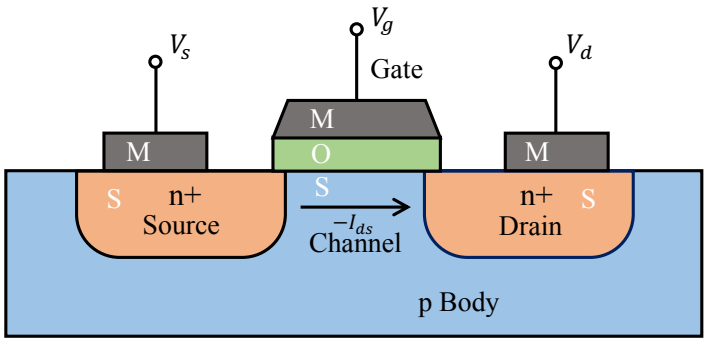

Figure 1.1: A standard MOSFET has three terminals, the source, the drain, and the gate. When a voltage is applied between the source and the drain terminals, current flows between them through an intermediary channel. The conductance of the channel is modulated using the gate terminal.

characteristic voltage that is proportional to kT:

$$I_{ds} \propto e^{qV_{gs}/\eta kT} \tag{1.1}$$

In the above equation, k is the Boltzmann constant, T is the temperature, and η is a dimensionless factor that we will explain in the next paragraph. The mechanism underlying this exponential dependence is depicted in Fig. 1.2.

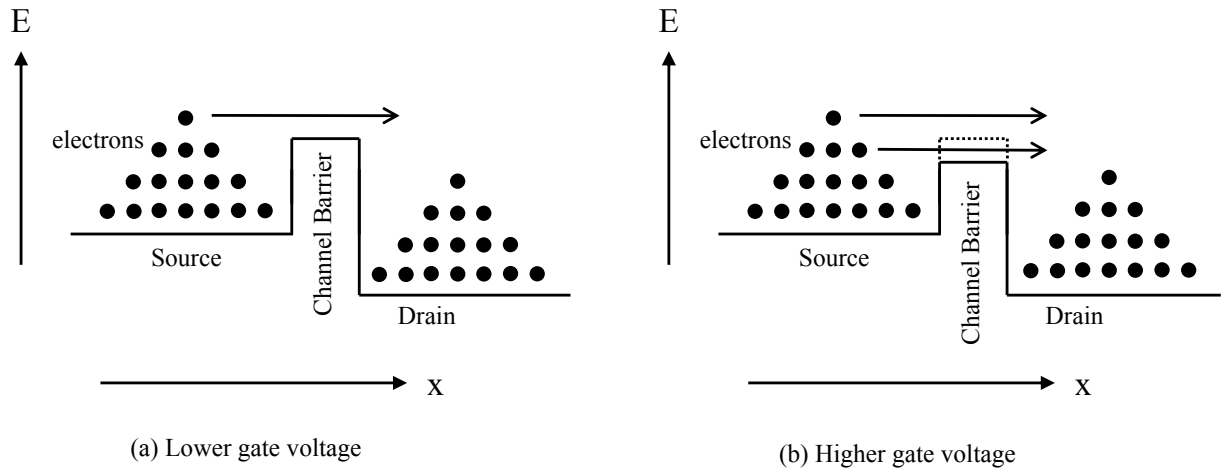

Figure 1.2: (a) When the gate voltage is low, the channel barrier is high, allowing only a few source electrons to the drain. (b) A higher gate voltage pushes the barrier down and allows an exponentially greater number of electrons to flow to the drain.

The MOSFET channel acts as a potential barrier to the many electrons in the $n+$ doped source and drain. Only the electrons in the source that have energies that are higher than the

barrier can participate in electrical conduction. The barrier is high when the gate voltage is zero, and gets gradually lowered as the gate voltage is increased. Since the Fermi occupation number looks like the Boltzmann factor far way from the Fermi level, the number of electrons that are free to participate in conduction as the barrier is lowered increases exponentially with the magnitude of barrier lowering. The fraction of the applied gate voltage that is dropped across the semiconductor body is given by the dimensionless factor $1/\eta$—this same η appears in Eq. (1.1). Since $1/\eta$ is a proper fraction, η is atleast 1.

Eq. (1.1) tells us that, in the subthreshold region at room temperature, the plot of $\log_{10}(I_{ds})$ vs V_{gs} is a straight line on which I_{ds} increases by a factor of 10 for the application of every 60ηmV of gate voltage. The quantity 60ηmV/decade is called the subthreshold slope V_{SS} of the device and it captures the amount of voltage that needs to be applied in order to turn the device ON. Even in the best case of $\eta = 1$, the subthreshold slope of the MOSFET can only be 60mV/decade, and no lower (we get back to this point later in this section). This fundamental statistical mechanical limit on the subthreshold slope of MOSFETs has important implications for the ON state operating voltage and power consumption of circuits as we see next.

Requirements of a good switch

A switch of industrial caliber is required to have three important properties:

1. high ON current—to drive the downstream logic components at high speed,
2. low OFF current, or equivalently, high ON-OFF ratio—to minimize power losses when the device is idle (also called the static power consumption), and
3. low ON voltage, or equivalently, low subthreshold slope—to minimize the total operating power of the circuit (also called the dynamic power consumption).

The MOSFET scores high on the first two criteria. According to the 2020 edition of the International Roadmap for Devices and Systems (IRDS) [1] published by the IEEE, the specifications of the FinFET, the state-of-the-art MOSFET in 2020, were: ON voltage = 0.7 V, V_{SS}=72 mV/decade, V_T=0.345 V, ON current = 484 μA/μm, OFF current = 100 pA/μm. From these numbers, the ON-OFF ratio is nearly 5×10^6. All these numbers are excellent, but the high ON voltage, caused by the high subthreshold slope, means that today's super-densely packed circuits consume large amounts of power in dynamic operation. Since MOSFETs can have a slope of only V_{SS}=60 mV/decade in the best case, the ON voltage will still have to be of the order of 0.5V in order to obtain an ON/OFF ratio of a million.

Getting around this fundamental limit will require groundbreaking new research in MOSFETs or totally new device concepts and mechanisms. One of the most seminal recent ideas in the first category is the use of ferroelectric materials as a **negative capacitance** in MOSFETs, first proposed in 2008 by Salahuddin et al [2]. The negative capacitance helps us achieve $\eta < 1$, pushing the subthreshold slope below $60mV$/decade. Ferroelectric negative capacitance transistors are now being fabricated and studied intensively around the world.

Novel transistor concepts; Tunnel transistors

Over the past few decades, there has also been a sustained research effort in the device research community in the second direction mentioned earlier, that is, in conceptualizing and building new transistors based on novel switching mechanisms different from the thermal barrier hopping mechanism of the MOSFET. **Tunnel transistors** and **electromechanical relays** are examples of novel devices that rely on mechanisms different from thermal barrier hopping for switching. The decade-long Center for Energy Efficient Electronics Science (E3S) was established by the National Science Foundation (NSF) in 2010 to study tunnel transistors and mechanical relays in more detail. My work in Part 1 of this thesis deals with the physics of tunnel transistors. **We show that tunnel transistors of the so-called 'energy-filtering' variety have to be constructed using narrow-band, heavy-effective mass wires to avoid large penalties in the subthreshold regime that arise from fundamental spectroscopy theory.**

1.2 Part 2: Physics-based optimization—Ising solvers

Optimization is ubiquitous in today's world. Everyday applications of optimization range from aerodynamic design of vehicles and physical stress optimization of bridges to airline crew scheduling and delivery truck routing. Furthermore, optimization is also indispensable in machine learning, reinforcement learning, computer vision, and speech processing. Given the preponderance of massive datasets and computations today, there has been a surge of activity in the design of hardware accelerators for neural network training and inference [3].

In this context, physicists and electrical engineers can ask whether physics can be exploited to address optimization and contribute to large-scale industrial problems. Luckily, nature is so structured that almost every fundamental physics principle can be thought of in terms of optimization. Newton's laws, for instance, are recast as an optimization principle in Lagrangian mechanics where the paths of particles are those that minimize a certain time integral. This is the famous 'Principle of Least Action'. In dissipative systems, thermodynamic currents (fluxes) adjust themselves to given thermodynamic forces such that, in steady state, the power dissipation is minimized. This is the 'Principle of Minimum Power Dissipation' (also called Minimum Entropy Generation). In quantum mechanics, the measurement of a system observable leads to the collapse of the state to an eigenvector of the observable. Therefore, the measurement of quantum systems automatically solves the eigenvector problem for arbitrary Hermitian matrices, and it is common knowledge that the eigenvector problem can be recast as a variational problem. The slow reduction of temperature, or annealing, of crystals causes them to gradually adopt crystal configurations of lower and lower potential energy. Physical annealing, therefore, optimizes the potential energy with the use of thermal noise. This fact inspired algorithms like Simulated Annealing several decades ago. Finally, the adiabatic theorem in physics that tells us how the state of a system evolves when the interactions in the system are slowly varied. This principle, which is also referred to as

'quantum annealing' in the quantum computing community, is exploited in several quantum optimizers today.

In due course, the research community may learn how to use each of these principles efficiently to build industrial-scale non-digital optimization accelerators and solvers that offer time and energy benefits in problem solving. Let us consider the principle of minimum power dissipation in dissipative physical systems such as resistive electrical circuits. It was shown by Onsager [4] that the equations of linear systems, like those of resistor networks, can be thought of as minimizing a power dissipation function $f(i_1, i_2, \ldots, i_n)$ for currents i_n in various branches of the resistor network. This means that, in steady state, the currents in the circuit are such that the power dissipation is minimized. If one designs a physical system that has the same functional form for the power dissipation as the mathematical merit function of interest, the circuit itself will find the minimum of the merit function.

Optimization is generally accompanied by constraints. For example, perhaps the constraint is that the final answers must be restricted to be ± 1. Such a digitally constrained optimization produces answers compatible with any digital computer. So, if we could construct a physical system that performs optimization subject to this binary constraint, we could then directly integrate it with the digital pipeline of mainstream computation. As a step in this direction, a series of physics-based machines have been developed in the physics and engineering community in the past decade to approximately solve the **Ising problem**, a difficult optimization problem with binary constraints.

Ising solvers

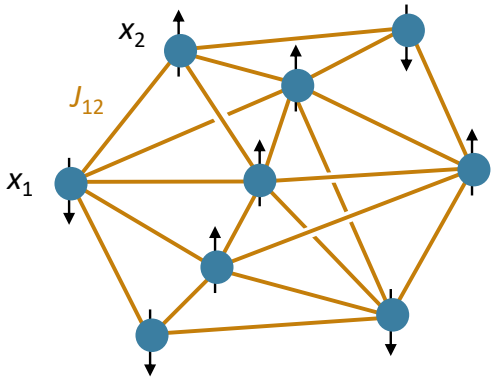

Figure 1.3: The Ising problem setup. A collection of N spins, which can each individually only point up ($x_i = 1$ if the i-th spin is up) or down ($x_i = -1$), interact via pairwise interaction energies $-J_{ij}x_ix_j$. The problem is to find the vector of orientations of the N spins, $\boldsymbol{x}$, that minimizes the sum of all the pairwise energies.

The Ising challenge is to find the minimum energy configuration of a large set of magnets. It is **NP**-hard when the magnets are restricted to two orientations, North pole up or down

[5]. The problem setup is depicted in Fig. 1.3.

Mathematically, let the orientation of the i-th spin be represented by a binary variable x_i which is $+1$ if the spin is pointed along $+z$ and -1 otherwise. The interaction energy between the i-th and j-th spins is specified by the interaction strength J_{ij} and their respective orientations x_i and x_j. Further, the i-th spin could also interact with a local magnetic field specified by h_i. The problem is to find the assignment $\boldsymbol{x}^*$ that minimizes the total Ising magnetic energy:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}:\ x_i=\pm 1\ \forall\ i}{\arg\min} \left(-\sum_{i=1}^{N} h_i x_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij} x_i x_j \right) = \underset{\boldsymbol{x}:\ x_i=\pm 1\ \forall\ i}{\arg\max} \left(\boldsymbol{h^T x} + \frac{1}{2}\boldsymbol{x^T J x} \right) \quad (1.2)$$

This problem is **NP**-hard, and consequently, there is no known polynomial-time algorithm that solves it exactly.

The recently proposed physics-based Ising solvers in the literature were implemented on a range of platforms including coupled optical parametric oscillators, RLC electrical circuits, coupled exciton-polaritons, and silicon photonic coupler arrays. **The main insights in our work are that most of these Ising solvers use hardware based on the Principle of Minimum Power Dissipation and that almost all of them implement the well-known Lagrange Multipliers method for constrained optimization**.

We recognized that the solution procedure in each case is fundamentally based on the application of Lagrange multipliers to the Ising problem with different physical quantities in each system playing the role of the Lagrange multipliers. The machines inherently perform gradient descent on the Lagrange function. For this reason, they can become stuck in local optima. However, they have the advantage that they can possibly converge to good local optima orders-of-magnitude faster, and in a more energy-efficient manner, than conventional digital chips that are limited by latency and the energy cost. Moreover, recent work [6] has demonstrated that these systems are also adaptable toward more advanced techniques for escaping from local optima and cleverly exploring the search space. Our viewpoint is that, these machines can, at the very least, help us perform rapid, energy-efficient searches for local optima, and thus narrow down on the global optimum.

Applications to machine learning

The potential time and energy savings these kinds of optimization solvers seem to offer can have a major impact on contemporary machine learning. Artificial intelligence today runs on neural network models with billions of parameters that are chosen by minimizing the error of the neural network on massive datasets of image, video, or natural language samples. The training process can be very energy-intensive [7] and there are already several start-ups building specialized low-power accelerator hardware for this purpose. Examples of non-digital hardware in the industry and academic literature are the CMOS matrix multipliers of Mythic [8], the optical beam-splitter multipliers of Shen et al. [3], and the highly-efficient

photodetector matrix multipliers of Hamerly et al. [9]. In this spirit, we present a straightforward generalization of physics-based Ising solvers to the problem of least-squares linear regression at the end of Part 2.

Part I

Spectral lineshape theory for tunnel transistors and optical absorption

Chapter 2

Tunnel Transistors and Spectroscopy

Spectroscopy is the branch of physical science that deals with predicting and understanding the frequency response of systems to external perturbations. An example of such a frequency response is the absorption coefficient, $\alpha(\omega)$, that quantifies the extent to which the light incident on a system is absorbed by it at different frequencies. The exact mathematical shape of the frequency response function is commonly called the spectral line shape. A Lorentzian line shape function $1/\left((\omega-\omega_0)^2+(\Gamma/2)^2\right)$, where ω_0 is the peak resonance frequency and Γ is the associated 'linewidth', is often encountered in theoretical calculations. However, this form rarely applies in experimental condensed matter physics. In the solid state, the optical spectrum of electronic transitions usually decays exponentially—this is the so-called Urbach tail, $e^{-(\omega-\omega_0)/\Gamma_U}$, where Γ_U is the characteristic Urbach decay frequency [10]. The importance of the exponential Urbach tail, especially for optical communications, is discussed in Chapter 3.

In this chapter, we introduce another instance in which an exponentially decaying spectrum has important consequences for technology—the switching behavior of the tunnel field-effect transistor (tFET). Tunnel transistors are attractive as an alternative to the MOSFET because they are not fundamentally limited by the thermal subthreshold slope limit of 60 mV/dec [11]. These devices were first conceived of in the late 1970s [12] and several early theoretical proposals and experimental demonstrations followed in the ensuing couple of decades [13–17], although none of them exhibited subthermal subthreshold slopes. In the 2000s and the 2010s, a large variety of device structures and material systems under various thermal, stress, and other physical conditions were investigated [18–35], and we now have several theoretical proposals, backed by simulations [36–42], and experimental demonstrations [43–58] of devices with subthreshold slopes either below 60 mV/dec or very close to it. Unfortunately, subthermal subthreshold slopes are achieved only for a few decades of current or at very low current levels.

In tFETs, there are two mechanisms which can be exploited to obtain improvements in the subthreshold slope [59]. They are: (1) the tunneling distance can be modulated [60–62], (2) the energy levels can be aligned or misaligned between the source and drain—this is sometimes called the energy-filtering mechanism [63, 64].

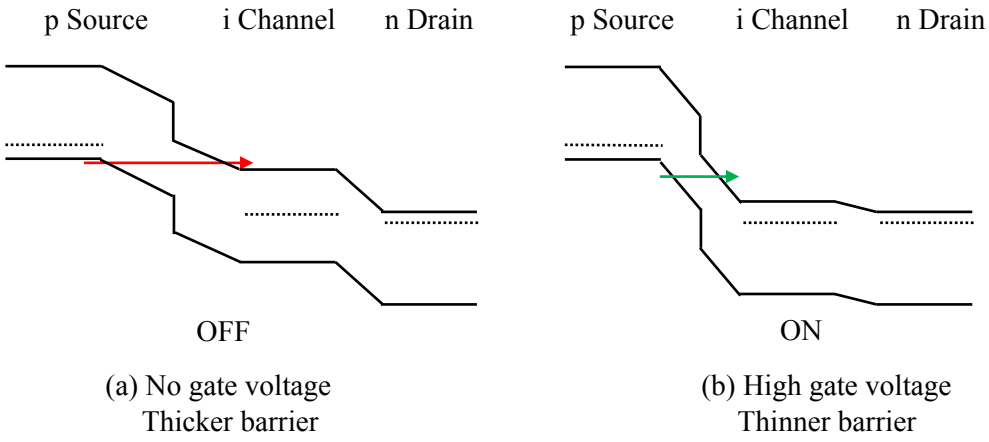

Figure 2.1: Tunnel distance modulating tunnel transistors. The barrier between the source and the channel is large in the absence of gate voltage and the device is OFF (left). A positive gate voltage narrows the barrier and turns the device ON (right).

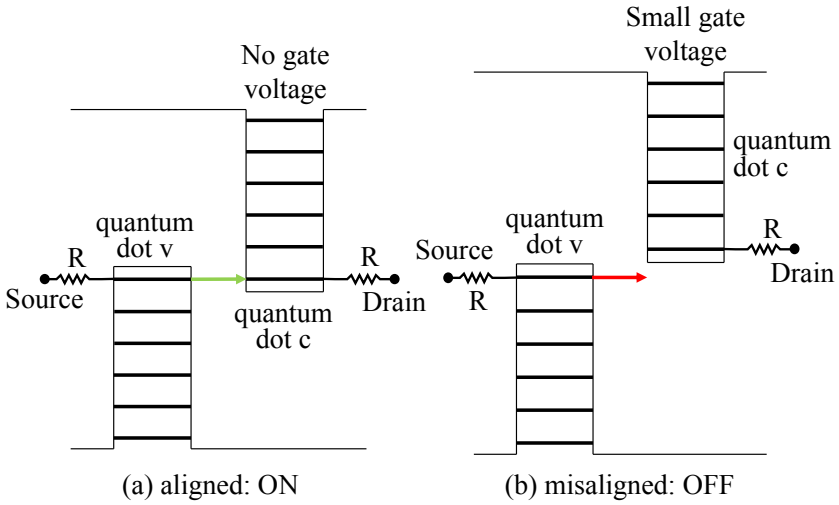

Figure 2.2: Naïve picture of energy-filtering tunnel transistors. A slight offset in the drain and source quantum dots leads to the current between them immediately going to zero so that energy conservation remains satisfied.

Tunnel distance modulation, shown in Fig. 2.1 is more common, but its response to control signals becomes less steep and more gradual at precisely the higher conductance densities that are actually needed for driving wires in a circuit [65]. Thus, the preferred mechanism for switching is energy filtering—alignment (ON) or misalignment (OFF) of two energy levels or band edges.

In energy filtering, the channel comprises two quantum dots, a source dot coupled to the source contact and a drain dot coupled to the drain, as shown in Fig. 2.2. Agarwal et al. [66] discuss a variety of other geometries for energy-filtering tunnel transistors, but the dot-to-dot (0D-0D) tunneling case suffices for our purposes. Returning to our picture, the drain dot is also capacitively coupled to a gate terminal, not shown in the figure. Since the dots are nanoscopic, the energy levels in the dots are quantized. Adjusting the gate voltage allows us to control the alignment of the energy levels, and consequently, the state of the switch—when the levels are aligned, current flows and the device is ON, while misaligned levels block the flow of current and turn the device OFF.

Unfortunately, energy levels are never perfectly sharp. The exact shape of the spectral line shape plays a crucial role in determining the current/voltage response of such transistors as they are switched from on (aligned) to off (misaligned). In other words, the spectral shape of the energy levels decides the subthreshold slope of tFETs. This is not widely recognized, though some prior work has been done in this regard [67, 68]. In our work, we study the problem of spectral line shape in tunnel transistors and deduce criteria that are necessary for a tFET to have exponential subthreshold behavior. The remainder of this chapter is the basis of my IEEE publication [69].

2.1 The performance of energy-filtering transistors depends on lineshape

Essence of this section: In this section, we argue that the performance evaluation of quantum dot energy-filtering transistors has to take into account the so-called **energy broadening** or **spectral lineshapes** of the energy levels in the quantum dots. The concept of a spectral lineshape is introduced and discussed and the procedure to compute it is presented. The exact lineshape determines the amount of leakage current when the device is turned off.

The energy-filtering tunnel transistor, at first sight, seems to achieve the goal of sharp switching with ease. When no gate voltage is applied on the drain-side quantum dot, the energy levels of the source-side and the drain-side quantum dots are aligned and there is current flow from one side to the other. This is the ON state. To turn the device off, one only requires the application of an infinitesimally small gate voltage to break the alignment of the energy levels in the two quantum dots and shut off the flow of current completely. The device is then in the OFF state. These situations are depicted in Fig. 2.2. If this were the case, one would be able to design switches that turned on and off on the application of a negligible amount of voltage, rendering the present-day MOSFET completely obsolete.

Unfortunately, this simplistic interpretation of the tunnel transistor mechanism, though very appealing, is incomplete. Since the quantum dots are in close contact with the the source and drain wires, their energy levels hybridize with those of the metallic wires to produce new stationary states that are delocalized over both the dot and the wire. In other

words, the spatially confined wavefunctions that were originally the stationary states of the isolated quantum dot, with well-defined sharp energies, are now linear superpositions of the new stationary states of the joint dot-wire system. Consequently, the spatially confined quantum dot wavefunctions no longer have well-defined energies. Instead, one would obtain a distribution of values when the energy of an electron in such a state is measured. This effect is called 'energy-level broadening' or 'spectral broadening'. Spectral broadening is by no means a surprising phenomenon, and the exact same effect occurs in RLC electrical circuits and damped optical cavities. A perfectly lossless LC circuit supports oscillations precisely at one frequency. However, the addition of a resistor, R, to the circuit allows it to support other neighboring frequencies too, albeit at lower strengths than the natural frequency. Similarly, a Fabry-Pérot cavity that is perfectly sealed off from its surroundings supports optical oscillations of only those special frequencies that satisfy the boundary conditions at the walls of the cavity. Once the cavity is exposed to optical loss, either through the introduction of absorbing material in the cavity or by making the cavity walls permeable, its frequency response broadens along the frequency axis and the cavity begins to support many more frequencies than before. In summary, the energy filtering response function of the drain-side quantum dot is perfectly sharp only in the complete absence of external disturbances. Once the quantum dot is augmented with a channel of electron escape, in the form of a metallic wire, the energy-filter response gets broadened.

The broadening of the energy-filtering response has spectacular implications for our earlier näive understanding of the working of an energy-filtering tunnel transistor. The transistor conducts current through the tails of the broadened quantum dot energy levels even after the levels have been displaced through the application of a gate voltage. This is shown in Fig. 2.3. The exact mathematical shape of the broadening, which we shall call 'lineshape', determines the nature of the current reduction as the quantum dot energy levels become more and more misaligned. That is, the drain-current-vs-gate-voltage ($I_d - V_g$) characteristics of a tunnel transistor can be calculated precisely if one has access to the quantum dot lineshapes.

Computation of lineshapes

In this subsection, we derive an expression for the lineshape of a quantum dot energy level that is in contact with a metallic reservoir. That is, we want to compute the energy distribution of an originally unperturbed quantum dot state in the new joint dot-wire system.

For the sake of clarity, we shall focus only on the broadening of the lowest energy eigenstate of the quantum dot. To find the broadening of this level, one simply has to place an electron in this state at $t = 0$ and observe the evolution of the electron wavefunction as time t progresses. It turns out that the electron decays gradually into the continuum of wire states from the dot. The Fourier transform of this time-domain decay of electron amplitude from the dot to the wire gives us the energy distribution (or lineshape) of the initial state. We remark here that this procedure is again exactly analogous to the way electromagnetic cavity lineshapes are obtained from the details of the time-domain decay of electric fields

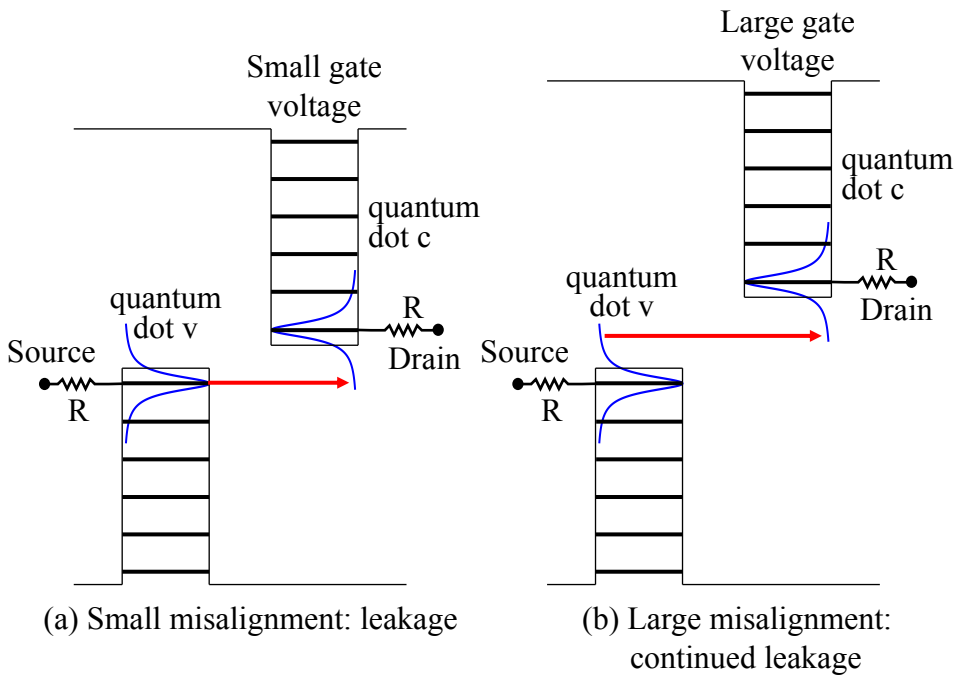

Figure 2.3: More realistic picture of energy-filtering tunnel transistors. Current keeps flowing between the source and drain dots even after they are offset by a gate voltage because of spectral broadening of the energy levels. Current flows between the energy tails and contributes to off-state leakage.

inside the cavity. We explain next why the Fourier transform of the time-domain decay of the electron wavefunction's amplitude on the initial quantum dot state should give us that initial state's lineshape.

Let this initial state (the lowest eigenstate of the unperturbed quantum dot) be denoted $|\psi\rangle$. Since $|\psi\rangle$ is no longer a stationary state of the joint dot-wire system, an electron that starts out in $|\psi\rangle$ at $t = 0$ evolves in time as $|\psi(t)\rangle$. Let the true eigenstates of the joint dot-wire system by denoted $|\phi_i\rangle$. Since the $|\phi_i\rangle$ form a basis, we can now express $|\psi\rangle$ in terms of them, using a set of complex coefficients c_i:

$$|\psi\rangle = \sum_i c_i |\phi_i\rangle \tag{2.1}$$

It follows that $|\psi(t)\rangle$ can be expressed as:

$$|\psi(t)\rangle = \sum_i c_i e^{-iE_i t/\hbar} |\phi_i\rangle \tag{2.2}$$

It is clear that $|\psi(0)\rangle = |\psi\rangle$. From the above formulae, we note that the energy distribution of $|\psi(t)\rangle$ doesn't change with time. The total probability that the state $|\psi(t)\rangle$ yields a value

E upon the performance of an energy measurement is equal to $\sum_{i:E_i=E} |c_i|^2$, independent of the time t.

A function that naturally captures information about these total probabilities is the **autocorrelation function** for state evolution. The autocorrelation function, which we shall denote by $C(t_1, t_2)$, is simply defined as the dot product between the states of the electron at two different times $\langle\psi(t_1)|\psi(t_2)\rangle$. Setting $t_1 = t$ and $t_2 = 0$, we get:

$$C(t, 0) = \langle\psi(t)|\psi(0)\rangle = \sum_i |c_i|^2 e^{iE_i t/\hbar} \tag{2.3}$$

The coefficient of $e^{iEt/\hbar}$ in the above expression is given by $\sum_{i:E_i=E} |c_i|^2$, which is exactly what we were looking for. To extract this information as an energy distribution, or lineshape, we perform the Fourier transform on $C(t, 0)$ to obtain $\chi(\omega)$:

$$\chi(\omega) = \int_{-\infty}^{\infty} C(t, 0) e^{-i\omega t} dt = 2\pi \sum_i |c_i|^2 \delta\left(\omega - \frac{E_i}{\hbar}\right) \tag{2.4}$$

Choosing a change of variables $\omega = E/\hbar$, we get the following distribution over energies:

$$L(E) = \frac{\chi(E/\hbar)}{2\pi\hbar} = \sum_i |c_i|^2 \delta\left(E - E_i\right) \tag{2.5}$$

For a discrete spectrum of true eigenenergies, E_i, we observe that $L(E)$ is a train of Dirac delta functions with the total probability of each energy being the coefficient of a Dirac delta function centered at that energy. For continuous spectra, $L(E)$ is a probability distribution function over energy.

Now that we agree that $L(E)$ gives us the lineshape we have been looking for, how do we compute $\langle\psi(t)|\psi(0)\rangle$ for the dot-wire system in the tunnel transistor? For this purpose, let us move from the basis of eigenstates of the joint dot-wire system to that of the eigenstates of the separate unperturbed dot and wire. This 'unperturbed' basis consists of $|\psi\rangle$ contributed by the dot and the set $\{|k\rangle\}$ contributed by the wire. k here represents the quantum numbers, such as k-vector and band index, that are needed to uniquely identify the eigenstates of the unperturbed wire. It has to be kept in mind that these states are not eigenstates of the joint dot-wire system. The unperturbed wire states $\{|k\rangle\}$ are orthogonal to the unperturbed quantum dot state $|\psi\rangle$. Therefore, in this 'uncoupled' basis, we have:

$$|\psi(0)\rangle = |\psi\rangle \tag{2.6}$$

$$|\psi(t)\rangle = d_1(t)\,|\psi\rangle + \sum_k d(k, t)\,|k\rangle \tag{2.7}$$

$$\langle\psi(t)|\psi(0)\rangle = d_1^*(t) \tag{2.8}$$

where $d_1(t)$ is the coefficient of the wavefunction at time t on $|\psi\rangle$ and $d(k, t)$ is the coefficient of the wavefunction at time t on $|k\rangle$. Therefore, $C(t, 0)$ is simply the amplitude of the electron wavefunction, at time t, on the starting state (but with complex conjugation)!

The conclusion of this section is that the energy distribution of a state is the Fourier transform of its autocorrelation function. The autocorrelation function is obtained by placing an electron in the state of interest at $t = 0$ and noting down the time-evolution of the electron wavefunction's amplitude on that starting state, $d_1(t)$, and taking its complex conjugate.

2.2 Standard derivation of Lorentzian lineshapes; Fermi's Golden Rule

Essence of this section: In this section, we show how Lorentzian lineshapes are derived from Schrödinger's equation using standard approximations. The key point is that Lorentzian lineshapes are inevitable if the electron tunnels from a quantum dot into a wire with a **broad density of states**.

Next, we start from Schrödinger's equation and derive the autocorrelation function, $C(t, 0)$, for an electron that is initially placed on the quantum dot and leaks into the wire as time passes. We will perform some standard approximations along the way, and finally derive a purely exponential time-domain decay of $d_1(t)$, or equivalently, $C(0, t)$. The derivation in this section is borrowed from the excellent Quantum Mechanics textbook by Cohen-Tannoudji et al. [70].

Setup

Let the Hamiltonian of the unperturbed quantum dot and wire that are initially placed far apart be H_0. The unperturbed dot state, $|\psi\rangle$, and the set of unperturbed wire states, $\{|k\rangle\}$, are eigenstates of H_0. They satisfy:

$$H_0 |\psi\rangle = E_1 |\psi\rangle \tag{2.9}$$

$$H_0 |k\rangle = E(k) |k\rangle \tag{2.10}$$

where the energy of the unperturbed state $|\psi\rangle$ is E_1 and the energy of the unperturbed wire state $|k\rangle$ is $E(k)$.

Since the dot and the wire need to be in close proximity for a working device, the perturbed Hamiltonian has an additional interaction term, V. The net Hamiltonian is then given by:

$$H = H_0 + V \tag{2.11}$$

We shall assume that the interaction V has non-zero matrix elements only between the dot state and a wire state. That is, $\langle\psi|V|\psi\rangle = 0$ and $\langle k|V|k\rangle = 0$ for every k. Only the elements $\langle k|V|\psi\rangle$ are allowed to be non-zero.

Schrödinger equation

As earlier, let the electron wavefunction at time t be $|\psi(t)\rangle$, with the initial condition being $|\psi(0)\rangle = |\psi\rangle$. The unperturbed wire states are the set $\{|k\rangle\}$. The Schrödinger equation is:

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{i}{\hbar}|\psi(t)\rangle \tag{2.12}$$

Using the uncoupled basis to expand $|\psi(t)\rangle$ as in Eq. 2.7, we get the following for the coefficients $d_1(t)$ and $d(k,t)$:

$$\frac{d}{dt}d_1(t) = -\frac{i}{\hbar}E_1 d_1(t) - \frac{i}{\hbar}\int dk\ d(k,t)\,\langle\psi|V|k\rangle \tag{2.13}$$

$$\frac{d}{dt}d(k,t) = -\frac{i}{\hbar}E(k)d(k,t) - \frac{i}{\hbar}d_1(t)\,\langle k|V|\psi\rangle \tag{2.14}$$

We now make a change of variables that is popularly called the 'Interaction Picture'. It is nothing but the use of an integrating factor. Setting $\tilde{d_1}(t) = d_1(t)e^{iE_1t/\hbar}$ and $\tilde{d}(k,t) = d(k,t)e^{iE(k)t/\hbar}$, we get:

$$\frac{d}{dt}\tilde{d_1}(t) = -\frac{i}{\hbar}\int dk\ \langle\psi|V|k\rangle\,\tilde{d}(k,t)e^{i(E_1-E(k))t/\hbar} \tag{2.15}$$

$$\frac{d}{dt}\tilde{d}(k,t) = -\frac{i}{\hbar}\,\langle k|V|\psi\rangle\,\tilde{d_1}(t)e^{-i(E_1-E(k))t/\hbar} \tag{2.16}$$

The variables $\tilde{b}_1(t)$ and $\tilde{b}(k,t)$ can also be interpreted as 'slowly varying amplitudes' because they are what remain when the fast natural oscillatory phase factors are removed from $b_1(t)$ and $b(k,t)$. Integrating (2.16) from $t=0$ to $t=t$, and recalling that $d(k,0) = \tilde{d}(k,0) = 0$ for all k, we get:

$$\tilde{d}(k,t) = -\frac{i}{\hbar}\,\langle k|V|\psi\rangle\int_0^t d\tau\ e^{-i(E_1-E(k))\tau/\hbar}\tilde{d_1}(\tau) \tag{2.17}$$

Substituting (2.17) into (2.15), and setting $\tau = t - \tau$ in the integral, we obtain:

$$\frac{d}{dt}\tilde{d_1}(t) = -\frac{1}{\hbar^2}\int_0^t d\tau\ \int dk\ |\,\langle\psi|V|k\rangle\,|^2 e^{i(E_1-E(k))\tau/\hbar}\tilde{d_1}(t-\tau) \tag{2.18}$$

$$\implies \frac{d}{dt}\tilde{d_1}(t) = -\int_0^t d\tau\ M(\tau)\tilde{d_1}(t-\tau) = -\int_0^t d\tau\ M(t-\tau)\tilde{d_1}(\tau) \tag{2.19}$$

On the right side above, we have a convolution between a memory function, given by $M(t) = \frac{1}{\hbar^2}\int dk\ |\,\langle\psi|V|k\rangle\,|^2 e^{i(E_1-E(k))t/\hbar}$, and $\tilde{d_1}(t)$. Let us make the simplifying assumption that the matrix element $\langle\psi|V|k\rangle = V(E(k))$ depends only on the energy $E(k)$ of the unperturbed state $|k\rangle$. This assumption doesn't change the essential nature of the calculation, and it can be easily removed with some extra notation. Then, we can rewrite the memory function as:

$$M(t) = \frac{1}{\hbar^2}\int dE\ \rho(E)|V(E)|^2 e^{i(E_1-E)t/\hbar} = r(t) + is(t) \tag{2.20}$$

where $\rho(E)$ is the density of states in the wire, and $r(t)$ and $s(t)$ are the real and imaginary parts of $M(t)$.

Infinitely broad final density of states leads to pure exponential decays of amplitude

We shall first solve Eq. (2.19) for the simplest case of $\rho(E) = \rho$ and $|V(E)| = |V|$ being constant and spread out over an infinite energy range, from $E = -\infty$ to $E = -\infty$. The memory function $M(t)$ is then:

$$M(t) = \frac{1}{\hbar^2}\int_{-\infty}^{\infty} dE\ \rho(E)|V(E)|^2 e^{i(E_1-E)t/\hbar} \tag{2.21}$$

$$= \frac{\rho|V|^2}{\hbar^2}\int_{-\infty}^{\infty} dE\ e^{i(E_1-E)t/\hbar} = \frac{2\pi}{\hbar}\rho|V|^2\delta(t) = \Gamma\delta(t) \tag{2.22}$$

where we have defined $\Gamma = \frac{2\pi}{\hbar}\rho|V|^2$. Plugging this into Eq. (2.19) gives:

$$\frac{d}{dt}\tilde{d}_1(t) = -\frac{2\pi}{\hbar}\rho|V|^2\int_0^t d\tau\ \delta(t-\tau)\tilde{d}_1(\tau) = -\frac{\pi}{\hbar}\rho|V|^2\tilde{d}_1(t) = -\Gamma\tilde{d}_1(t)/2 \tag{2.23}$$

where the integral of the Dirac delta only has half the strength since the integral doesn't straddle $\tau = t$. Thus $d_1(t)$, the amplitude of the electron's wavefunction on the dot state, decays in a purely exponential manner:

$$d_1(t) = e^{-\Gamma t/2}e^{-iE_1 t/\hbar} \tag{2.24}$$

Markov and time-independent coefficient approximations also lead to pure exponential decays

We can derive a similar exponential formula in a more general context where the energy range is not infinite but simply much broader compared to $\hbar\Gamma$, the energy unit we just encountered that captures both the strength of the perturbation and the magnitude of the available density of states for tunneling.

Let us assume that the wire density of states $\rho(E)$ and the matrix element $|V(E)|$ are broad along the E axis, that is, that they are of appreciable magnitude over a large range of energy values ΔE. If ΔE is sufficiently large, it is enough for t to cross a certain small threshold time ($t > \Delta t$) to render the memory function $M(t)$ approximately equal to 0. The reason for this is the 'high' frequency oscillation $e^{i(E_1-E)t/\hbar}$ that modulates $\rho(E)|V(E)|^2$ in the integrand in Eq. (2.20). The product, which is still highly oscillatory, roughly integrates out to 0, making $M(t) \approx 0$ for $t > \Delta t$. This assumption that the final density of states is broad enough, resulting in a sharp memory function, is called the **Markov approximation**.

Since the memory function $M(t)$ is non-zero only over a small time interval $[0, \Delta t]$, the memory inherent in (2.19) is short. If we further assume that V is only a weak perturbation, we can safely assert that the slowly varying amplitudes $\tilde{d}_1(t)$ and $\tilde{d}(k,t)$ change very little in the time span Δt in which the memory function is non-zero. This means $\tilde{d}_1(t-\tau) \approx \tilde{d}_1(t)$

in (2.19) and we can pull it out of the integral. We get:

$$\frac{d}{dt}\tilde{d_1}(t) = -\left(\int_0^t d\tau\ M(\tau)\right)\tilde{d_1}(t) \tag{2.25}$$

From (2.19) to (2.25), the right hand side has reduced from a linear convolution to a memoryless linear term with a time-dependent coefficient.

The next approximation in the standard approach involves turning the time-dependent coefficient into a time-independent quantity. We are not aware of a name for this approximation, and shall simply call it the **'time-independent coefficient approximation'**. The approximation involves asserting that one is interested in the decay of the electron at large enough times t, and not times that are extremely close to $t = 0$. In that case, since $M(t)$ is anyway approximately zero for $t > \Delta t$, we can replace the integral in (2.25) from $\tau = 0$ to $\tau = t$ with the full integral of $M(\tau)$ from $\tau = 0$ to $\tau = \infty$. This integral can be computed as follows:

$$\int_0^\infty d\tau\ M(\tau) = \frac{1}{\hbar^2}\int dE\ \rho(E)|V(E)|^2 \int_0^\infty d\tau\ e^{i(E_1-E)\tau/\hbar} \tag{2.26}$$

$$= \frac{1}{\hbar^2}\int dE\ \rho(E)|V(E)|^2 \left(\pi\hbar\delta\left(E_1 - E\right) + \frac{i\hbar}{E_1 - E}\right) \tag{2.27}$$

$$= \left(\frac{\pi}{\hbar}\rho(E_1)|V(E_1)|^2 + i\frac{1}{\hbar}\int dE\ \frac{\rho(E)|V(E)|^2}{E_1 - E}\right) \tag{2.28}$$

$$= \left(\frac{\Gamma}{2} + i\frac{\Delta}{\hbar}\right) \tag{2.29}$$

where we introduced a rate constant Γ and an energy offset Δ defined as follows:

$$\Gamma = \frac{2\pi}{\hbar}\rho(E_1)|V(E_1)|^2 \tag{2.30}$$

$$\Delta = \int dE\ \frac{\rho(E)|V(E)|^2}{E_1 - E} \tag{2.31}$$

Plugging (2.29) into (2.25), we get the following simple form:

$$\frac{d}{dt}\tilde{d_1}(t) = -\left(\frac{\Gamma}{2} + i\frac{\Delta}{\hbar}\right)\tilde{d_1}(t) \tag{2.32}$$

Recalling that $\tilde{d_1}(0) = 1$, this yields:

$$d_1(t) = e^{-\left(\frac{\Gamma}{2} + i\frac{\Delta}{\hbar}\right)t} e^{-iE_1 t/\hbar} \tag{2.33}$$

Exponential decay holds as $t \to \infty$ irrespective of the memory function: It should be kept in mind that, if the memory function is not 'thin' enough, the time-independent coefficient approximation may not hold at even moderately large times t. However, since a physical memory function decays to 0, the approximation will always hold as $t \to \infty$. This implies that exponential decay of amplitude, with the rate given by Fermi's golden rule, is always accurate at sufficiently large times.

Getting the full autocorrelation function from one-sided time decay

We have only computed $C(0,t) = d_1(t)$ for $t \geq 0$ so far. This result can be extended to $t < 0$ in a straightforward fashion if one uses the unitary property of time-evolution in quantum mechanics. Let $U(t)$ be the unitary operator that propagates a wavefunction forward in time by an amount of time t. Then, we have:

$$C(0,t) = \langle\psi(0)|\psi(t)\rangle = \langle\psi(0)|U(t)|\psi(0)\rangle \tag{2.34}$$

For negative time, we have the following:

$$C(0,-t) = \langle\psi(0)|U(-t)|\psi(0)\rangle = \langle\psi(0)|U^{\dagger}(t)|\psi(0)\rangle = (\langle\psi(0)|U(t)|\psi(0)\rangle)^{*} \tag{2.35}$$

where t is a positive quantity. Therefore, for $t < 0$, we have $C(0,t) = e^{\left(\frac{\Gamma}{2} - i\frac{\Delta}{\hbar}\right)t} e^{-iE_1 t/\hbar}$ which results in the following formula for $C(t,0)$ for all time t:

$$C(t,0) = e^{-\frac{\Gamma}{2}|t|} e^{i\frac{\Delta}{\hbar}t} e^{iE_1 t/\hbar} \text{ for all } t \tag{2.36}$$

Lorentzian lineshape

The calculation is completed by taking the Fourier transform of Eq. (2.36). The phase factors on the right hand side of Eq. (2.36) simply cause the lineshape to shift along the E axis. Since the amplitude part, $e^{-\frac{\Gamma}{2}|t|}$, is real and even in t, its Fourier transform can be computed as:

$$\int_{-\infty}^{0} e^{-\frac{\Gamma}{2}|t|} e^{-i\omega t} dt + \int_{0}^{\infty} e^{-\frac{\Gamma}{2}|t|} e^{-i\omega t} dt = \int_{0}^{\infty} e^{-\frac{\Gamma}{2}|t|} e^{i\omega t} dt + \int_{0}^{\infty} e^{-\frac{\Gamma}{2}|t|} e^{-i\omega t} dt \tag{2.37}$$

$$= 2\mathrm{Re}\left\{\int_{0}^{\infty} e^{-\frac{\Gamma}{2}t} e^{-i\omega t} dt\right\} \tag{2.38}$$

$$= \frac{\Gamma}{\omega^2 + \left(\frac{\Gamma}{2}\right)^2} \tag{2.39}$$

Plugging in the shifts induced by the phase factors, we get for the lineshape:

$$L(E) = \frac{1}{\pi} \frac{\frac{\hbar\Gamma}{2}}{\left(E - (E_1 + \Delta)\right)^2 + \left(\frac{\hbar\Gamma}{2}\right)^2} \tag{2.40}$$

Eq. (2.40) is the famed **Lorentzian lineshape** that is central to spectroscopy. It is peaked near the original unperturbed energy level E_1—with a small shift Δ due to the interaction with the wire energy levels—and falls off as $1/E^2$ in the tails on both sides. We will see next that this 'heavy' tail is detrimental to the $I_d - V_g$ performance of tunnel transistors.

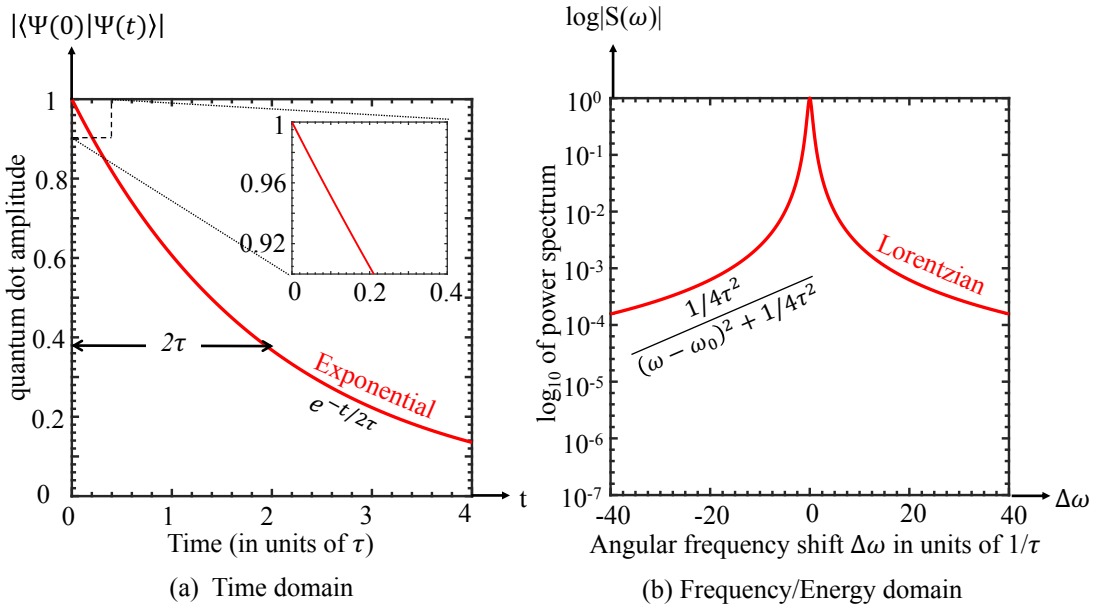

Figure 2.4: (a) The pure exponential decay of amplitude from a quantum dot level into a wire with a broad density of states. (b) The lineshape is proportional to the Fourier transform of the time domain decay. The Fourier transform of a pure exponential time decay is a Lorentzian function in frequency. Lorentzians have slow decay $1/\omega^2$ at large ω.

2.3 $I_d - V_g$ characteristics from lineshape

Essence of this section: In this section, we show how to compute the current flow I_d between two broadened quantum dots when a drain-source voltage V_{ds} is applied across them and a gate voltage V_g is applied on the second quantum dot. The final current expression takes into account the lineshapes of the two levels, the strength of the interaction between them, the offset between the two levels caused by the gate voltage, and their electron occupations which are determined by the offset caused by the drain-source voltage between the drain and source quasi-Fermi levels.

We now tackle the problem of computing the $I_d - V_g$ characteristics of a tunnel transistor from the lineshapes of the quantum dots that comprise it. To do this, we refer to Fig. 2.5. We have so far dealt with the broadening of the quantum dot levels that is induced by the dot-wire interaction X. Current flows through this device from the source-side quantum dot to the drain-side quantum dot due to the interaction M. Since the effect of X is already included in our lineshapes, the inputs to our current calculation will be M, $L_1(E)$, and $L_2(E)$, the latter two functions being the lineshapes of the source-side and the drain-side quantum dots respectively.

For convenience, we will assume that the Fermi levels in the source and drain are such

that the source-side dot is fully filled and the drain-side dot is fully empty. This assumption is removed towards the end of this section.

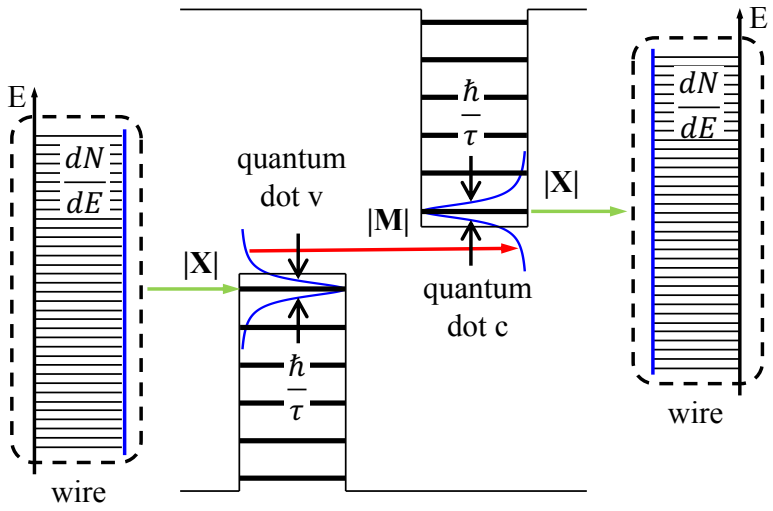

Figure 2.5: The full tunnel transistor. The source wire is coupled to the source quantum dot via the interaction X as is the drain wire to its own quantum dot. This interaction X led to the broadening of the quantum dot's energy levels. The quantum dots themselves interact via the coupling M and it is this interaction that leads to current in the device.

Sharp source dot into broadened drain dot

Let us inspect the case of a single electron leaking from a sharp source dot to a broadened drain dot through the interaction M. The decay of the electrons wavefunction's amplitude on the source-dot can be computed in exactly the same way as discussed in the previous section. M would take the place of X, and the lineshape of the drain dot, $L_2(E)$, would take the place of the final density of states $\rho(E)$. The amplitude on the source dot falls off as:

$$d_1(t) = e^{-\left(\frac{\Gamma}{2}+i\frac{\Delta}{\hbar}\right)t} e^{-iE_1 t/\hbar} \tag{2.41}$$

where $\Gamma = \frac{2\pi}{\hbar} L_2(E_1)|M(E_1)|^2$ and $\Delta = \int dE\ \frac{L_2(E)|M(E)|^2}{E_1 - E}$. That means the probability that the electron is on the source-side falls off with t as $e^{-\Gamma t}$. Interpreting this further:

$$P(\text{electron is on the drain side at time } t) = 1 - e^{-\Gamma t} \tag{2.42}$$

$$\implies P(\text{electron left the source side before time } t) = 1 - e^{-\Gamma t} \tag{2.43}$$

$$\implies P(\text{electron left the source side between time } t \text{ and } t + dt) = \Gamma e^{-\Gamma t} dt \tag{2.44}$$

The last line of reasoning essentially tells us that the time at which the electron leaves the source and goes to the drain is a random variable that is exponentially distributed with

parameter Γ. The mean of an exponential distribution with parameter Γ is simply $1/\Gamma$. Therefore, $1/\Gamma$ is the average time taken by an electron in the source-dot to tunnel into the drain-dot. Since one unit of charge is transported in time $1/\Gamma$ on average, and there are two spin states per energy level, the average current is:

$$I = \frac{-2e}{1/\Gamma} = -2\frac{2\pi e}{\hbar} L_2(E_1)|M(E_1)|^2 \tag{2.45}$$

Broadened source dot into broadened drain dot

For the case of electron tunneling from a broadened source dot to a broadened drain dot, we write down the tunneling current originating from an energy range dE_1 around E_1 on the source-side to the drain dot, and then integrate it over all E_1 to obtain the full current. Since the 'number of states' in the source-dot between energies E_1 and $E_1 + dE_1$ is $L_1(E_1)dE_1$, we multiply Eq. (2.45) with this amount to get:

$$I(e^- \text{ originating between } E_1 \ \&\ E_1 + dE_1) = \frac{-2eL_1(E_1)dE_1}{1/\Gamma} \tag{2.46}$$

$$= -dE_1 \ 2\frac{2\pi e}{\hbar} L_1(E_1)L_2(E_1)|M(E_1)|^2 \tag{2.47}$$

Therefore, the full current is:

$$I_{\text{full}} = -2\frac{2\pi e}{\hbar} \int dE \ L_1(E)L_2(E)|M(E)|^2 \tag{2.48}$$

Once simply tacks the factor $(f_1(E) - f_2(E))$ onto the integrand, where $f_1(E)$ and $f_2(E)$ are the Fermi occupations of the source and drain respectively, to take partial occupation of the quantum dots into account. Further, the gate voltage V_g serves to offset the drain-dot levels from the source-dot levels.

$$I_{\text{full}} = -2\frac{2\pi e}{\hbar} \int dE \ L_1(E)L_2(E + eV_g)|M(E)|^2 \left(f_1(E) - f_2(E)\right) \tag{2.49}$$

Equation (2.49) is the important equation that takes as input the interaction M and the two lineshapes $L_1(E)$, $L_2(E)$, and outputs the $I_d - V_g$ characteristic of the device. When the Lorentzian lineshape obtained in Eq. (2.40) is plugged into Eq. (2.49), a Lorentzian shaped $I_d - V_g$ plot is obtained. This is disastrous for devices whose primary aim is to beat MOSFETs. A Lorentzian $I_d - V_g$ characteristic not only fails to surpass the MOSFET's subthreshold slope of 60mV/dec, it is not even exponentially shaped in the first place! The sequence of derivations we have seen so far seem to indicate that tunnel transistors are doomed to failure.

2.4 Deriving non-Lorentzian lineshapes — Exact solution using Laplace transforms

Essence of this section: In this section, we derive the exact formula for the lineshape of a quantum dot level coupled to a wire that has an arbitrary density of states via an arbitrary coupling interaction.

In order to be able to design sharper-switching tunnel transistors, with exponential lineshapes and $I_d - V_g$ curves, we have to first develop a deeper understanding of the origin of Lorentzian lineshapes. We have already seen that coupling the dot to a wire possessing a broad density of states leads to the validity of the Markov and time-independent coefficient approximations and eventually to purely exponential time-domain decays and frequency-domain Lorentzians. An immediate course of action would then be to study the dot-wire system when the wire has a narrower density of states, that is, when most of its states lie in a restricted range of energies.

One could imagine that solving Eq. (2.19), reproduced below, exactly would lead to expressions for $\tilde{d_1}(t)$ that would help us deduce the conditions that the matrix element $\mathbf{X}$ and the wire density of states $\rho(E)$ should satisfy in order for exponential lineshapes to emerge.

$$\frac{d}{dt}\tilde{d_1}(t) = -\int_0^t d\tau\ M(\tau)\tilde{d_1}(t-\tau) \tag{2.50}$$

To that end, we now present a standard exact solution of Eq. (2.50) using Laplace transforms. The derivation in this section is adapted from the textbook on atom-photon interactions by Cohen-Tannoudji et al. [71].

We start out by recalling the definition of the Laplace transform $f^L(s)$ of a function $f(t)$:

$$f^L(s) = \int_0^\infty dt\ f(t)e^{-st} \tag{2.51}$$

Representing the Laplace transform of $\tilde{d_1}(t)$ by $\tilde{d_1}^L(s)$ and that of $M(t)$ by $M^L(s)$, we perform the Laplace transform on both sides of Eq. (2.50):

$$s\tilde{d_1}^L(s) - \tilde{d_1}(0) = -M^L(s)\tilde{d_1}^L(s) \tag{2.52}$$

$$\tilde{d_1}^L(s) = \frac{1}{M^L(s)+s} \tag{2.53}$$

We used the fact that the initial condition $\tilde{d_1}(t=0) = 1$ while moving from Eq. (2.52) to Eq. (2.53). Since we know the exact form of $M(t)$, reproduced next, we can compute $M^L(s)$:

$$M(t) = \frac{1}{\hbar^2}\int dE'\ \rho(E')|V(E')|^2 e^{i(E_1-E')t/\hbar} \tag{2.54}$$

$$\implies M^L(s) = \frac{1}{\hbar^2}\int dE'\ \rho(E')|V(E')|^2 \frac{1}{s - i\frac{E_1-E'}{\hbar}} \tag{2.55}$$

All that remains now is to express the Fourier transform of $d_1^*(t)$ in terms of the Laplace transform of $\tilde{d}_1(t)$:

$$\int_{-\infty}^{\infty} dt\ d_1^*(t)e^{-i\omega t} = \int_{-\infty}^{0} dt\ \tilde{d_1}^*(t)e^{iE_1t/\hbar}e^{-i\omega t} + \int_{0}^{\infty} dt\ \tilde{d_1}^*(t)e^{iE_1t/\hbar}e^{-i\omega t} \tag{2.56}$$

$$= \left(\int_{0}^{\infty} dt\ \tilde{d_1}^*(-t)e^{-iE_1t/\hbar}e^{i\omega t}\right) + \lim_{\eta\to 0}\left(\tilde{d_1}^L\left(-i\left(\omega - \frac{E_1}{\hbar}\right) + \eta\right)\right)^* \tag{2.57}$$

$$= \left(\int_{0}^{\infty} dt\ \tilde{d}_1(t)e^{-iE_1t/\hbar}e^{i\omega t}\right) + \lim_{\eta\to 0}\left(\tilde{d_1}^L\left(-i\left(\omega - \frac{E_1}{\hbar}\right) + \eta\right)\right)^* \tag{2.58}$$

$$= \lim_{\eta\to 0}\tilde{d_1}^L\left(-i\left(\omega - \frac{E_1}{\hbar}\right) + \eta\right) + \lim_{\eta\to 0}\left(\tilde{d_1}^L\left(-i\left(\omega - \frac{E_1}{\hbar}\right) + \eta\right)\right)^* \tag{2.59}$$

$$= 2\mathrm{Re}\left\{\lim_{\eta\to 0}\tilde{d_1}^L\left(-i\left(\omega - \frac{E_1}{\hbar}\right) + \eta\right)\right\} \tag{2.60}$$

In moving from (2.57) to (2.58), we used $d_1(-t) = d_1^*(t)$ from Eq. (2.35). In moving from (2.56) to (2.57), the fact that the Laplace transform of $\tilde{d_1}^*(t)$ at s is equal to the complex conjugate of the Laplace transform of $\tilde{d}_1(t)$ at s^* was used. The algebra is continued as follows:

$$\int_{-\infty}^{\infty} dt\ d_1^*(t)e^{-i\omega t} = \lim_{\eta\to 0} 2\mathrm{Re}\left\{\frac{1}{-i\left(\omega - \frac{E_1}{\hbar}\right) + \eta + \frac{1}{\hbar^2}\int dE'\ \rho(E')|V(E')|^2 \frac{1}{-i\left(\omega - \frac{E_1}{\hbar}\right) + \eta - i\frac{E_1 - E'}{\hbar}}}\right\} \tag{2.61}$$

$$= \lim_{\eta\to 0} 2\mathrm{Re}\left\{\frac{1}{-i\left(\omega - \frac{E_1 + \Delta(\hbar\omega)}{\hbar}\right) + \eta + \frac{\Gamma(\hbar\omega)}{2}}\right\} \tag{2.62}$$

$$= \frac{\Gamma(\hbar\omega)}{\left(\omega - \frac{E_1 + \Delta(\hbar\omega)}{\hbar}\right)^2 + \left(\frac{\Gamma(\hbar\omega)}{2}\right)^2} \tag{2.63}$$

where we defined:

$$\Gamma(E) = \frac{2\pi}{\hbar}\rho(E)|V(E)|^2 \tag{2.64}$$

$$\Delta(E) = \int dE'\ \rho(E')|V(E')|^2 \frac{1}{E - E'} \tag{2.65}$$

The final lineshape $L(E)$ is:

$$L(E) = \frac{1}{\pi}\frac{\frac{\hbar\Gamma(E)}{2}}{\left(E - (E_1 + \Delta(E))\right)^2 + \left(\frac{\hbar\Gamma(E)}{2}\right)^2} \tag{2.66}$$

Eq. (2.66) is an *exact* expression, with no approximation whatsoever. It looks exactly like the Lorentzian lineshape we derived earlier, Eq. (2.40), except for the important difference that $\Gamma(E)$ and $\Delta(E)$ are not constant but are now dependent on E. Near the peak of the lineshape, for $E \sim E_1$, both $\Gamma(E)$ and $\Delta(E)$ are roughly constant and are equal to $\Gamma(E_1)$ and $\Delta(E_1)$ respectively. This gives us the Lorentzian shape near the center of the lineshape. However, the shape of the tails depends on the behavior of $\Gamma(E)$ and $\Delta(E)$ far away from the peak, and this is intimately connected to the structure of the wire density of states, $\rho(E)$, and the dot-wire interaction strength, $|V(E)|$. For instance, if the wire density of states $\rho(E)$ falls off exponentially on either side of E_1, then the lineshape $L(E)$ is going to be exponential on either side of E_1 too.

2.5 Deriving non-Lorentzian lineshapes — Tunneling from a dot into a wire that has a Lorentzian density of states

Essence of this section: In this section, we derive the lineshape of a quantum dot level coupled to a wire that has a Lorentzian density of states and find that it falls off as $1/E^4$ instead of the typical Lorentzian decay, $1/E^2$. We also discuss a useful equivalence between this system and another system that consists of *two* quantum dots and a wire with constant density of states.

It is difficult to draw conclusions from Eq. (2.66) because of the seeming opacity of the functions $\Gamma(E)$ and $\Delta(E)$. Therefore, we shall put the general formula aside and work out a special case now.

Let us assume the wire density of states takes on the Lorentzian form $\rho(E) = \frac{B}{E^2+A^2}$ for some energy parameters A and B. Further, let the energy of the dot level, E_1, be at the center of this Lorentzian, that is, $E_1 = 0$. Next, let the matrix element, $|V(E)| = |V|$, be a constant. Then, the corresponding memory function is going to be:

$$M(t) = \frac{|V|^2}{\hbar^2} e^{iE_1 t/\hbar} \int_{-\infty}^{\infty} dE \; \frac{B}{E^2 + A^2} e^{-iEt/\hbar} \tag{2.67}$$

$$= \frac{|V|^2}{\hbar^2} \frac{\pi B}{A} e^{-A|t|/\hbar} e^{iE_1 t/\hbar} = \frac{|V|^2}{\hbar^2} \frac{\pi B}{A} e^{-A|t|/\hbar} \tag{2.68}$$

We next use the method of Laplace transforms to solve Eq. (2.19) exactly with this particular memory function. Denoting the Laplace transform of $M(t)$ by $M^L(s)$ and that of $\tilde{d}_1(t)$ by

$\tilde{d_1}^L(s)$, we have:

$$M^L(s) = \frac{|V|^2}{\hbar^2}\frac{\pi B}{A}\frac{1}{s+\frac{A}{\hbar}} \tag{2.69}$$

$$\tilde{d_1}^L(s) = \frac{s+\frac{A}{\hbar}}{s^2 + s\frac{A}{\hbar} + \frac{|V|^2}{\hbar^2}\frac{\pi B}{A}} \tag{2.70}$$

Denoting the two complex roots of $s^2 + s\frac{A}{\hbar} + \frac{|V|^2}{\hbar^2}\frac{\pi B}{A} = 0$ by $-z_1$ and $-z_2$, and using standard inverse Laplace transform formulas, we get the following expression for $\tilde{d_1}(t)$, for $t > 0$:

$$\tilde{d_1}(t) = \frac{z_2 e^{-z_1 t} - z_1 e^{-z_2 t}}{z_2 - z_1} \tag{2.71}$$

The autocorrelation function $C(t,0)$, for all t, is given by:

$$C(t,0) = d_1^*(t) = \begin{cases} \left(z_2^* e^{-z_1^* t} - z_1^* e^{-z_2^* t}\right) / \left(z_2^* - z_1^*\right), & t \geq 0 \\ \left(z_2 e^{z_1 t} - z_1 e^{z_2 t}\right) / \left(z_2 - z_1\right), & t < 0 \end{cases} \tag{2.72}$$

The nature of the roots $-z_1$ and $-z_2$ of $s^2 + s\frac{A}{\hbar} + \frac{|V|^2}{\hbar^2}\frac{\pi B}{A} = 0$ depends on whether the discriminant $D = A^2 - 4|V|^2\frac{\pi B}{A}$ is positive, negative, or zero. If D is positive, it means the width of the Lorentzian, A, is larger than the dot-wire interaction strength, $|V|$, which implies that the system is in the 'weak coupling' regime. The equation has two real roots, and the decay $C(t,0)$ is exponential. If D is negative, on the other hand, it means the width of the Lorentzian, A, is smaller than the dot-wire interaction strength, $|V|$, and the system is said to be in the 'strong coupling' regime. The equation has two complex conjugate roots, and the decay $C(t,0)$ is a damped Rabi oscillation.

Irrespective of the regime however, eyeballing Eq. (2.72) reveals that $C(t,0)$ decays quadratically at small t, and not linearly as was the case when the wire had constant density of states (Fig. 2.4). $C(t,0)$ now has a continuous first derivative in t, while there was a discontinuity in the first derivative of $\tilde{d_1}(t)$, Eq. (2.36), for the constant wire density of states case. The close connection between the continuity of the derivatives of $\tilde{d_1}(t)$ and lineshape will be discussed in the next section.

The lineshape $L(E)$ obtained from the above autocorrelation is:

$$L(E) = \frac{1}{\pi\hbar}\mathrm{Re}\left\{\lim_{\eta\to 0}\tilde{d_1}^L(-i\omega+\eta)\right\} \tag{2.73}$$

$$= \frac{\frac{B}{\hbar^2}\frac{|V|^2}{\hbar^2}}{\left(\omega^2 - \frac{|V|^2}{\hbar^2}\frac{\pi B}{A}\right)^2 + \omega^2\frac{A^2}{\hbar^2}} \tag{2.74}$$

$$= \frac{B|V|^2}{\left(E^2 - |V|^2\frac{\pi B}{A}\right)^2 + E^2 A^2} \tag{2.75}$$

The lineshape $L(E)$ now decays as $1/E^4$ instead of as $1/E^2$ in the tail! We still have a slow, reciprocal polynomial lineshape but it now falls off faster than before. The immediate, and correct, interpretation of this result is that since a constant wire density of states produces a Lorentzian lineshape on the quantum dot, it seems reasonable that a Lorentzian wire density of states should produce a squared-Lorentzian-like lineshape. A different way of thinking about and interpreting the result would be to carefully look at the analytical behavior of $\tilde{d}_1(t)$. Specifically, the first derivative of $\tilde{d}_1(t)$ is continuous, unlike in the case of the constant wire density of states. One could speculate that there is a connection between the emergence of continuity in the first derivative of $\tilde{d}_1(t)$ and the sharpening of the aysmptotic tail of the lineshape from $1/E^2$ to $1/E^4$. This is indeed the case, as we discuss in Section 2.6.

Reinterpreting the dot and Lorentzian wire system as a 2-dot and flat wire system

Before proceeding to the discussion of the smoothness of $\tilde{d}_1(t)$, we provide a useful reinterpretation of the system we just studied, the quantum dot coupled to a wire with a Lorentzian density of states. This remolding of the problem will be of use when we attempt to build a good tunnel transistor.

We know that a quantum dot coupled to a wire with constant density of states assumes a Lorentzian lineshape. What this means is that when the joint energy eigenstates of the coupled dot-wire system are expressed as linear combinations of the old uncoupled dot and wire states, the coupled eigenstates with energy closest to the uncoupled dot energy have the highest contribution from the uncoupled dot state (measured in terms of the absolute squared coefficient in the linear combination) and the coupled eigenstates with energy farther away from the uncoupled dot energy have lower contribution from the uncoupled dot state. Moreover, the contribution falls off in a Lorentzian fashion.

Every physical system that was originally interacting only with the uncoupled quantum dot before the dot was coupled to the wire (which has constant density of states) now interacts with the joint eigenstates of the coupled dot-wire system. The strength of the physical system's interaction with a joint eigenstate depends on how much of that eigenstate was formed from the uncoupled dot state. In other words, if one were to plot the energies of the coupled eigenstates of the dot-wire system on the x-axis, and the interaction strength of the physical system with those eigenstates on the y-axis, one would get a Lorentzian function. It is as if the physical system were interacting with a different wire which had only a Lorentzian density of states! The bottomline is: A quantum dot that is coupled to a wire with constant density of states appears effectively like a wire with an effective Lorentzian density of states to any physical system that was originally interacting with the quantum dot. The wire 'dresses' the dot and makes it appear like a wire with Lorentzian density of states. This equivalence is shown in Fig. 2.6. The systems in the red dotted boxes in parts (a) and (b) are exactly equivalent to one another from the point of view of the left quantum dot.

When this interpretation is used, care should be taken regarding the normalization of the Lorentzians. The Lorentzian wire in Fig. 2.6 (a) generally has many states in it whereas the Lorentzian lineshape in Fig. 2.6 (b) has only 1 state in it. In order to move from picture (a) to picture (b), one has to set:

$$A = \frac{\hbar}{2}\frac{2\pi}{\hbar}|X''|^2\rho = \frac{\hbar}{2}\frac{1}{\tau_p} \tag{2.76}$$

$$\frac{\pi B}{A}|X|^2 = |X'|^2 \tag{2.77}$$

where ρ is the constant density of states of the wire in Fig. 2.6 (b) and τ_p is the time constant for the decay of electrons from the intermediary quantum dot to the constant density of states wire in Fig. 2.6 (b).

We end this section by quickly revisiting 'weak' and 'strong' coupling. If the wire were absent in Fig. 2.6 (b), an electron starting out on the left quantum dot would undergo Rabi oscillations between the two quantum dots. The addition of the wire induces 'damping' to the system because the electron now has a choice of escaping from the 2 quantum dot system into the wire. If the coupling $|X'|$ between the left and the right quantum dots in Fig. 2.6 (b) is weaker than the coupling $|X''|$ between the wire and the right quantum dot, the system is in 'weak coupling'. In the equivalent (a) picture, it means the width of the Lorentzian is larger than $|X|$. The electron undergoes exponential decay from the left dot into the red dotted box. If the coupling $|X'|$ between the left and the right quantum dots in Fig. 2.6 (b) is stronger than the coupling $|X''|$ between the wire and the right quantum dot, the system is in 'strong coupling'. In the equivalent (a) picture, it means the width of the Lorentzian is smaller than $|X|$. The electron undergoes damped sinusoidal decay from the left dot into the red dotted box.

2.6 Smoothness of $\tilde{d_1}(t)$ and spectral lineshape tails

Essence of this section: In this section, we show that jump discontinuities in either $\tilde{d_1}(t)$ or any of its time derivatives lead to slowly decaying, reciprocal polynomial tails in the spectral lineshape. Conversely, a time evolution $\tilde{d_1}(t)$ that is infinitely differentiable leads to spectral lineshapes that decay faster than any reciprocal polynomial. The conditions under which $\tilde{d_1}(t)$ is infinitely differentiable, that is, $\tilde{d_1}(t)$ possesses continuous derivatives of all orders, are derived.

We start out by expressing $\tilde{d_1}(t)$ in polar form:

$$\tilde{d_1}(t) = A(t)e^{i\phi(t)} \tag{2.78}$$

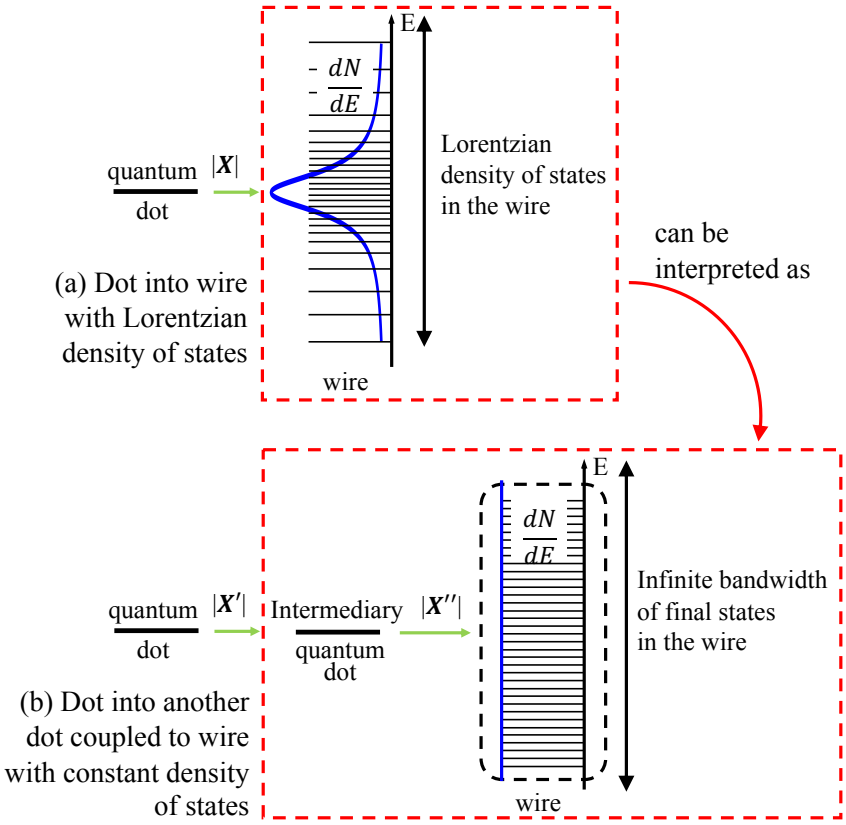

Figure 2.6: From the point of view of an electron leaking from a quantum dot into a wire, these two systems are equivalent: (a) The quantum dot is coupled to a wire with a Lorentzian density of states, and (b) The quantum dot is coupled to an intermediary quantum dot that is then coupled to a wire with a broad flat density of states through an appropriate interaction $|X''|$. The electron decay amplitude $d_1(t)$, and the resultant lineshape $L(E)$ induced on the main quantum dot, will be the same in both cases. Moreover, $L(E)$ goes as $1/E^4$, sharper than the $1/E^2$ that we get in the absence of an intermediary quantum dot.

where $A(t)$ is the magnitude function and $\phi(t)$ is the phase function. Eq. (2.35) tells us that $\tilde{d}_1(t) = \tilde{d}_1^{\,*}(-t)$. Consequently, we have:

$$A(t) = A(-t) \tag{2.79}$$

$$\phi(t) = -\phi(-t) \tag{2.80}$$

$A(t)$ is an even function of time while $\phi(t)$ is an odd function.

Next, let us make the reasonable assumption that the matrix element $V(E)$ and the wire density of states $\rho(E)$ are 'physically well-behaved' functions of E. By 'physically well-behaved', we mean that $V(E)$ and $\rho(E)$ lead to a solution $\tilde{d}_1(t)$ of Eq. (2.19) that is infinitely differentiable for all $t > 0$. Then, $A(t)$ and $\phi(t)$ are infinitely differentiable for $t > 0$ too. We

shall also assume that the right-hand derivatives of both $A(t)$ and $\phi(t)$ (of all orders) exist at $t = 0$. All these conditions were indeed true for both the flat wire density of states case and the Lorentzian wire density of states case that we studied. From the assumptions we have made, we can write a Taylor expansion for both $A(t)$ and $\phi(t)$ for $t \geq 0$:

$$A(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k}{dt^k} A(t)\bigg|_{t=0} t^k, \quad t \geq 0 \tag{2.81}$$

$$\phi(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k}{dt^k} \phi(t)\bigg|_{t=0} t^k, \quad t \geq 0 \tag{2.82}$$

We know that $A(t)$ is an even function of t, but this doesn't automatically mean that $A(t)$ has to have only even powers of t in its Taylor expansion. It could have odd powers of t too, and we already saw an example of this when we studied the case of constant wire density of states. There, the amplitude function was $A(t) = e^{-\Gamma t}$ for $t \geq 0$ but $A(t) = e^{-\Gamma|t|}$ for all t. The exponential function has both odd and even powers in its Taylor series. Therefore, the way to construct an even function from a one-sided Taylor series that has odd powers is to simply change t to $|t|$. In other words, if $A(t)$ and $\phi(t)$ are given by Eqs. (2.81) and (2.82) for $t \geq 0$, the corresponding Taylor expansions for $t < 0$ will be:

$$A(t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{d^k}{dt^k} A(t)\bigg|_{t=0} t^k, \quad t < 0 \tag{2.83}$$

$$\phi(t) = \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{k!} \frac{d^k}{dt^k} \phi(t)\bigg|_{t=0} t^k, \quad t < 0 \tag{2.84}$$

Differentiability of $A(t)$ and $\phi(t)$

Before proceeding, we make a few observations about $A(t)$ and $\phi(t)$ and their derivatives. First, both functions are continuous for all t. Next, from the Taylor series for positive time and negative time, it is clear that the k-th derivative of $A(t)$ for every even k exists at $t = 0$ because the left-hand even derivatives and the right-hand even derivatives are equal to one another at $t = 0$. For odd k, on the other hand, the left-hand derivative and the right-hand derivative differ by a - sign at $t = 0$. Therefore, if $A(t)$ has an odd power, k, in its Taylor series, the k-th derivative of $A(t)$ will be discontinuous at $t = 0$. Similar conclusions hold for $\phi(t)$, but with odd and even switched, due to it being an odd function of t. If $\phi(t)$ has an even power, k, in its Taylor series, the k-th derivative of $\phi(t)$ will be discontinuous at $t = 0$.

Discontinuities in either $\tilde{d}_1(t)$ or its time derivatives lead to Lorentzian-like tails

As alluded to earlier, it turns out that Lorentzian and other reciprocal polynomial lineshapes arise from *jump discontinuities* in t of either $\tilde{d}_1(t)$ or any of its time derivatives. A function

$x(t)$ is said to have a jump discontinuity at a point t_0 if the left-hand side limit and the right-hand side limit both exist at that point but are not equal to each other. Whenever we say 'discontinuity' in this chapter, we are referring only to jump discontinuities.

Since $\tilde{d_1}(0) = 1$ and $\tilde{d_1}(t) = \tilde{d_1}^*(-t)$, there is no discontinuity in $\tilde{d_1}(t)$ itself at any time t. However, there could be discontinuities in its time derivatives. We present the connection between discontinuities in the derivatives of $\tilde{d_1}(t)$ and the shape of $L(E)$ first through the examples we have already encountered.

Constant wire density of states

In this case, we had $A(t) = e^{-\Gamma|t|/2}$ and $\phi(t) = 0$. $A(t)$ is even in t, and $\phi(t)$ is odd, so these expressions pass the sanity checks. The lineshape $L(E)$ derived from this $\tilde{d_1}(t)$ was a Lorentzian shape, decaying as $1/E^2$ at large E. It should be noted that, while $\phi(t)$ is obviously infinitely differentiable for all t, the amplitude function, $A(t)$, is not. When considered as a function of all t, both positive and negative, $A(t)$ has discontinuities at $t = 0$ in all its odd derivatives. For instance, the first derivative of $A(t)$ is $-\Gamma/2$ immediately to the right of $t = 0$ but $\Gamma/2$ immediately to the left of $t = 0$. The first derivative of $A(t)$, therefore, has a jump discontinuity at $t = 0$. It can be checked that this is case with all the odd derivatives of $A(t)$. One can check that the discontinuities in the odd derivatives appear because of the presence of non-zero odd powers of t in the Taylor series of $A(t)$ for $t \geq 0$. If all the odd powers were absent, there would have been no discontinuity in $A(t)$ whatsoever, and $\tilde{d_1(t)}$ would have been infinitely differentiable. These odd powers are the cause of the heavy Lorentzian tail in the Fourier domain.

Lorentzian wire density of states, weak coupling

In this case, we had $A(t) = \left(z_2 e^{-z_1|t|} - z_1 e^{-z_2|t|}\right) / (z_2 - z_1)$ and $\phi(t) = 0$. $A(t)$ is even in t, and $\phi(t)$ is odd, so these expressions pass the sanity checks. The lineshape $L(E)$ derived from this $\tilde{d_1}(t)$ was like a squared Lorentzian and decayed as $1/E^4$ at large E. As before, $\phi(t)$ is infinitely differentiable for all t but the amplitude function is not. $A(t)$ has discontinuities at $t = 0$ in all its odd derivatives except the first derivative. The first derivative is zero at $t = 0$ on both sides. This can again be deduced directly by looking at the Taylor series of $A(t)$ for $t \geq 0$. The series contains powers of t for all odd numbers strictly greater than 1. These discontinuities are the cause of the heavy reciprocal polynomial tail, $1/E^4$, in the Fourier domain. However, since the linear power is absent in the Taylor series, the amplitude function in this case is slightly smoother than the $A(t)$ in the constant density of states case. This slight extra smoothness is the reason for the sharpening of the spectral lineshape tail from $1/E^2$ to $1/E^4$.

General case

Now, we generalize our observations from the previous subsection. To keep matters simple, we assume that the phase $\phi(t)$ is a simple linear function of time $\Delta t/\hbar$ that arises from

the energy shift Δ of the dot energy level caused by the wire energy levels. We had first encountered this energy shift towards the end of Section 2 of this chapter. Since $\Delta t/\hbar$ is an odd function of time t, it is a legitimate phase function, and since it is infinitely differentiable, we can forget about it altogether and focus on $A(t)$.

Why should discontinuities in the derivatives of $A(t)$, which arise from odd powers of t in the Taylor expansion, lead to heavy tails? To answer this, we look at the simplest example of a function with a jump discontinuity, the Heaviside step function, $H(t)$, and its Fourier transform, $H^F(\omega)$:

$$H(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{2.85}$$

$$H^F(\omega) = \pi\delta(\omega) - i\frac{1}{\omega} \tag{2.86}$$

$H^F(\omega)$ already has a slowly decaying, reciprocal polynomial tail, $\frac{i}{\omega}$. This is fundamentally the origin of heavy tails in functions with discontinuities in their derivatives.

Let us say the k-th derivative of $A(t)$ has a jump discontinuity at $t = 0$. From Eqs. (2.81) and (2.83), we see that k can only be odd. There cannot be a discontinuity in an even derivative. Since $A(t)$ decays exponentially at large times, we know that $A(t)$ is absolutely integrable from $-\infty$ to ∞, along with all its derivatives. Using the absolute integrability of $A(t)$ and all its derivatives, and the fact that there are no discontinuities in the even derivatives, it can be proven using tools from Fourier analysis that the Fourier transform of $A(t)$ has tails that decay as $1/\omega^{k+1}$. The proof is discussed by Saichev and Woyczynski in [72].

This is an exact generalization of what we have encountered in our calculations. The constant density of states case had a discontinuity in the first derivative, and this led to a $1/E^2$ lineshape. Similarly, the constant density of states case had a discontinuity in the third derivative, and this led to a $1/E^4$ lineshape.

Absence of discontinuities in either $\tilde{d}_1(t)$ or its derivatives leads to rapidly decaying lineshapes; Schwartz space

We have learned that the *presence* of discontinuities in $\tilde{d}_1(t)$ or its derivatives leads to slow, reciprocal polynomial lineshapes. But what about the converse? Does the *absence* of discontinuities in $\tilde{d}_1(t)$ or its derivatives leads to rapidly decaying lineshapes? Pleasantly, it turns out that the converse is indeed true for functions like $\tilde{d}_1(t)$. The set of functions that are infinitely differentiable and fall off faster than any reciprocal polynomial is called the **Schwartz space** in math. This definition rules out functions like the Lorentzian and other power law decays, but includes Gaussian and hyperbolic secant functions. Since $\tilde{d}_1(t)$ decays exponentially at large t, and is assumed to be infinitely differentiable, it belongs to the Schwartz space.

It is well known in Fourier theory that the Fourier transform of a Schwartz space function is also a Schwartz space function. This means the lineshape that results from an infinitely differentiable $\tilde{d_1}(t)$ will be **rapidly decaying in E**, i.e., it will decay faster than any reciprocal polynomial of E! In the next section, we construct a physically inspired Schwartz space candidate for the amplitude $A(t)$ of $\tilde{d_1}(t)$.

A derivation of infinitely differentiable $\tilde{d_1}(t)$ using the Markov approximation with time-dependent coefficients

We return to the derivation of Sec. 2.2 to try and derive an infinitely differentiable $\tilde{d_1}(t)$. In particular, we recall the exact Eq. (2.19):

$$\frac{d}{dt}\tilde{d_1}(t) = -\int_0^t d\tau\ M(\tau)\tilde{d_1}(t-\tau) \tag{2.87}$$

and the result of performing the Markov approximation on it, Eq. (2.25):

$$\frac{d}{dt}\tilde{d_1}(t) = -\left(\int_0^t d\tau\ M(\tau)\right)\tilde{d_1}(t) \tag{2.88}$$

Let us now try to stay at this level of approximation and avoid making the 'time-independent coefficient approximation' that we made in Sec. 2.2. Physically speaking, this stance corresponds to asserting that the memory function $M(t)$ is 'thin' enough that $\tilde{d_1}(t)$ doesn't evolve much in that duration, but not so negligible in width that it can be approximated by a Dirac delta function (Sec. 2.2 essentially reduces the memory function to a Dirac delta function). Since we know $M(t)$ exactly, we can integrate it to obtain:

$$\frac{d}{dt}\tilde{d_1}(t) = -\left[f(t) + ig(t)\right]\tilde{d_1}(t) \tag{2.89}$$

where we used $f(t)$ and $g(t)$ to abbreviate the functions that resulted from the integration of $M(t)$. $M(t)$ is reproduced here for our convenience:

$$M(t) = \frac{1}{\hbar^2}\int dE\ \rho(E)|V(E)|^2 e^{i(E_1-E)t/\hbar} = r(t) + is(t) \tag{2.90}$$

It is clear that $f(t)$ and $g(t)$ are the time integrals of $r(t)$ and $s(t)$ respectively. If we assume that $\rho(E)$ and $|V(E)|^2$ are symmetric on either side of the initial dot energy E_1, then the imaginary part $s(t)$ will be 0. Setting $E_1 = 0$, we have:

$$f(t) = \int_0^t d\tau M(\tau) = \int_0^t d\tau\left(\frac{1}{\hbar^2}\int dE\ \rho(E)|V(E)|^2 e^{-iE\tau/\hbar}\right) \tag{2.91}$$

$$= \int_0^t d\tau\left(\frac{1}{\hbar^2}\int dE\ \rho(E)|V(E)|^2\cos\left(E\tau/\hbar\right)\right) \tag{2.92}$$

$$= \frac{1}{\hbar^2}\int dE\ \rho(E)|V(E)|^2\frac{\sin\left(Et/\hbar\right)}{E/\hbar} \tag{2.93}$$

As t goes to ∞, $f(t)$ tends to $\Gamma/2$ as before. However, the interesting thing to note is that $f(t)$ is a linear combination of sine functions of t indicating that it could be an *infinitely differentiable, odd* function of time if $\rho(E)$ and $|V(E)|^2$ are chosen 'appropriately'. This infinite differentiability will eventually lead to rapidly decaying lineshapes. The 'appropriate' choice of $\rho(E)$ and $|V(E)|^2$ will be discussed momentarily. For now, integrating Eq. (2.89) yields:

$$\tilde{d_1}(t) = e^{-F(t)} \tag{2.94}$$

where $F(t) = \int_0^t d\tau f(\tau)$. If $f(t)$ is an infinitely differentiable, odd function of t, then $F(t)$ will be an **infinitely differentiable**, **even** function of t. Then, concomitantly, $e^{-F(t)}$, will be infinitely differentiable and even in t too. Further, at large t, $e^{-F(t)}$ decays asymptotically as $\sim e^{-\Gamma|t|/2}$. Therefore, $e^{-F(t)}$ will belong to the Schwartz space, which means the lineshape that results from Fourier transforming it will be rapidly decaying in E! In other words, when we work in this approximation, we can confidently rule out Lorentzians and all other reciprocal polynomial decays in E if $f(t)$ is infinitely differentiable and odd.

Since $f(t)$ is the time integral of $M(t)$, it will be infinitely differentiable and odd if and only if $M(t)$ is infinitely differentiable and even. We also know that $M(t)$ is absolutely integrable, meaning that the integral of its absolute value from $t = 0$ to $t = \infty$ is finite and equal to $\Gamma/2$. Finally, we note that $M(t)$ is the Fourier transform of $\rho(E)|V(E)|^2$. Fourier theory tells us that, if $p(\omega)$ is a function and $p^F(t)$ its Fourier transform, then the absolute integrability of $p^F(t)$ implies that $p(\omega)$ is continuous, and the infinite differentiability of $p^F(t)$ implies that $p(\omega)$ decays more rapidly than any reciprocal polynomial of ω. Therefore, it seems plausible that, if the wire density of states $\rho(E)$ and the dot-wire interaction $|V(E)|$ are chosen such that $\rho(E)|V(E)|^2$ is continuous in E and narrow, then the resulting lineshape $L(E)$ on the quantum dot will be rapidly decaying too. For $\rho(E)|V(E)|^2$ to be narrow, the density of states and the interaction matrix element cannot both simultaneously be broad functions of E. In our previous examples, both $\rho(E)$ and $|V(E)|$ were broad, leading to heavy reciprocal polynomial tails.

The arguments of this section seem to indicate that it is essential to choose either a narrow-band wire at the outset, or set the dot-wire interaction such that only a narrow band of wire states interact with the dot, in order to obtain rapidly decaying lineshapes on the dot. This is what we will explore in more detail in the upcoming sections.

2.7 Proposed tunnel transistor design and its analysis

Essence of this section: In this section, we combine insights from the previous sections to argue that inserting a narrow-band wire between the main quantum dots and their corresponding wire contacts in the tunnel transistor will result in lineshapes that decay more rapidly than any reciprocal power law. An original candidate function $A_c(t)$ for the true, complicated, $d_1(t)$ is proposed, its physical meaning is discussed, and the lineshape it produces is shown to decay exponentially in E.

In this section, we combine our insights from Secs. 2.2, 2.5, and 2.6. The summary of what we have learned so far is:

1. Coupling a quantum dot to a wire with an infinitely broad, constant density of states leads to purely exponential decay of electron amplitude from the dot into the wire.
2. The initial time decay of $\tilde{d}_1(t)$ is linear in t, and the odd powers of t in the pure exponential decay lead to lineshapes with heavy Lorentzian tails that go as $1/E^2$.
3. Coupling a quantum dot to a wire with a narrower, Lorentzian density of states leads to exponential decay of electron amplitude from the dot into the wire, but with no linear power in t. The time decay is quadratic at small t. The odd powers of t from 3 onwards still exist.
4. The improved smoothness of $\tilde{d}_1(t)$ for this case leads to sharper tails, $1/E^4$.
5. Coupling a quantum dot to Lorentzian wire is equivalent to coupling the quantum dot to another dot which is itself coupled to a wire with a broad, constant density of states.

These points suggest a procedure to construct a sharper lineshape. Coupling a quantum dot to a wire with constant density of states leads to $\tilde{d}_1(t)$ with all odd powers of t. Coupling a quantum dot to a wire with constant density of states, **but through an intermediary quantum dot**, gets rid of the linear power of t in $\tilde{d}_1(t)$, though all the other odd powers of t remain. Why don't we keep adding intermediary quantum dots one after the other between the main quantum dot and the wideband wire? Each intermediary quantum dot should get rid of one additional odd power of t.

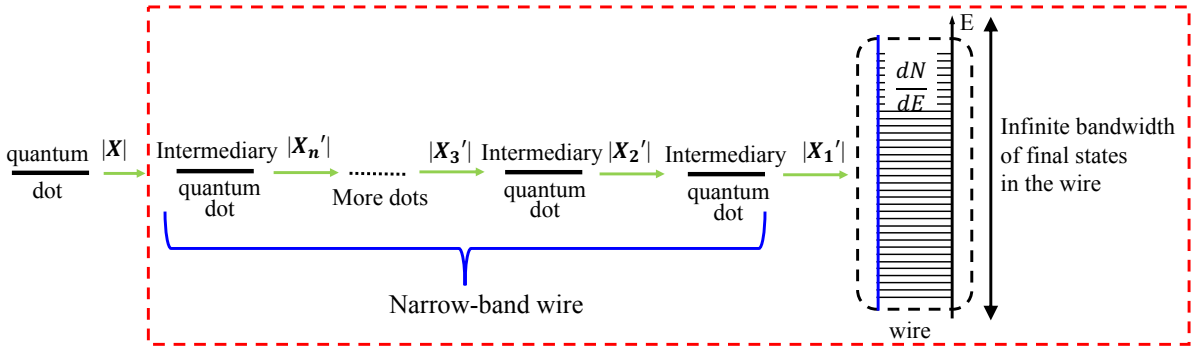

Figure 2.7: Since the addition of one intermediary quantum dot between the main quantum dot and the wire sharpened the tails of the lineshape from $1/E^2$ to $1/E^4$, one can sharpen it further by adding a sequence of intermediary dots. The series of intermediary dots will effectively form a narrow-band wire between the main quantum dot and the broadband contact wire.

If we place many intermediary dots in succession, they effectly form a wire in their own right. This is illustrated in Fig. 2.7. The interaction $|X_1'|$ cannot be stronger than the inter-atom interactions within the infinite-bandwidth wire, else the first intermediary quantum dot will simply become a part of the wide-band wire. Further, we had reasoned that each successive intermediary quantum dot would sharpen the lineshape of the main quantum dot. For this argument to go through, each intermediary quantum dot has to be weakly coupled to its preceding intermediary quantum dot. Since the entire chain of intermediary quantum dots is only weakly held together, and the tight-binding method tells us that the bandwidth depends on the inter-atom coupling, we deduce that the bandwidth of this chain will be small. That is, the chain of intermediary quantum dots acts like a narrow-band wire.

Now that the wire has been constructed, we simply have to compute $\tilde{d}_1(t)$ for the decay of the electron from the leftmost quantum dot into the narrow-band wire. In order to avoid the detailed calculations that would be necessitated by an exact analysis, we obtain direct physical intuition about the decay by first proposing an ansatz for the decay, $A_c(t)$, and then giving it physical meaning by matching its parameters with the cases we have studied so far. This is the subject of the next two subsections.

A candidate $A_c(t)$ for the decay from the dot to the narrow-band wire

When a quantum dot is coupled to such a narrow-band wire, an electron starting out on it will undergo quantum amplitude decay $\tilde{d}_1(t)$ that is infinitely differentiable. We assume that the phase part of $\tilde{d}_1(t)$ is a simple energy shift $e^{-i\Delta t/\hbar}$. The amplitude part $A(t)$ has to decay exponentially at large t. We propose the following function, $A_c(t)$, as a candidate ansatz for $A(t)$ that captures the essential physics of the system:

$$A_c(t) = \exp\left(\frac{b}{2} - \frac{1}{2}\sqrt{b^2 + \frac{t^2}{\tau^2}}\right) \tag{2.95}$$

where b is a dimensionless parameter and τ is a time constant. We shall assign physical meaning to these parameters and interpret them in the next subsection. It is easily verified that this function decays exponentially at large t, as a correct amplitude function should. The decay coefficient is $1/2\tau$. Further, $A_c(t)$ decays quadratically at small t, and not linearly, with the coefficient of t^2 in the Taylor series being $-1/4b\tau^2$.

The most attractive feature of this candidate function is that it is infinitely differentiable and decays exponentially asymptotically, which means it belongs to the Schwartz space. Its Fourier transform, $A_c^F(\omega)$, belongs to the Schwartz space too. The exact formula for $A_c^F(\omega)$, taken from the comprehensive table of integrals by Bateman [73], is:

$$A_c^F(\omega) = \frac{2b\tau e^{b/2}}{\sqrt{1+4\omega^2\tau^2}} K_1\left(\frac{b}{2}\sqrt{1+4\omega^2\tau^2}\right) \tag{2.96}$$

where $K_1(\cdot)$ is the $\nu = 1$ member of a family of functions called the modified Bessel functions of the second kind, $\{K_\nu(\cdot)\}_{\nu\in\mathbb{C}}$. For large ω, the famous mathematical functions handbook by Abramovich and Stegun [74] tells us that $K_1(\cdot)$ decays exponentially for large values of the argument. The asymptotic behavior of $A_c^F(\omega)$, for $\omega^2 \gg (4-b^2)/(4b^2\tau^2)$, is:

$$A_c^F(\omega) \approx \frac{2b\tau e^{b/2}}{\sqrt{1+4\omega^2\tau^2}}\sqrt{\frac{\pi}{b\sqrt{1+4\omega^2\tau^2}}}e^{-\frac{b}{2}\sqrt{1+4\omega^2\tau^2}} \tag{2.97}$$

$$= \mathcal{O}\left(e^{-b\tau|\omega|}\right) \tag{2.98}$$

where $\mathcal{O}$ is the standard asymptotic 'big-oh' notation used in computer science. In the opposite limit, for small values of the argument, the modified Bessel function $K_1(p)$ goes as $1/p$. This means that, for small ω and b, $A_c^F(\omega)$ has the following asymptotic expression:

$$A_c^F(\omega) \approx \frac{4\tau}{1+4\omega^2\tau^2} \tag{2.99}$$

We recover the familiar Lorentzian form of lineshape close to $\omega = 0$. We chose the candidate function $A_c(t)$ for precisely this reason. The function $A_c^F(\omega)$ looks like a traditional Lorentzian near its center $\omega = 0$ but has exponential decays in its far tails.

The final lineshape $L_c(E)$ that is produced by the Fourier transform $A_c^F(\omega)$ is:

$$L_c(E) = \frac{1}{2\pi\hbar}A_c^F(E/\hbar) = \frac{1}{2\pi}\frac{2be^{b/2}}{\sqrt{\frac{\hbar^2}{\tau^2}+4E^2}}K_1\left(\frac{b\tau}{2\hbar}\sqrt{\frac{\hbar^2}{\tau^2}+4E^2}\right) \tag{2.100}$$

For large $|E|$, the asymptotic form is:

$$L_c(E) = \mathcal{O}\left(e^{-\frac{|E|}{\hbar/b\tau}}\right) \tag{2.101}$$

$L_c(E)$ has exponential decays in E at both ends of the spectrum—let us call it a double-sided exponential—with a characteristic decay energy of $\hbar/(b\tau)$.

Physical interpretation of the candidate $A_c(t)$

As already indicated earlier, $A_c(t)$ is simply our mathematical candidate function that captures the essential nature of the true amplitude decay from the dot to the narrow-band wire. In order to infuse this abstract function with physical meaning, we need to define the parameters b and τ in physical terms. Our approach will be to apply $A_c(t)$ to the simpler case of amplitude decay into a Lorentzian wire, which was solved in Sec 2.5 (Eq. (2.72), denoted by $A_l(t)$ in this subsection). Since $A_c(t)$ has two parameters, we need to match two temporal properties of $A_c(t)$ with the corresponding properties of $A_l(t)$ to express both b and τ in terms of the physical parameters of the Lorentzian wire case. Once we have these expressions, we shall take a leap of faith and generalize their meaning to the case of any narrow-band wire.

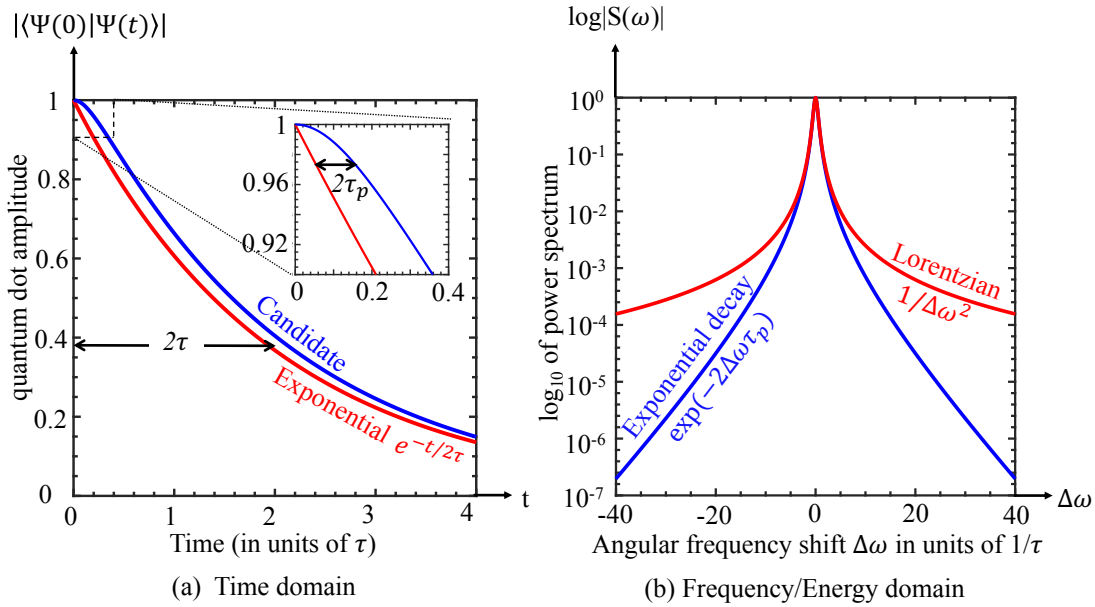

Figure 2.8: (a) The candidate function $A_c(t)$ is exponentially decaying at large t in accordance with Fermi's Golden Rule, but is infinitely differentiable. This brings about an initial quadratic decay near $t = 0$ as shown in the inset. (b) This seemingly small change in time-domain has startling consequences for the frequency-domain. The tails of the candidate $A_c^F(\omega)$ are now steep exponentials compared to the heavy Lorentzian tails in the normal Fermi Golden Rule case.

Our first step is to set the long-time asymptotic decay coefficients of $A_c(t)$ and $A_l(t)$ equal to one another. Next, we set the small-time quadratic decay coefficients equal to one another. In terms of the parameters in the picture, Fig. 2.6 (b), we get:

$$\frac{1}{2\tau} = 2\frac{|X'|^2}{\hbar^2}\tau_p \tag{2.102}$$

$$-\frac{1}{4b\tau^2} = -\frac{|X'|^2}{2\hbar^2} \tag{2.103}$$

where $|X'|$ is the interaction strength between the main quantum dot and the intermediary quantum dot and τ_p is the decay time constant from the intermediary quantum dot to the wide-band wire. Equivalently, $\hbar/\tau_p$ is the Full-Width-at-Half-Maximum of the Lorentzian lineshape induced on the intermediary dot by the wide-band wire. These equations together imply that $A_c^F(\omega)$ decays with the constant $b\tau = 2\tau_p$. Since we have physically motivated choices of b and τ now, we compare $A_c(t)$ and pure exponential decay on the same plot in Fig. 2.8 (a), and their respective Fourier transforms in Fig. 2.8 (b).

We generalize this exact definition of τ and b to the dot-and-narrow-band-wire system in Fig 2.7. The role of $|X'|$ in Eqs. (2.102) ad (2.103) will be played by the interaction $|X|$

between the main quantum dot and the narrow-band wire in Fig 2.7. The role of $\hbar/\tau_p$ will be played by the effective Full-Width-at-Half-Maximum of the narrow-band wire.

It might be surprising at first sight that the decay constant of $A_c^F(\omega)$, given by $b\tau = 2\tau_p$, depends only on the bandwidth of the narrow-band wire and has no dependence at all on the coupling strength between the main quantum dot and the narrow-band wire. In the Lorentzian lineshapes we studied earlier in this chapter, the lineshape became broader as the coupling to the wire got stronger. It only seems reasonable that a similar effect should be observable here. This confusion is resolved when one notes that, though the dot-narrow-band-wire coupling does not affect the behavior of $A_c^F(\omega)$ as $\omega \to \infty$, it does affect its form close to the origin. Eq. (2.99) tells us that the lineshape behaves like a Lorentzian as $\omega \to 0$. As the dot-narrow-band-wire coupling strengthens, this Lorentzian broadens, so we recover the familiar relation between 'linewidth' and the dot-wire coupling strength. It is only the far tail of $A_c^F(\omega)$ that is independent of the dot-narrow-band-wire coupling.

2.8 Final structure and $I_d - V_g$ characteristics

Essence of this section: In this section, we present the $I_d - V_g$ characteristics of the modified tunnel transistor device with a narrow-band wire inserted between the main quantum dots and their corresponding wire contacts. The characteristics are computed using the candidate function $A_c(t)$ for the true $d_1(t)$. A physical platform for the fabrication of such a device is also briefly discussed.

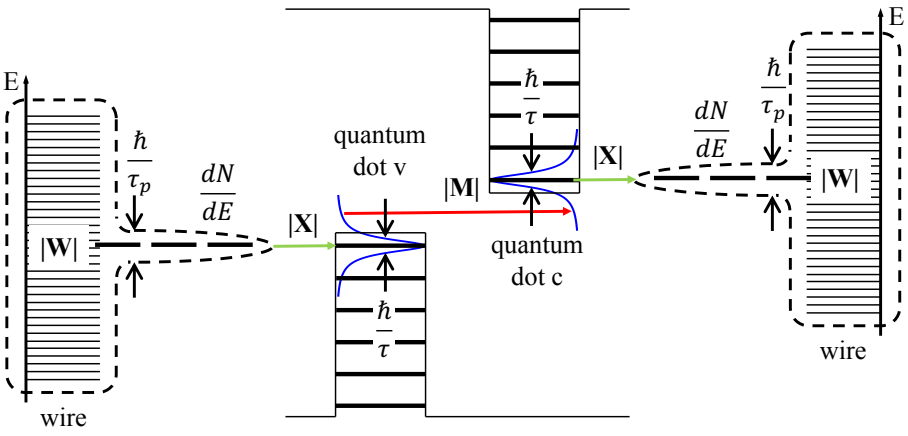

Figure 2.9: Energy band depiction of our proposed narrow-band tunnel transistor. The source and drain quantum dots are coupled to each other via interaction $|M|$. Each of them is coupled to its associated narrow-band wire via interaction $|X|$. The narrow-band wires couple to the broadband contact wires via the interaction $|W|$. The width of the density of states of the narrow-band wires is $\hbar/\tau_p$.

Our final design for the tunnel transistor simply consists of the inclusion of a narrow-band wire between the main quantum dots and their respective wide-band wire contacts. The inclusion of the narrow-band wire sharpens the lineshapes of the main quantum dots and leads to sharper I-V characteristics. An energy band depiction of this structure is given in Fig. 2.9. The source and drain quantum dots are coupled to each other via interaction $|M|$, each of them is coupled to its associated narrow-band wire via interaction $|X|$, and the narrow-band wires couple to the broadband contact wires via the interaction $|W|$. The width of the density of states of the narrow-band wires is $\hbar/\tau_p$. In [66] and [75], it was shown that the matrix element $|M|$ coupling the two quantum dots cannot be made arbitrarily large. $|M|$ must be smaller than the dot linewidth, $|M| < \hbar/4\tau$. If $|M|$ were any larger, strong coupling between the two dots would occur. Under strong coupling, the electron would oscillate back and forth between the quantum dots leading to the disappearance of the energy-filtering-based switching mechanism and the reemergence of the thermal 60mV/decade slope.

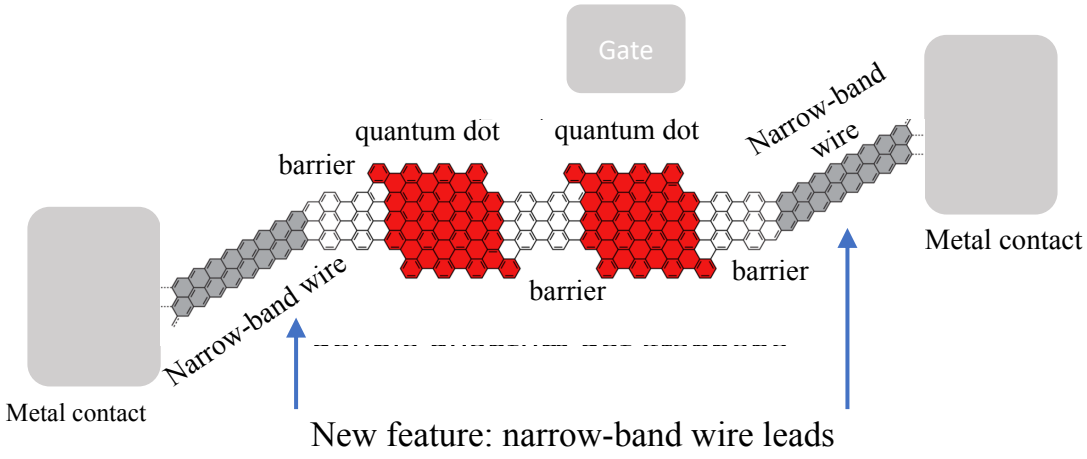

Figure 2.10: Physical construction of our proposed narrow-band tunnel transistor. All the components of the transistor, the quantum dots, the tunnel barriers, and the narrow-band wires, can be made using graphene nanoribbons of the appropriate size and width. The drain-side quantum dot is controlled via a gate terminal.

In order to avoid leakage current arising from tunneling through defect states, we would like our devices to be defect-free. Defects are one of the banes of present-day bulk tunnel transistors. To overcome these challenges, scientists have investigated the performance of tunnel transistors constructed in the bottom-up fashion from the molecular level. Based on the extensive work done on graphene and its derivatives, we propose that the tunnel transistor should be built from graphene nanoribbons. Recent results from Mutlu et al. indicate that we now have the ability to synthesize precisely controlled, defect-free graphene nanoribbons [76]. Equally important for our proposal is the question of whether it is possible to synthesize wires with narrow bands. It is known that zigzag graphene nanoribbons can

be designed to possess nearly flat bands in their electronic structure, giving rise to peaked, narrow density of states [77, 78].

Further, the width of the ribbons and the edge type can also be modulated to produce insulating and conducting states of graphene. In summary, the entirety of the tunnel transistor, including the main quantum dots, the insulating tunnel barriers, and the narrow-band wires, can be fabricated in the graphene nanoribbon platform. A schematic of such a transistor is depicted in Fig. 2.10.

$I_d - V_g$ characteristics

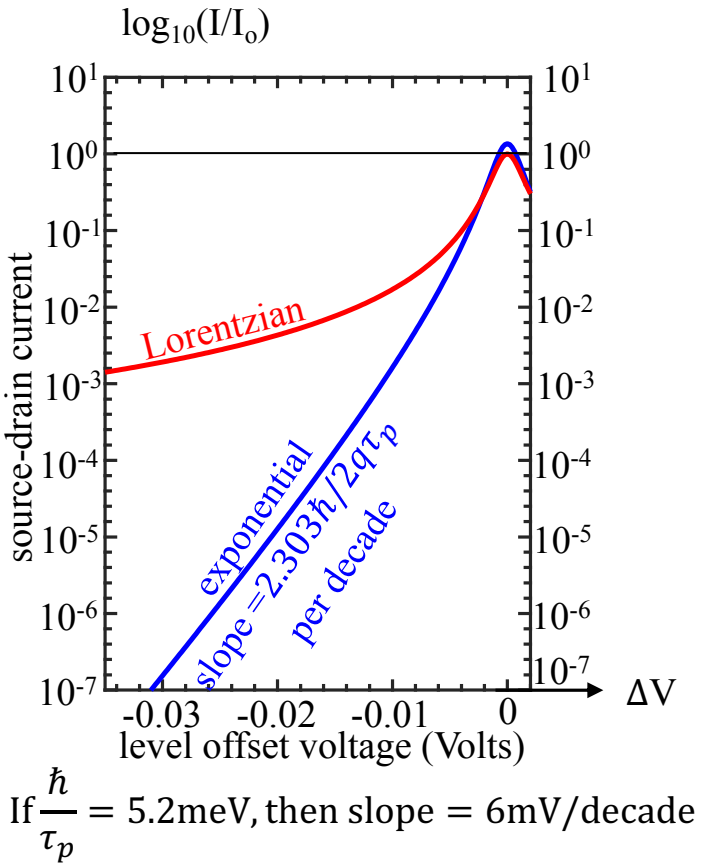

Figure 2.11: Based on the overlap between the source and drain dot spectral lineshapes, the subthreshold current-voltage characteristic is shown. The exponential spectral tails produce a leakage that can be characterized in terms of steepness or voltage swing: "millivolts per decade." For this numerical example, $\hbar/2q\tau_p = 2.6mV$, $\tau = 500fs = 4\tau_p$, and the steepness $2.303\hbar/2q\tau_p = 6mV/$decade. The Lorentzian spectrum has unacceptable leakage. The current I_0 used to normalize the two plots is the full on-state current for aligned Lorentzian quantum dots.

Lastly, we present the $I_d - V_g$ characteristics for the narrow-band wire tunnel transistor. The lineshape induced on the quantum dots by the narrow-band wire is modeled by the candidate lineshape function, $(1/(2\pi\hbar))\, A_c^F(E/\hbar)$, that we discussed in detail in the previous section. We recall from Sec. 2.3 that the $I_d - V_g$ characteristics are obtained from the source-dot-drain-dot interaction strength $|M(E)|$ and the lineshape functions of the two quantum dots as follows:

$$I_{\text{full}}(V_g) = -2\frac{2\pi e}{\hbar}\int dE\ L_1(E)L_2(E+eV_g)|M(E_1)|^2\,(f_1(E)-f_2(E)) \tag{2.104}$$

Temperature T appears in the above expression essentially only through the Fermi factors $f_1(E)$ and $f_2(E)$. For now, we will just set $T = 0$ and assume that the source and drain Fermi levels are so widely separated that the source dot is fully occupied and the drain dot is fully unoccupied. The case of non-zero T is discussed at the end of this subsection. We also set the inter-dot coupling to a constant, $|M(E)| = |M|$. Then, the expression for the current reduces to a simple convolution of the source-side and drain-side lineshapes. It is easily shown that the convolution of two exponential functions in ω, each with a decay constant of $1/2\tau_p$, is itself exponential with a decay constant $1/2\tau_p$. Shifting from the ω domain (frequency) to energy and then to the V domain (voltage) through the introduction of $\hbar$ and q, the decay constant becomes $\hbar/2q\tau_p$. This means that the subthreshold slope, in units of V/decade, is given by:

$$V_{SS} = 2.303\frac{\hbar}{2q\tau_p} \tag{2.105}$$

This expression says that the subthreshold slope of the device depends only on the width of the density of states of the narrow-band wire, $\frac{\hbar}{\tau_p}$. The narrower the bandwidth of the narrow-band wire, the smaller the subthreshold slope will be according to our model. In Fig. 2.11, we plot the IV curve that results from taking a narrow-band wire that has a bandwidth of 5.2 meV. The resulting sub-threshold slope is shown to be a remarkable 6 mV/decade. We look forward to future tunnel transistor work using narrow-band wires by the experimental community in order to confirm our predictions.

Eq. (2.104) can be used to compute the peak current, that is the current at 0 level offset $V_g = 0$, under the same assumptions of $T = 0$, source fully occupied and drain empty, and $|M|$ being constant. We have:

$$I_{\text{peak}} \approx -2\frac{2\pi e}{\hbar}|M|^2\frac{\tau}{\pi\hbar} \tag{2.106}$$

where τ is the exponential decay lifetime of the dots.

Effect of temperature on the sub-threshold slope

When the temperature T is non-zero, the function $f_1(E) - f_2(E)$ has exponential tails in E at both extremes, $E \to \infty$ and $E \to -\infty$, with slope $V_T = 60$mV/decade.

Since $L_1(E)$ is a double-sided exponential with slope given by Eq. (2.105), it is clear that $N(E) = L_1(E)\,(f_1(E) - f_2(E))$ is a double-sided exponential function with slope

$\frac{1}{V_N} = \frac{1}{V_{SS}} + \frac{1}{V_T}$. Again assuming that $|M(E)| = |M|$ for all E, the right-hand side of Eq. (2.104) reduces to a convolution of $N(E)$, a double-sided exponential with slope V_N, and $L_2(E)$, a double-sided exponential with slope V_{SS}. It is easy to verify that the convolution of two double-sided exponential functions is another double-sided exponential function with an asymptotic slope that is equal to the *larger* of the two slopes. That is, when two double-sided exponentials are convolved, the slower one dominates and the passes on its slope to the resultant convolution. Since V_N is smaller than V_{SS}, the resultant $I_d - V_g$ curve has an asymptotic slope given by V_{SS}V/decade even at finite temperatures. The asymptotic subthreshold slope of the device in the far tails is essentially independent of temperature!

2.9 Conclusion

The conclusions of this chapter are:

1. Quantum dots that are coupled to their surroundings hybridize with them and experience an effective broadening of their original sharp energy levels. The shape of the broadening is related to the time-domain behavior of the wavefunction of an electron placed on the quantum dot.

2. The $I_d - V_g$ performance of a tunnel transistor composed of quantum dots coupled to wires depends on the exact shape of the energy broadening or lineshape of the quantum dots.

3. Coupling source and drain quantum dots directly to broadband contact wires results in Lorentzian broadening of the quantum dots. Lorentzian lineshapes lead to unacceptable leakage in the off state.

4. Inserting an intermediary quantum dot between the main quantum dots and their contacts smoothens the time-domain electron wavefunction and leads to a sharper-than-Lorentzian lineshape.

5. Smoothness of the time-domain decay of the wavefunction is intimately connected to the tails of the energy-domain lineshape.

6. Inserting a series of intermediary quantum dots, effectively narrow-band wires, between the main quantum dots and their contacts leads to lineshapes that decay faster than any reciprocal power law in E.

7. The subthreshold slope of the $I_d - V_g$ characteristic depends on the bandwidth of the narrow-band wire. The narrower the bandwidth, the lower the subthreshold slope.

8. Graphene nanoribbons seem to be a promising platform for the fabrication of high-purity, narrow-band wire tunnel transistors.

Chapter 3

Optical Spectroscopy and the Quest for the Urbach Tail

We turn next to the problem of optical absorption in crystals. The absorption coefficient, $\alpha(\omega)$, quantifies the extent to which the light incident on a system is absorbed by it at different frequencies. Theoretical calculations often yield Lorentzian absorption line shapes of the form $1/\left((\omega-\omega_0)^2+(\Gamma/2)^2\right)$, where ω_0 is the peak resonance frequency and Γ is the associated 'linewidth'. However, in the solid state, the optical spectrum of electronic transitions usually decays exponentially below the bandgap via the so-called Urbach tail, $e^{(\omega-\omega_0)/\Gamma_U}$ (valid for $\omega < \omega_0$), where $\hbar\omega_0$ is the bandgap of the material and Γ_U is the characteristic Urbach decay frequency [10].

A sector where this exponential falloff has far-reaching consequences is optical communications. The Urbach tail permits glass to be very transparent in the infrared—this essential property of glass makes possible the long-distance optical communication that underlies the internet today. If the electronic transitions in glass had a Lorentzian line shape instead, the infrared optical absorption would never quite turn off due to the slow decay $\sim 1/\Delta\omega^2$ and the attenuation length at the communication frequencies would be ~ 1 meter. Lorentzian line shape in glass would be a catastrophe for modern civilization, which relies upon optical fibers for worldwide connectivity. Urbach tails are found not only in amorphous glass but also in almost every crystalline semiconductor, even when the sample is defect-free.

Despite its universality and critical importance to telecommunications, there is no physics consensus for a simple, common origin of Urbach tail spectra, though many different physical models have been proposed in the past [79–85]. Mahr [86] and Keil [84] compute absorption lineshapes by performing a weighted-averaging of the absorption shape for each possible crystal lattice distortion due to thermal vibrations but do not obtain satisfactory Urbach tails. Halperin and Lax [87] only derive impurity-induced band-tails. Dow and Redfield [85] deduce the absorption by computing the effect of internal electric microfields on excited electron-hole pairs. Sumi and Toyozawa [88] compute the absorption using the Laplace transform approach that we ourselves adopt in this chapter, but perform different approximations and do not succeed in obtaining a clear exponential shape with the right temperature de-

pendence. John et al. [81] propose a complex theory for disordered crystals by modeling the phonon dynamics via random Gaussian fields.

In this chapter, we contribute to the question of the Urbach tail by applying the technique of 'Markov approximation with time-dependent coefficients' that we encountered in the previous chapter to various model optical systems. In the process, we succeed in obtaining rapidly decaying absorption spectral lineshapes that belong to the Schwartz class (introduced in Chapter 2) though we are still unable to pinpoint the exact exponential decay coefficient and the exact temperature dependence. Future novel work on this problem using our approach and approximation might eventually lead to a transparent universal derivation of the Urbach tail.

3.1 Absorption coefficient $\alpha(\omega)$ in terms of the susceptibility $\chi(\omega)$

Absorption $\alpha(\omega)$

The absorption coefficient of a material is a measure of the fraction of the incident light that remains unabsorbed in a material after propagating a unit distance in it. More precisely, if I_0 is the incident light intensity, $I(x)$ is the intensity at distance x in the medium, and the absorption coefficient of the medium is α, we have:

$$I(x) = I_0 e^{-\alpha x} \tag{3.1}$$

α has units of reciprocal distance and usually has a strong dependence on the frequency of the incident light. The function $\alpha(\omega)$, called the absorption spectrum or the absorption lineshape, is very important in optical engineering.

Susceptibility $\chi(\omega)$

The susceptibility is the linear response coefficient that captures the effect of incident electric fields E on the electric dipole moment density or polarization of the material P. Again, this quantity depends on the frequency of the electric field, ω. The linear response relation is given by:

$$P(\omega) = \epsilon_0 \chi(\omega) E(\omega) \tag{3.2}$$

The response coefficient is, in general, a complex number:

$$\chi(\omega) = \chi'(\omega) + i\chi''(\omega) \tag{3.3}$$

Eq. (3.2) can be rewritten in the time domain as a convolution:

$$P(t) = \epsilon_0 \int_{-\infty}^{\infty} d\tau \; \chi(t-\tau) E(\tau) \tag{3.4}$$

Since real-life processes have to be causal, $\chi(t)$ is non-zero only for $t \geq 0$. Further, the material is assumed to be passive, that is, it is not driven or pumped in any form. Therefore, there is no optical gain. Then, one can use the ideas of complex contour integration and derive the famous Kramers-Kronig relations between the real and imaginary parts of $\chi(\omega)$:

$$\chi'(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} d\omega' \ \frac{\chi''(\omega')}{\omega' - \omega} \tag{3.5}$$

$$\chi''(\omega) = -\frac{1}{\pi} \int_{-\infty}^{\infty} d\omega' \ \frac{\chi'(\omega')}{\omega' - \omega} \tag{3.6}$$

The meaning of these relations is that the real and imaginary parts of $\chi(\omega)$ cannot behave arbitrarily. One does not have independent control over the two of them. Fixing one of them as a function of ω decides the other immediately through these relations. A derivation of the Kramers-Kronig relations is given in the optics textbook by Amnon Yariv [89].

The relation between $\alpha(\omega)$ and $\chi(\omega)$

It turns out that the absorption coefficient $\alpha(\omega)$ is linearly related to the *imaginary part* of the susceptibility $\chi''(\omega)$. The simple derivation of this fact is as follows.

A wave $e^{i(\omega t - kx)}$ propagating in a lossy medium can be described using a complex wave number $k = k'(\omega) - ik''(\omega)$. The imaginary part means that the field amplitude falls off exponentially with decay coefficient $k''(\omega)$. The intensity, therefore, has a decay coefficient $\alpha(\omega) = 2k''(\omega)$, since it is the square of the field amplitudes. The wave number can be expressed in terms of a complex refractive index as:

$$k'(\omega) - ik''(\omega) = \frac{\omega}{c} \left(n'(\omega) + in''(\omega) \right) \tag{3.7}$$

where $n'(\omega)$ and $n''(\omega)$ are the real and imaginary refractive indices of the $e^{i\omega t}$ wave. Moreover, it is known that $n^2 = \epsilon_r = 1 + \chi$. Comparing real and imaginary parts on both sides, we have:

$$n'^2 - n''^2 = 1 + \chi' \tag{3.8}$$

$$2n'n'' = \chi'' \tag{3.9}$$

Therefore, the decay coefficient α is:

$$\alpha(\omega) = 2k''(\omega) = -2\frac{\omega}{c} n''(\omega) = -\frac{\omega}{c} \frac{\chi''(\omega)}{n'(\omega)} \tag{3.10}$$

We assume that the real refractive index $n'(\omega)$ is roughly constant in the frequency range of interest. Then, $\alpha(\omega)$ becomes directly proportional to $\chi''(\omega)$. The rest of the chapter will focus on computing $\chi''(\omega)$.

3.2 Lorentz-Lorenz model

Essence of this section: In this section, we present the simple classical Lorentz-Lorenz model for the refractive index and derive from it the standard Lorentzian lineshape for $\chi''(\omega)$. We need to go beyond this simple model to get exponential lineshapes like the Urbach tail.

The Lorentz-Lorenz model is a simple classical model for the susceptibility $\chi(\omega)$ that predicts reciprocal polynomial lineshapes that look Lorentzian near the peak. We quickly discuss this model to see what we are up against in our quest for the Urbach tail.

We focus on one electron in the medium, described by its position $x(t)$. The net force experienced by the electron includes the electric field force, the Drude damping due to the resistance of the medium, and a possible spring force that keeps the electron bound to its nucleus (this force is very important in insulators, where free electrons are a rarity). Newton's force equation tells us:

$$m\ddot{x} = qE - m\omega_0^2 x - m\frac{1}{\tau}\dot{x} \tag{3.11}$$

where $q = -e$ is the electron charge, m is the electron mass, ω_0 is the spring's natural frequency, and $1/\tau$ is the resistive damping rate. If the electric field is sinusoidal and has frequency ω, we can write $x(t) = Xe^{i\omega t}$ for some amplitude X. Plugging this in, we get:

$$X = \frac{qE/m}{(\omega_0^2 - \omega^2) + i\frac{\omega}{\tau}} \tag{3.12}$$

The dipole moment of the separated electron-nucleus pair (let the nucleus have only one proton) is then just $p = qX$. If there are N such pairs in a unit volume in the material, the susceptibility $\chi(\omega)$ will be given by:

$$\chi(\omega) = \frac{Nq^2/m}{(\omega_0^2 - \omega^2) + i\frac{\omega}{\tau}} \tag{3.13}$$

The imaginary part of this expression is:

$$\chi''(\omega) = -\frac{\frac{Nq^2}{m}\frac{\omega}{\tau}}{(\omega_0^2 - \omega^2)^2 + \left(\frac{\omega}{\tau}\right)^2} \tag{3.14}$$

We see that the absorption lineshape decays as $1/\omega^3$, a reciprocal polynomial, at large frequencies. Further, there is no hint of temperature dependence at all in the above expression. It should be recalled that the Urbach tail is an exponential decay with an energy decay coefficient that is approximately $kT/2$. In order to derive **exponential** Urbach tails that correctly capture the **temperature dependence**, it seems essential that more nuanced physics be included in the computation.

3.3 Statistical models of lineshape based on dipole moment fluctuations; Brownian motion; Kubo's lineshape

Essence of this section: In this section, we introduce the fact that the absorption spectrum of a system is intimately connected to the random equilibrium fluctuations of its dipole moment. A specific statistical model of lineshape originally proposed by Kubo is presented, and the reasons why it does not lead to Urbach tails are discussed.

One of the fundamental discoveries in the history of non-equilibrium statistical mechanics was the 'fluctuation-dissipation theorem'. This result gives the relationship between the linear response of a driven, out-of-equilibrium system, and the fluctuations of the system variables *at equilibrium*. We will derive this theorem in the **next section**, Sec. 3.4. For now, however, we will simply take on faith the fact that the susceptibility $\chi(\omega)$ of a system is directly related to equilibrium time-domain fluctuations of its dipole moment p.

$\chi(\omega)$ is in fact proportional to the Fourier transform of the autocorrelation function $\langle p(t)p(0)\rangle$ of the equilibrium time-domain dipole moment fluctuations of the system. The autocorrelation function of a system variable is a 2-time statistical average, meaning, it measures how correlated that system variable is at two different times. The dipole moment autocorrelation $R(t_1, t_2)$ is the equilibrium average of the product of the system's dipole moments at time t_1 and t_2: $R(t_1, t_2) = \langle p(t_1)p(t_2)\rangle$. The angular brackets denote ensemble averaging of this quantity over several copies of the system.

Such an autocorrelation function is computable if one has a model for the dynamics of the variable under question—we are interested in the dipole moment here. The dipole moment $p(t)$ of a system evolves both due to external driving and due to noisy perturbations from the system's surroundings. This type of dynamics is similar to the Brownian motion of a particle suspended in a liquid. Stochastic differential equations are a natural way of describing Brownian motion, and can similarly be used to form a statistical model for the evolution of the dipole moment. We take a brief look at Brownian motion next.

Brownian motion

The dipole moment $p(t)$ of a system evolves both due to external driving and due to noisy perturbations from the system's surroundings. This type of dynamics is similar to the Brownian motion of a particle suspended in a liquid. In Brownian motion, the particle's motion involves directed drift that is caused by structured external driving forces, and random diffusion caused by the particle's interaction with the random thermal motion of the liquid's molecules. Such dynamics is well described using *stochastic* differential equations that represent the effect of the reservoir's random thermal motion on the particle using a random process in time $n(t)$. The following is an example of a stochastic differential equation that

can be used to describe the particle's velocity:

$$\frac{dv(t)}{dt} = -\alpha v(t) + n(t) \tag{3.15}$$

α is a viscosity coefficient or resistance that the particle experiences, and the $n(t)$ is the random process that dictates how the environment affects the particle's acceleration.

In equilibrium, the particle has zero average velocity, but the mean-squared velocity is non-zero and is proportional to kT, where k is the Boltzmann constant and T is the ambient temperature. The particle's mean-squared velocity can also be computed from Eq. (3.15) in terms of the viscosity α, and the strength A of the noise process $n(t)$. Equating the two expressions for the mean-squared velocity yields a relationship between the dissipation coefficient α, the noise strength A, and the temperature T. This is the famous Einstein relation that connects drift (mobility) and diffusion (diffusivity). The simplest treatment of Brownian motion assumes that $n(t)$ is Gaussian white noise. It is then fairly straightforward to explicitly derive an Einstein relation.

Kubo's model

To derive the dipole moment autocorrelation $R(t_1, t_2) = \langle p(t_1)p(t_2)\rangle$, we need a model for the dynamics of the dipole moment. A popular model for the dipole moment is the one presented by Kubo in [90]. This model succeeds in capturing the effect of motional narrowing that is observed in Nuclear Magnetic Resonance and other spectroscopic settings. Kubo's model is described in Appendix 2. From the Appendix, the formula for the dipole autocorrelation is:

$$\langle p(0)p(t)\rangle = p(0)^2 \exp\left(-\Delta^2\tau_c^2\left(e^{-t/\tau_c} + \frac{t}{\tau_c} - 1\right)\right) \tag{3.16}$$

where Δ is the width of the background Gaussian frequency distribution and $1/\tau_c$ is the rate at which the dipole moment 'jumps' around. The meaning of these terms and parameters is explained in Appendix 2.

Discussion

It can be verified that the above formula, when Taylor expanded about $t = 0$, has cubic terms in $|t|$. This indicates a discontinuity in the third derivative of the autocorrelation function at $t = 0$. We saw in Chapter 2 that discontinuities in any of the derivatives of a function generates reciprocal polynomial tails in its Fourier transform. Therefore, this model cannot produce asymptotic exponential tails in $\chi(\omega)$ for any choice of the two parameters involved. One fundamental cause of this discontinuity is the pure linear decay of the angular frequency ω with rate $1/\tau_c$ in the equations in the Appendix. The real-world autocorrelation has a quadratic behavior at short times that is not captured by such a linear decay. In the remaining sections, we will move away from models that involve ad hoc assumptions like pure

linear decay terms and instead turn to more rigorous quantum mechanical treatments. To capture the 'smooth' behavior of the dipole autocorrelation function in time more precisely, we will rigorously solve for the quantum mechanical evolution of the dipole moment in several model systems in an attempt to tease out an Urbach tail.

3.4 Linear response and autocorrelations

Essence of this section: In this section, we derive the important Green-Kubo formula that connects (1) the linear response of a system to a driving force, and (2) the statistical fluctutations of the system in equilibrium, before the driving force was applied. We will specifically relate the susceptibility $\chi(\omega)$ to the Fourier transform of the dipole autocorrelation function.

In this section, we derive the important Green-Kubo formula that connects the (1) linear response of a system to a driving force and the (2) statistical fluctutations of the system in equilibrium, before the driving force was applied. This relation is also called the 'fluctuation-dissipation theorem' and dates back in its original form to Onsager [4].

Although the derivation is completely general, the notation will be oriented towards optical response, that is, the connection between the (1) linear response of the dipole moment to driving electric fields, which is simply the susceptibility, and (2) the equilibrium fluctuations of the dipole moment. We shall also switch to quantum mechanics for the derivation because the rest of the chapter involves quantum mechanical calculations.

Setup

Let the Hamiltonian of the equilibrium system be H_0. The driving force is the classical electric field, $E(t) = \left(E^* e^{-i\omega t} + E e^{i\omega t}\right)/2$. The system, which is charge-neutral, interacts with the electric field through its dipole moment, p, resulting in an interaction energy of $V = -p \cdot E(t)$. We shall assume that the electric field is linearly polarized along a particular direction, so only the dipole moment operator along that direction plays a role. The full Hamiltonian is:

$$H = H_0 + V = H_0 - pE(t) \tag{3.17}$$

Schrödinger equation; Interaction picture

The Schrödinger equation in density matrix notation is:

$$\frac{d\rho(t)}{dt} = -\frac{i}{\hbar}[H, \rho(t)] = -\frac{i}{\hbar}[H_0 + V, \rho(t)] \tag{3.18}$$

The version of quantum mechanics with constant operators and time-dependent states is called the Schrödinger picture. In order to get rid of the equilibrium Hamiltonian H_0 and to

understand the novel dynamics caused by the driving $-pE(t)$, we perform the 'Interaction picture' change of variables. This change of variables effectively amounts to removing the fast oscillations caused by H_0 from the density matrix and looking only at the slowly varying envelope of the wavefunction caused by $-pE(t)$. As an aside, we mention that if one were to shift the entire time-dependence, both fast and slow, to the operators and make the state constant (not just slowly-varying), then we are said to be using the 'Heisenberg picture'. All of these are simply changes of variables, there is no difference in their physical content.

We denote the operators after the variable change with tildes. The interaction picture operator $\tilde{A}(t)$ that corresponds to an arbitrary Schrödinger operator A is:

$$\tilde{A}(t) = e^{i\frac{H_0}{\hbar}t} A e^{-i\frac{H_0}{\hbar}t} \tag{3.19}$$

The density matrix in the interaction picture is:

$$\tilde{\rho}(t) = e^{i\frac{H_0}{\hbar}t} \rho(t) e^{-i\frac{H_0}{\hbar}t} \tag{3.20}$$

With this change of variables, the Schrödinger equation becomes:

$$\frac{d\tilde{\rho}(t)}{dt} = -\frac{i}{\hbar}[\widetilde{V}(t), \tilde{\rho}(t)] \tag{3.21}$$

Linear response

Since we are interested only in linear response, we will look at the first order change in $\tilde{\rho}(t)$ caused by the driving. Expressing $\tilde{\rho}(t)$ as $\tilde{\rho}(t) = \tilde{\rho}^{(0)}(t) + \tilde{\rho}^{(1)}(t) + \dots$ where $\tilde{\rho}^{(i)}(t)$ depends on the i-th power of the perturbation $\tilde{V}$, we have, for the zeroth and first orders:

$$\frac{d\tilde{\rho}^{(0)}}{dt} = 0 \tag{3.22}$$

$$\frac{d\tilde{\rho}^{(1)}}{dt} = -\frac{i}{\hbar}[\widetilde{V}(t), \tilde{\rho}^{(0)}(t)] \tag{3.23}$$

Let us say the electric field was turned on at time $t = 0$. The density matrix at $t = 0$ should then be the equilibrium density matrix of the system, $\rho(0) = e^{-H_0/kT}/Z$, where Z is the normalization constant, and T is the temperature. From this, it is also clear that $\tilde{\rho}^{(0)}(0) = e^{-H_0/kT}/Z$ and $\tilde{\rho}^{(1)}(0) = 0$.

Solving (3.22) gives us $\tilde{\rho}^{(0)}(t) = \tilde{\rho}^{(0)}(0) = \rho(0)$. Plugging this into (3.23) and integrating from 0 to t, we get:

$$\tilde{\rho}^{(1)}(t) = -\frac{i}{\hbar}\int_0^t [\widetilde{V}(\tau), \rho(0)] d\tau \tag{3.24}$$

The expected dipole moment of the system is obtained by taking the trace of the product of the dipole operator and the density matrix:

$$\langle p(t) \rangle = \mathrm{Tr}\left\{\tilde{p}(t)\tilde{\rho}(t)\right\} \tag{3.25}$$

Then, to first order in the perturbation, the expected dipole moment at time t is:

$$\begin{aligned}
\langle p(t)\rangle &= \mathrm{Tr}\left\{\tilde{p}(t)\tilde{\rho}^{(1)}(t)\right\} && (3.26)\\
&= -\frac{i}{\hbar}\int_0^t \mathrm{Tr}\left\{\tilde{p}(t)\left[\widetilde{V}(\tau),\rho(0)\right]\right\}d\tau && (3.27)\\
&= -\frac{i}{\hbar}\int_0^t \mathrm{Tr}\left\{\left[\tilde{p}(t),\widetilde{V}(\tau)\right]\rho(0)\right\}d\tau && (3.28)\\
&= \frac{i}{\hbar}\int_0^t \mathrm{Tr}\left\{\left[\tilde{p}(t),\widetilde{p}(\tau)\frac{1}{2}\left(Ee^{i\omega\tau}+E^*e^{-i\omega\tau}\right)\right]\rho(0)\right\}d\tau && (3.29)\\
&= \frac{i}{\hbar}\int_0^t \mathrm{Tr}\left\{\left[\tilde{p}(t-\tau),\widetilde{p}(0)\right]\rho(0)\right\}\frac{1}{2}\left(Ee^{i\omega\tau}+E^*e^{-i\omega\tau}\right)d\tau && (3.30)\\
&= \frac{i}{\hbar}\frac{E}{2}e^{i\omega t}\int_0^t d\tau\ \mathrm{Tr}\left\{\left[\tilde{p}(\tau),\widetilde{p}(0)\right]\rho(0)\right\}e^{-i\omega\tau}\\
&\quad + \frac{i}{\hbar}\frac{E^*}{2}e^{-i\omega t}\int_0^t d\tau\ \mathrm{Tr}\left\{\left[\tilde{p}(\tau),\widetilde{p}(0)\right]\rho(0)\right\}e^{i\omega\tau} && (3.31)
\end{aligned}$$

In the above sequence, we assumed that $\rho(0)$ is diagonal in the unperturbed eigenstate basis which is indeed the case. The steady state dipole moment, after all the transients die down, is obtained by setting $t\to\infty$ in the above expression. The susceptibility $\chi(\omega)$ is then given by:

$$\chi(\omega) = \frac{i}{\hbar}\int_0^\infty \mathrm{Tr}\left\{\left[\tilde{p}(\tau),\widetilde{p}(0)\right]\rho(0)\right\}e^{-i\omega\tau}d\tau \tag{3.32}$$

Since we have $\tilde{p}(\tau) = e^{i\frac{H_0}{\hbar}t}Ae^{-i\frac{H_0}{\hbar}t}$, $\tilde{p}(\tau)$ is in fact the Heisenberg dipole operator at time τ in the absence of the light probe. The trace on the right-hand side above is then the average of the dipole-dipole commutator at two different times in thermal equilibrium. By thermal equilibrium, we mean that the system is not driven in any manner. The input driving electric field $E(t)$ is absent, as is the interaction term V. Since we are computing a 2-time average of the same physical quantity, we will call it an autocorrelation. Therefore, $\chi(\omega)$, which is the linear response of the material polarization to input light, is related to the equilibrium dipole-dipole autocorrelation function. We have successfully related linear response to equilibrium fluctuations.

$\chi''(\omega)$ is the Fourier transform of the dipole autocorrelation

The imaginary part of $\chi(\omega)$, which is what we are interested in, is given by:

$$\chi''(\omega) = \frac{1}{2\hbar}\int_{-\infty}^\infty \mathrm{Tr}\left\{\left[\tilde{p}(\tau),\widetilde{p}(0)\right]\rho(0)\right\}e^{-i\omega\tau}d\tau \tag{3.33}$$

The right-hand side is in fact a Fourier transform. In summary, the procedure to get the optical absorption lineshape $\alpha(\omega)$ is to compute the dipole-dipole commutator autocorrelation, take its Fourier transform, and scale it by the relevant prefactors.

Building up to absorption in a semiconductor

Trying to derive Urbach tails by directly applying this machinery to semiconductors and insulators is not directly tractable because of the complexity of those systems. It is well-understood that phonons, or mechanical vibrations, are the cause of the Urbach tail in crystals. Our approach toward this problem will be to study a few simple electronic systems that are coupled to phonon reservoirs and then attempt to draw more general conclusions from those results. In the rest of the chapter, we shall study optical absorption of (1) an electron in a harmonic potential well that is coupled to a bath of phonons, and (2) an electron in a 2-level atom that is coupled to a bath of phonons. Though these examples do not directly lead to Urbach tails, they are very instructive.

Since Eq. (3.33) involves the time-evolution of the dipole operator, we will use the Heisenberg equation of motion for operators. The Heisenberg picture involves capturing the dynamics in the operators, instead of in the state as the Schrödinger picture does. The Heisenberg equation of motion for an operator $A(t)$, which is completely equivalent to the Schrödinger equation for states, is given by:

$$\frac{d}{dt}A(t) = \frac{i}{\hbar}\left[H, A(t)\right] \tag{3.34}$$

where H, the Hamiltonian of the system, is assumed to be time-independent.

3.5 Warm-up: Absorption in a harmonic potential well

Essence of this section: In this section, we apply the Green-Kubo formula to the simple case of optical absorption by an electron in a harmonic potential well as a warm-up problem. This same procedure will be consistently applied to the other systems in the rest of the chapter.

Our very first application of the dipole autocorrelation formula to find $\chi''(\omega)$ will be for the optical absorption of an electron in a harmonic potential well. The harmonic oscillator has discrete equally spaced levels, so the absorption should simply be a set of delta functions at various harmonics of the harmonic oscillator natural frequency ω_0. When we consider optical absorption due to the dipole interaction alone, we in fact get absorption only at the natural frequency ω_0.

There is only one electron in the well

Hamiltonian

The Hamiltonian of the harmonic oscillator, H_0, is:

$$H_o = \hbar\omega_0 \left(b^\dagger b + \frac{1}{2} \right) \tag{3.35}$$

where $b/b^\dagger$ are the destruction/construction operators of the electron's oscillator, and ω_0 is the natural frequency of the electron's oscillator.

Heisenberg equations

The Heisenberg equation for the harmonic oscillator destruction operator is:

$$i\hbar \frac{d}{dt} b(t) = \hbar\omega_0 b(t) \tag{3.36}$$

which yields:

$$b(t) = be^{-i\omega_0 t} \tag{3.37}$$

Dipole commutator autocorrelation

Now, the dipole operator of the electron is $p = -eX$, where X is the electron's position operator. In a harmonic oscillator, the position can be expressed in terms of the creation and destruction operators giving us:

$$p(t) = -\frac{e\sqrt{\hbar}}{\sqrt{2m_e\omega_0}} \left(b(t) + b^\dagger(t) \right) \tag{3.38}$$

We are interested in the average value of the commutator of $p(t)$ and $p(0)$:

$$\langle [p(t), p(0)] \rangle = \frac{e^2\hbar}{2m_e\omega_0} \left(\langle [b(t), b] \rangle + \left\langle \left[b(t), b^\dagger \right] \right\rangle + \left\langle \left[b^\dagger(t), b \right] \right\rangle + \left\langle \left[b^\dagger(t), b^\dagger \right] \right\rangle \right) \tag{3.39}$$

The first and last terms on the right-hand side of Eq. (3.39) are clearly 0 due to the form of $b(t)$ in Eq. (3.37). From the relation $\left[b, b^\dagger \right] = 1$, we see that the second term on the right-hand side of Eq. (3.39) is $e^{-i\omega_0 t}$ while the third term on the right-hand side of Eq. (3.39) is $-e^{i\omega_0 t}$. Plugging these in and using Eq. (3.33), the imaginary part of $\chi(\omega)$ is:

$$\chi''(\omega) = \frac{e^2\pi}{2m_e\omega_0} \left(\delta\left(\omega + \omega_0\right) - \delta\left(\omega - \omega_0\right) \right) \tag{3.40}$$

Discussion

$\chi''(\omega)$ is composed of delta functions at ω_0 and $-\omega_0$, indicating that there is sharp absorption of light at $\hbar\omega = \hbar\omega_0$, the energy difference between adjacent states of the harmonic oscillator. The oscillator levels are not broadened because there is no reservoir or environment that is interacting with and disturbing the oscillator. One curious feature of Eq. (3.40) is that it contains no temperature dependence at all—one would have expected the electron thermal occupations of the oscillator levels to have played a role in the absorption, but that temperature dependence is evidently missing here. The reason is that our present analysis only considered a single electron in the harmonic well. We rectify this in the next subsection by filling the harmonic well with several non-interacting electrons.

There are several non-interacting electrons in the well

The natural way to express the Hamiltonian of a multi-electron system is via the use of Fermionic construction and destruction operators—this approach is also commonly known as the *second quantization formalism*. Just like their bosonic counterparts, the fermionic construction operators add electrons to modes/energy levels while the destruction operators remove electrons from energy levels. One key difference is that the fermionic operators satisfy *anti-commutation* relations as opposed to the commutation relations satisfied by the bosonic operators. This difference is fundamentally due to the fact that bosons can be exchanged between states with no change to the wavefunction while fermion exchange leads to a phase factor of -1 in the wavefunction. If $c_k/c_k^\dagger$ are the destruction/construction operators for an electronic state $|k\rangle$, they satisfy the anticommutation relation $\left\{c_k, c_k^\dagger\right\} = c_k c_k^\dagger + c_k^\dagger c_k = 1$.

Hamiltonian

The Hamiltonian of the harmonic oscillator in second quantization is:

$$H_o = \sum_{n=0}^{\infty} \hbar\omega_0 \left(n + \frac{1}{2}\right) c_n^\dagger c_n \tag{3.41}$$

where $c_n/c_n^\dagger$ are the destruction/construction operators for the n-th energy level in the harmonic oscillator, and ω_0 is the natural frequency of the electron's oscillator.

Heisenberg equations

The Heisenberg equations are:

$$i\hbar\frac{d}{dt}c_k(t) = \hbar\omega_0 \left(k + \frac{1}{2}\right) c_k(t) \tag{3.42}$$

which yields:

$$c_k(t) = c_k e^{-i\omega_0\left(k+\frac{1}{2}\right)t} \tag{3.43}$$

Dipole commutator autocorrelation

Recalling that the dipole operator for a single electron is given by:

$$p(t) = -\frac{e\sqrt{\hbar}}{\sqrt{2m_e\omega_0}}\left(b(t) + b^\dagger(t)\right) \tag{3.44}$$

one can write the sum of the dipole operators of all the electrons in the system as:

$$p_{\text{all}}(t) = -\frac{e\sqrt{\hbar}}{\sqrt{2m_e\omega_0}}\left(\sum_{k=0}^{\infty}\sqrt{k+1}\left(c_k^\dagger(t)c_{k+1}(t) + c_{k+1}^\dagger(t)c_k(t)\right)\right) \tag{3.45}$$

The average value of the commutator of $p(t)$ and $p(0)$ is then:

$$\begin{aligned}\langle[p(t),p(0)]\rangle =& \frac{e^2\hbar}{2m_e\omega_0}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sqrt{k+1}\sqrt{l+1}\left(\left\langle\left[c_k^\dagger(t)c_{k+1}(t), c_l^\dagger c_{l+1}\right]\right\rangle\right. \\ &+ \left\langle\left[c_k^\dagger(t)c_{k+1}(t), c_{l+1}^\dagger c_l\right]\right\rangle + \left\langle\left[c_{k+1}^\dagger(t)c_k(t), c_l^\dagger c_{l+1}\right]\right\rangle \\ &\left.+ \left\langle\left[c_{k+1}^\dagger(t)c_k(t), c_{l+1}^\dagger c_l\right]\right\rangle\right)\end{aligned} \tag{3.46}$$

Patient algebra on the commutators reveals that the first and last terms reduce to terms of the form $\left\langle c_k c_{k+2}^\dagger\right\rangle$, $\left\langle c_{k+1}^\dagger c_{k-1}\right\rangle$, and the like. All these averages, which correspond to 'coherences' between different harmonic oscillator energy levels, can be set to 0 in thermal equilibrium. On the other hand, the second term, $\left\langle\left[c_k^\dagger(t)c_{k+1}(t), c_{l+1}^\dagger c_l\right]\right\rangle$, reduces to something like $\left\langle c_k^\dagger c_k - c_{k+1}^\dagger c_{k+1}\right\rangle$, which can be expressed in terms of the thermal occupations of the energy levels. The third term also reduces to a similar form. The final expression for the dipole commutator autocorrelation is:

$$\langle[p(t),p(0)]\rangle = \frac{e^2\hbar}{2m_e\omega_0}\left(e^{-i\omega_0 t} - e^{i\omega_0 t}\right)\sum_{k=0}^{\infty}(k+1)\left(\langle n(k)\rangle - \langle n(k+1)\rangle\right) \tag{3.47}$$

$$= \frac{e^2\hbar}{2m_e\omega_0}\left(e^{-i\omega_0 t} - e^{i\omega_0 t}\right)\sum_{k=0}^{\infty}(k+1)\left(\frac{1}{1+e^{\frac{\hbar\omega_0(k+0.5)-\mu}{kT}}} - \frac{1}{1+e^{\frac{\hbar\omega_0(k+1.5)-\mu}{kT}}}\right) \tag{3.48}$$

where $\langle n(k)\rangle$ is the thermal occupation of the k-th energy level. We will assume that the spacing between the energy levels, $\hbar\omega_0$, is much larger than kT, as is the case for absorption in semiconductors. Then, only the two levels, k_0 and k_0+1, that are on either side of the Fermi level μ contribute significantly to the summation above.

$$\langle[p(t),p(0)]\rangle \approx \frac{e^2\hbar}{2m_e\omega_0}\left(e^{-i\omega_0 t} - e^{i\omega_0 t}\right)(k_0+1)\left(\langle n(k_0)\rangle - \langle n(k_0+1)\rangle\right) \tag{3.49}$$

The imaginary part of $\chi(\omega)$ is then:

$$\chi''(\omega) \approx \frac{e^2\pi}{2m_e\omega_0}\left(\delta(\omega+\omega_0) - \delta(\omega-\omega_0)\right)(k_0+1)\left(\langle n(k_0)\rangle - \langle n(k_0+1)\rangle\right) \tag{3.50}$$

Discussion

$\chi''(\omega)$ again has the same form as before—it has delta functions at ω_0 and $-\omega_0$. However, the correct dependence of the absorption on the difference of the thermal populations of the lower and upper levels is now recovered.

A key lesson to be learned here is that, though second quantization calculations recover the exact dependence of the absorption lineshape on the electronic thermal occupations, the single electron calculation was accurate enough to capture the essential shape of the absorption. The electron thermal occupation dependence is a very intuitive correction that can be tacked onto the lineshape later on, in the spirit of the $I_d - V_g$ curve calculation for finite temperature in Chapter 2. Further, the electron thermal occupation difference term seemingly has no effect on the shape of the absorption, and, above it all, is nearly 1 for semiconductors at room temperature. It cannot possibly play a role in the creation of the Urbach tail. For this reason, all our calculations in the succeeding sections will be in the single-electron picture.

3.6 Absorption by an electron in a harmonic potential well coupled to a phonon reservoir

Essence of this section: In this section, we apply the Green-Kubo formula to the slightly more involved case of optical absorption by an electron in a harmonic potential well that is coupled to a reservoir of phonons. The harmonic potential well is coupled to a large crystal which is at thermal equilibrium. The phonons in the crystal interact with the electron and broaden its absorption spectrum.

Our next $\chi''(\omega)$ calculation will be for the optical absorption of an electron in a harmonic potential well that is coupled to a phonon bath as depicted in Fig 3.1. The electron is attached to its home stationary atom via a spring. This harmonic oscillator has the well-known equally-spaced energy spectrum. Its absorption spectrum consists of a sharp Dirac delta function at the natural frequency of the oscillator ω_0. When the electron is weakly connected to a lattice of springs, the vibrations of the lattice hybridize with, and consequently broaden, the originally sharp energy levels of the harmonic electron. This also broadens the absorption spectrum, which will be computed in this section.

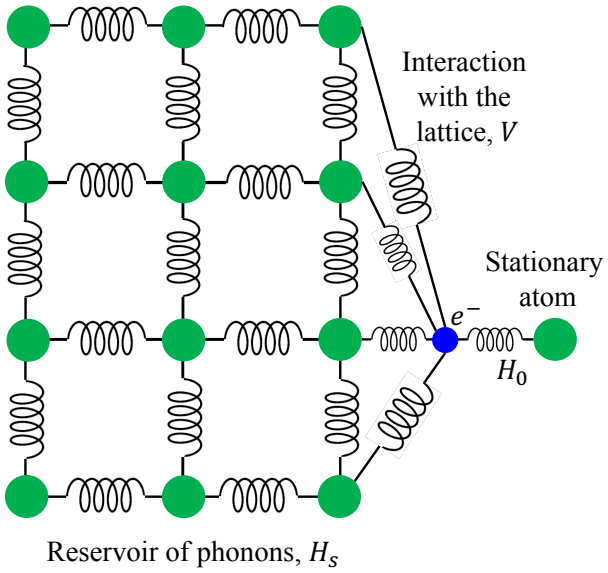

Figure 3.1: The electron is in a harmonic potential well and is also coupled to a large crystal which is at thermal equilibrium. The phonons in the crystal interact with the electron and broaden its absorption spectrum. The phonons are described by the Hamiltonian H_s, the harmonic well by H_0, and their interaction by V.

Hamiltonian

The Hamiltonian of the damped harmonic oscillator is sum of three terms, the harmonic oscillator's own Hamiltonian, H_0, the reservoir term, H_s, and the coupling term, V:

$$H_o = \hbar\omega_0 \left(b^\dagger b + \frac{1}{2} \right) \tag{3.51}$$

$$H_s = \sum_i \hbar\omega_i \left(a_i^\dagger a_i + \frac{1}{2} \right) \tag{3.52}$$

$$V = \sum_i \left(g_i b^\dagger a_i + g_i^* b a_i^\dagger \right) \tag{3.53}$$

where $b/b^\dagger$ are the destruction/construction operators of the electron's oscillator, $a_i/a_i^\dagger$ are the destruction/construction operators of the i-th phonon mode, ω_0 is the natural frequency of the electron's oscillator, ω_i are the phonon frequencies, and the g_i are the coupling strengths between the electron and the i-th phonon. In the coupling term V, we are considering only those processes that result in the creation of an electron excitation when a phonon is destroyed, and the destruction of an electron excitation when a phonon is created. When the offset between ω_i and ω_0 is large, this is only an approximation, and other processes such as the simultaneous creation of electron and phonon excitations should also be considered. This is done in Appendix A.

Heisenberg equations

The Heisenberg equations for the operators are:

$$i\hbar\frac{d}{dt}b(t) = \hbar\omega_0 b(t) + \sum_i g_i a_i(t) \tag{3.54}$$

$$i\hbar\frac{d}{dt}a_i(t) = \hbar\omega_i a_i(t) + g_i^* b(t) \tag{3.55}$$

Introducing 'slowly-varying' operators:

$$b(t) = \hat{b}(t)e^{-i\omega_0 t} \tag{3.56}$$

$$a_i(t) = \hat{a_i}(t)e^{-i\omega_i t} \tag{3.57}$$

The Heisenberg equations become:

$$i\hbar\frac{d}{dt}\hat{b}(t) = \sum_i g_i\hat{a_i}(t)e^{i(\omega_0-\omega_i)t} \tag{3.58}$$

$$i\hbar\frac{d}{dt}\hat{a_i}(t) = g_i^*\hat{b}(t)e^{i(\omega_i-\omega_0)t} \tag{3.59}$$

We integrate the $\hat{a_i}$ equations from 0 to t and substitute the result into the $\hat{b}$ equation to get:

$$\frac{d}{dt}\hat{b}(t) = -\int_0^t d\tau\kappa(\tau)\hat{b}(t-\tau) + \hat{F}(t) \tag{3.60}$$

where

$$\kappa(\tau) = \frac{1}{\hbar^2}\sum_i |g_i|^2 e^{i(\omega_0-\omega_i)\tau} \tag{3.61}$$

$$\hat{F}(t) = -\frac{i}{\hbar}\sum_i g_i\hat{a_i}(0)e^{i(\omega_0-\omega_i)t} \tag{3.62}$$

The Heisenberg equation for $\hat{b}(t)$, Eq. (3.60), is a linear equation with a memory kernel $\kappa(\tau)$ and an added 'noise' input term, $\hat{F}(t)$. Laplace transforms are a standard technique to solve such equations. In our notation, the Laplace transform of a function $x(t)$ will be $x^L(s)$. Following the standard procedure, the solution to (3.60) turns out to be:

$$\hat{b}(t) = \int_0^t h(t-s)\hat{F}(s)ds + h(t)\hat{b}(0) \tag{3.63}$$

$$\hat{b}^\dagger(t) = \int_0^t h^*(t-s)\hat{F}^\dagger(s)ds + h^*(t)\hat{b}^\dagger(0) \tag{3.64}$$

where $h(t)$ is the function whose Laplace transform is given by:

$$h^L(s) = \frac{1}{s + \kappa^L(s)} \tag{3.65}$$

The Laplace transform of $h^*(t)$ is denoted by $(h^*)^L(s)$ and given by:

$$(h^*)^L(s) = \frac{1}{s + (\kappa^L(s^*))^*} \tag{3.66}$$

$\kappa^L(s)$ is the Laplace transform of $\kappa(\tau)$, and is given by:

$$\kappa^L(s) = \frac{1}{\hbar^2} \sum_i \frac{|g_i|^2}{s - i(\omega_0 - \omega_i)} \tag{3.67}$$

Dipole commutator autocorrelation

Now, the dipole operator of the electron is $p = -eX$, where X is the electron's position operator. In a harmonic oscillator, the position can be expressed in terms of the creation and destruction operators giving us:

$$p(t) = -\frac{e\sqrt{\hbar}}{\sqrt{2m_e\omega_0}} \left(b(t) + b^\dagger(t)\right) \tag{3.68}$$

We are interested in the average value of the commutator of $p(t)$ and $p(0)$:

$$\langle [p(t), p(0)] \rangle = \frac{e^2\hbar}{2m_e\omega_0} \left(\langle [b(t), b] \rangle + \left\langle \left[b(t), b^\dagger\right] \right\rangle + \left\langle \left[b^\dagger(t), b\right] \right\rangle + \left\langle \left[b^\dagger(t), b^\dagger\right] \right\rangle \right) \tag{3.69}$$

Now, the first commutator in Eq. (3.69) is:

$$\langle [b(t), b] \rangle = e^{-i\omega_0 t} \left(\int_0^t h(t-s) \left\langle \left[\hat{F}(s), b\right] \right\rangle ds + h(t) \langle [b, b] \rangle \right) = 0 \tag{3.70}$$

The first term is zero because a_i commutes with b, and the second is zero because $[b, b] = 0$. A similar cancellation happens in the fourth commutator and renders it equal to zero. The second commutator in (3.69) is:

$$\left\langle \left[b(t), b^\dagger\right] \right\rangle = e^{-i\omega_0 t} \left(\int_0^t h(t-s) \left\langle \left[\hat{F}(s), b^\dagger\right] \right\rangle ds + h(t) \left\langle \left[b, b^\dagger\right] \right\rangle \right) \tag{3.71}$$

$$= h(t) e^{-i\omega_0 t} \tag{3.72}$$

A similar calculation for the third commutator of Eq. (3.69) yields:

$$\left\langle \left[b^\dagger(t), b\right] \right\rangle = -h^*(t) e^{i\omega_0 t} \tag{3.73}$$

Therefore, the full dipole commutator autocorrelation is:

$$\langle [p(t), p(0)] \rangle = \frac{e^2\hbar}{2m_e\omega_0} \left(h(t) e^{-i\omega_0 t} - h^*(t) e^{i\omega_0 t} \right) \tag{3.74}$$

Susceptibility $\chi(\omega)$

Applying Eq. (3.32) for the susceptibility, we have:

$$\chi(\omega) = \frac{i}{\hbar}\frac{e^2\hbar}{2m_{e0}}\int_0^\infty \left(h(\tau)e^{-i\omega_0\tau} - h^*(\tau)e^{i\omega_0\tau}\right)e^{-i\omega\tau}d\tau \tag{3.75}$$

$$= \frac{i}{\hbar}\frac{e^2\hbar}{2m_e\omega_0}\lim_{\eta\to 0}\left(h^L(i(\omega_0+\omega)+\eta) - (h^*)^L\left(i(\omega-\omega_0)+\eta\right)\right) \tag{3.76}$$

$$= \frac{i}{\hbar}\frac{e^2\hbar}{2m_e\omega_0}\lim_{\eta\to 0}\left(\frac{1}{i(\omega_0+\omega)+\eta+\kappa^L(i(\omega_0+\omega)+\eta)}\right. \tag{3.77}$$

$$\left. -\frac{1}{i(\omega-\omega_0)+\eta+\left(\kappa^L(-i(\omega-\omega_0)+\eta)\right)^*}\right) \tag{3.78}$$

Plugging in the expression for $\kappa^L(s)$, we continue:

$$\chi(\omega) = \frac{i}{\hbar}\frac{e^2\hbar}{2m_e\omega_0}\lim_{\eta\to 0}\left(\frac{1}{i(\omega_0+\omega)+\eta+\frac{1}{\hbar^2}\sum_i\frac{|g_i|^2}{i(\omega+\omega_i)+\eta}} - \frac{1}{i(\omega-\omega_0)+\eta+\frac{1}{\hbar^2}\sum_i\frac{|g_i|^2}{i(\omega-\omega_i)+\eta}}\right) \tag{3.79}$$

$$= \frac{i}{\hbar}\frac{e^2\hbar}{2m_e\omega_0}\lim_{\eta\to 0}\left(\frac{1}{i(\omega_0+\omega)+\eta+\frac{1}{\hbar^2}\sum_i|g_i|^2\left(\frac{-i}{\omega+\omega_i}+\pi\delta\left(\omega+\omega_i\right)\right)}\right. \tag{3.80}$$

$$\left. -\frac{1}{i(\omega-\omega_0)+\eta+\frac{1}{\hbar^2}\sum_i|g_i|^2\left(\frac{-i}{\omega-\omega_i}+\pi\delta\left(\omega-\omega_i\right)\right)}\right) \tag{3.81}$$

At this point, we define the 'generalized' decay rate and energy shift, $\Gamma(\omega)$ and $\Delta(\omega)$ respectively, as follows:

$$\Gamma(\omega) = \frac{1}{\hbar^2}\sum_i|g_i|^2\pi\delta\left(\omega+\omega_i\right) \tag{3.82}$$

$$\Delta(\omega) = \frac{1}{\hbar^2}\sum_i|g_i|^2\left(\frac{-1}{\omega+\omega_i}\right) \tag{3.83}$$

$\chi(\omega)$ then reduces to:

$$\chi(\omega) = \frac{ie^2}{2m_e\omega_0}\lim_{\eta\to 0}\left(\frac{1}{i\left(\omega_0+\omega+\Delta(\omega)\right)+\Gamma(\omega)+\eta} - \frac{1}{i\left(\omega-\omega_0-\Delta(-\omega)\right)+\Gamma(-\omega)+\eta}\right) \tag{3.84}$$

The imaginary part, $\chi''(\omega)$, is given by:

$$\chi''(\omega) = \frac{e^2}{2m_e\omega_0}\left(\frac{\Gamma(\omega)}{\left(\omega+\omega_0+\Delta(\omega)\right)^2+\Gamma^2(\omega)} - \frac{\Gamma(-\omega)}{\left(\omega-\omega_0-\Delta(-\omega)\right)^2+\Gamma^2(-\omega)}\right) \tag{3.85}$$

We have successfully derived an exact $\chi''(\omega)$ for our model.

Discussion

Eq. (3.85) looks like the general lineshape for the tunneling problem that was derived in Chap. 2. If the 'generalized' decay rate and energy shift are assumed to be constant in ω, Eq. (3.85) reduces to a difference of two Lorentzian lineshapes which yields a reciprocal polynomial absorption lineshape. However, the generalized quantities are usually not constant away from the peaks, so the tails should be expected to be non-Lorentzian.

A disappointing feature of Eq. (3.85), however, is that it too doesn't have any phonon-population-induced temperature dependence, just like the result of the Lorentz-Lorenz model. The simple linear dynamics of the physical operators in this system, which we were able to solve exactly using standard techniques, seems to be insufficient to introduce temperature dependence in the absorption lineshape. This motivates us to next look at a more complicated model system whose physical variables do indeed obey non-linear dynamics.

3.7 Optical absorption in a 2-level system coupled to a phonon reservoir

Essence of this section: In this section, we apply the Green-Kubo formula to the most complicated problem in this chapter—optical absorption by an electron in a 2-level system that is coupled to a reservoir of phonons. The phonons in the crystal interact with the 2-level system and broaden its absorption spectrum.

Relevance of absorption in 2-level systems

In this section, we will study the optical absorption of an electron in a 2-level system or atom that is perturbed by a phonon reservoir. This system can be used to approximate the absorption in a real semiconductor. When a probe light signal is incident on a semiconductor, transitions occur between conduction-band and valence-band states that have the same Bloch vectors k. Each such 'compatible' pair of valence and conduction band states is effectively a 2-level system. A zeroth-order method to obtain the full absorption spectrum of the crystal is to sum the spectra of all the individual 'compatible' pairs of valence band and conduction band states, that is, by summing the spectra of many 2-level systems. Of course, this is only an approximation, and doesn't take into account the additional broadening of conduction and valence band states in crystals that occurs due to intra-band electron scattering caused by the phonons. More importantly, however, we think one needs to exercise more care while adopting this kind of procedure. The full absorption spectrum is possibly not a simple sum of individual 2-level spectra but something like the squared absolute value of the sum of the individual complex square-root responses of all the component 2-level systems. Nevertheless, computing the 2-level system spectrum is likely to be an important step in the final derivation of semiconductor spectra, so we go ahead with it now.

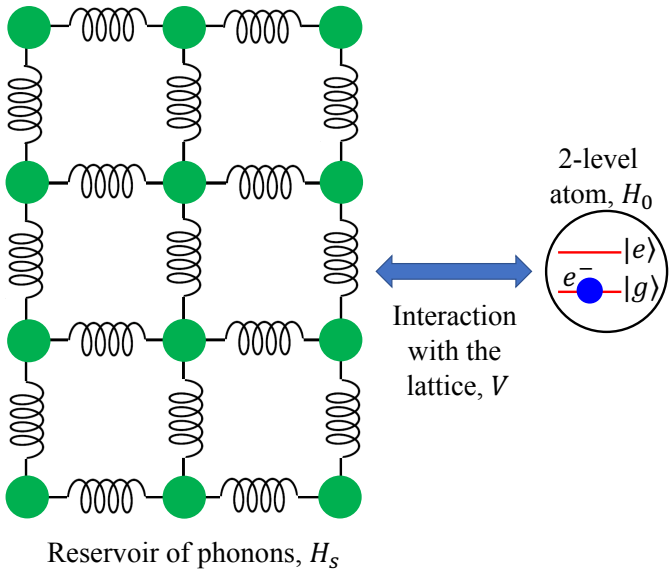

Figure 3.2: The electron is in a 2-level system and is also coupled to a large crystal which is at thermal equilibrium. The phonons in the crystal interact with the electron and broaden its absorption spectrum. The phonons are described by the Hamiltonian H_s, the 2-level system by H_0, and their interaction by V.

Hamiltonian

All 2-level system Hermitian operators can be expressed in terms of the Pauli matrices and the identity matrix. The three Pauli matrices are:

$$S_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; S_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \; S_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.86}$$

Let the energy of the excited state, $|e\rangle$, of the 2-level atom be E, and the energy of the ground state, $|g\rangle$, be $-E$. Then, the Hamiltonian H_0 of the 2-level system is:

$$H_0 = ES_z \tag{3.87}$$

The phonon bath has the same Hamiltonian as before:

$$H_s = \sum_i \hbar\omega_i \left(a_i^\dagger a_i + \frac{1}{2} \right) \tag{3.88}$$

For the interaction V, we shall again consider only those processes which involve the excitation of the 2-level atom upon the destruction of a phonon or the de-excitation of the 2-level atom upon the creation of a phonon. Defining the 2-level atom 'creation' and 'destruction'

operators, $S_+ = (S_x + iS_y)/2$ and $S_- = (S_x - iS_y)/2$, we can write V as:

$$V = \sum_i \left(g_i a_i S_+ + g_i^* a_i^\dagger S_- \right) \tag{3.89}$$

where g_i is the strength of the coupling between the 2-level atom and the i-th phonon mode. The full Hamiltonian is then $H = H_0 + H_s + V$.

Heisenberg equations

The Heisenberg equations are:

$$\frac{d}{dt}S_+(t) = \frac{i}{\hbar}\left[2ES_+(t) - \sum_i g_i^* a_i^\dagger(t) S_z(t)\right] \tag{3.90}$$

$$\frac{d}{dt}S_-(t) = \frac{i}{\hbar}\left[-2ES_-(t) + \sum_i g_i a_i(t) S_z(t)\right] \tag{3.91}$$

$$\frac{d}{dt}S_z(t) = \frac{2i}{\hbar}\left[-\sum_i g_i a_i(t) S_+(t) + \sum_i g_i^* a_i^\dagger(t) S_-(t)\right] \tag{3.92}$$

$$\frac{d}{dt}a_i(t) = \frac{i}{\hbar}\left[-\hbar\omega_i a_i(t) - g_i^* S_-(t)\right] \tag{3.93}$$

We follow the exact same procedure again and introduce the 'slowly-varying' operators:

$$a_i(t) = \hat{a}_i(t)e^{-i\omega_i t},\ S_+(t) = \hat{S}_+(t)e^{i2Et/\hbar},\ S_-(t) = \hat{S}_-(t)e^{-i2Et/\hbar} \tag{3.94}$$

After some algebra which involves plugging in these operators into the four equations, integrating the differential equation for $a_i(t)$ from 0 to t, and substituting the resultant $a_i(t)$ expression into the three system spin equations, we get:

$$\frac{d}{dt}\hat{S}_+(t) = -\frac{i}{\hbar}\sum_i g_i^* a_i^\dagger e^{i(\omega_i - 2E/\hbar)(t)} S_z(t) + \left(\frac{1}{\hbar^2}\int_0^t d\tau \sum_i |g_i|^2 e^{i(\omega_i - 2E/\hbar)\tau}\hat{S}_+(t-\tau)\right) S_z(t) \tag{3.95}$$

$$\frac{d}{dt}\hat{S}_-(t) = \frac{i}{\hbar}\sum_i g_i a_i e^{-i(\omega_i - 2E/\hbar)(t)} S_z(t) + \left(\frac{1}{\hbar^2}\int_0^t d\tau \sum_i |g_i|^2 e^{-i(\omega_i - 2E/\hbar)\tau}\hat{S}_-(t-\tau)\right) S_z(t) \tag{3.96}$$

$$\frac{d}{dt}S_z(t) = \frac{2i}{\hbar}\left(-\sum_i g_i a_i e^{-i(\omega_i - 2E/\hbar)(t)}\hat{S}_+(t) + \left(\frac{i}{\hbar}\int_0^t d\tau \sum_i |g_i|^2 e^{-i(\omega_i-2E/\hbar)\tau}\hat{S}_-(t-\tau)\right)\hat{S}_+(t) + \sum_i g_i^* a_i^\dagger e^{i(\omega_i-2E/\hbar)(t)}\hat{S}_-(t) + \left(\frac{i}{\hbar}\int_0^t d\tau \sum_i |g_i|^2 e^{i(\omega_i-2E/\hbar)\tau}\hat{S}_+(t-\tau)\right)\hat{S}_-(t)\right) \tag{3.97}$$

These equations indeed look quite unwieldy but a prominent feature that stands out upon casual inspection is their nonlinear nature. The terms that contain integrals all have products of spin operators in their integrands.

Markov approximation with time-dependent coefficients

Since it seems difficult to solve the coupled nonlinear equations directly, we recall the Markov approximation that we made in Chapter 2 and see if it can be applied. The first equation, for instance, has a memory-kernel convolution in parenthesis in the second term on the right-hand side, with an additional $S_z(t)$ outside the parenthesis. The same structure appears in the other equations. The memory function, $M(t)$, that we are referring to is:

$$M(t) = \frac{1}{\hbar^2}\sum_i |g_i|^2 e^{i(\omega_i - 2E/\hbar)t} = r(t) + is(t) \tag{3.98}$$

where we have labelled the real and imaginary part of the memory function $r(t)$ and $s(t)$ respectively. If we make the usual Markov assumption that the memory function is very thin in t, the $\hat{S}_+(t-\tau)$ in the integral in the first equation becomes approximately equal to $\hat{S}_+(t)$ for the purposes of the integration and can be pulled out of it. Next, the $\hat{S}_+(t)$ term multiplies the $S_z(t)$ term outside the integral and gives $-\hat{S}_+(t)$. We perform the same two steps, Markov approximation with on operator getting pulled out, and simplification of possible products of operators, in all the other equations too. Setting $f(t) = \int_0^t d\tau\ r(\tau)$ and $g(t) = \int_0^t d\tau\ s(\tau)$, we finally get:

$$\frac{d}{dt}\hat{S}_+(t) = -\frac{i}{\hbar}\sum_i g_i^* a_i^\dagger e^{i(\omega_i-2E/\hbar)(t)}S_z(t) - (f(t) + ig(t))\,\hat{S}_+(t) \tag{3.99}$$

$$\frac{d}{dt}\hat{S}_-(t) = \frac{i}{\hbar}\sum_i g_i a_i e^{-i(\omega_i-2E/\hbar)(t)}S_z(t) + (f(t) - ig(t))\,\hat{S}_-(t) \tag{3.100}$$

$$\frac{d}{dt}S_z(t) = 2\left(-\frac{i}{\hbar}\sum_i g_i a_i e^{-i(\omega_i-2E/\hbar)(t)}\hat{S}_+(t) - f(t) - ig(t)S_z(t) + \frac{i}{\hbar}\sum_i g_i^* a_i^\dagger e^{i(\omega_i-2E/\hbar)(t)}\hat{S}_-(t)\right) \tag{3.101}$$

If one were to apply the Markov approximation with time-independent coefficients instead, we would have gotten the same equations, but with $f(t)$ replaced by a rate Γ, and $g(t)$ replaced by a frequency shift Δ. We shall retain $f(t)$ but replace $g(t)$ with Δ towards the end of the next subsection.

Dipole commutator autocorrelation

The dipole operator of the 2-level system is given by $p = dS_+ + d^*S_-$, where d is the matrix element of the dipole moment between the excited and ground states, $d = \langle e|p|g\rangle$. The commutator of interest for optical absorption is $[dS_+(t) + d^*S_-(t), dS_+ + d^*S_-]$.

From the first two differential equations, it is clear that the computation of $[S_+(t), A]$ and $[S_-(t), A]$ for any A involves computing $[S_z(t), A]$. The equation that governs $[S_+(t), S_+]$ is:

$$\frac{d}{dt}\left[\hat{S}_+(t), S_+\right] = -\frac{i}{\hbar}\sum_i g_i^* a_i^\dagger e^{i(\omega_i - 2E/\hbar)(t)}\left[S_z(t), S_+\right] - (f(t) + ig(t))\left[\hat{S}_+(t), S_+\right] \tag{3.102}$$

The term $[S_z(t), S_+]$ is computed as follows:

$$\begin{aligned}\frac{d}{dt}\left[S_z(t), S_+\right] = 2\Bigg(&-\frac{i}{\hbar}\sum_i g_i a_i e^{-i(\omega_i - 2E/\hbar)(t)}\left[\hat{S}_+(t), S_+\right] - ig(t)\left[S_z(t), S_+\right] \\ &+\frac{i}{\hbar}\sum_i g_i^* a_i^\dagger e^{i(\omega_i - 2E/\hbar)(t)}\left[\hat{S}_-(t), S_+\right]\Bigg)\end{aligned} \tag{3.103}$$

There is a similar equation for $[S_z(t), S_-]$. Setting $G(t) = \int_0^t d\tau\ g(\tau)$ and solving Eq. (3.103) using the standard integrating factor technique gives us:

$$\begin{aligned}\left[S_z(t), S_+\right] =& 2S_+ e^{-2iG(t)} + 2e^{-2iG(t)}\int_0^t d\tau\ e^{2iG(\tau)}\left(-\frac{i}{\hbar}\sum_i g_i a_i e^{-i(\omega_i - 2E/\hbar)\tau}\left[\hat{S}_+(\tau), S_+\right]\right) \\ &+ 2e^{-2iG(t)}\int_0^t d\tau\ e^{2iG(\tau)}\left(\frac{i}{\hbar}\sum_i g_i^* a_i^\dagger e^{i(\omega_i - 2E/\hbar)\tau}\left[\hat{S}_-(\tau), S_+\right]\right)\end{aligned} \tag{3.104}$$

Plugging this expression into Eq. (3.102) gives us:

$$\begin{aligned}\frac{d}{dt}\left[\hat{S}_+(t), S_+\right] =& -(f(t) + ig(t))\left[\hat{S}_+(t), S_+\right] - \frac{i}{\hbar}\sum_i g_i^* a_i^\dagger e^{i(\omega_i - 2E/\hbar)(t)}\left(2S_+ e^{-2iG(t)}\right) \\ &+ 2e^{-2iG(t)}\int_0^t d\tau\ e^{2iG(\tau)}\left(-\frac{1}{\hbar^2}\sum_{ij} g_i^* g_j a_i^\dagger a_j e^{i(\omega_i - 2E/\hbar)t} e^{-i(\omega_j - 2E/\hbar)\tau}\left[\hat{S}_+(\tau), S_+\right]\right) \\ &+ 2e^{-2iG(t)}\int_0^t d\tau\ e^{2iG(\tau)}\left(\frac{1}{\hbar^2}\sum_{ij} g_i^* g_j^* a_i^\dagger a_j^\dagger e^{i(\omega_i - 2E/\hbar)t} e^{i(\omega_j - 2E/\hbar)\tau}\left[\hat{S}_-(\tau), S_+\right]\right)\end{aligned} \tag{3.105}$$

We now take the average of both sides of this expression in the equilibrium thermal density matrix, $\rho = e^{-H/kT}/Z$. Since the phonon bath is too large to be significantly affected by the 2-level system, it is reasonable to assume that the full density matrix can be decomposed into a tensor product, $\rho = \rho_0 \otimes \rho_s$. Here, ρ_0 and ρ_s are the reduced density matrices of the 2-level system and the phonon bath respectively, and ρ_s is given by $\rho_s = e^{-H_s/kT}/Z_s$.

Now, the computation of the averages of the terms in Eq. (3.105) becomes straightforward:

$$\left\langle a_i^\dagger S_+ \right\rangle = \left\langle a_i^\dagger \right\rangle \langle S_+ \rangle = 0 \tag{3.106}$$

$$\left\langle a_i^\dagger a_j \left[\hat{S}_+(\tau), S+\right] \right\rangle = \left\langle a_i^\dagger a_j \right\rangle \left\langle \left[\hat{S}_+(\tau), S+\right] \right\rangle = \langle n_i \rangle \, \delta_{ij} \left\langle \left[\hat{S}_+(\tau), S+\right] \right\rangle \tag{3.107}$$

$$\left\langle a_i^\dagger a_j^\dagger \left[\hat{S}_-(\tau), S+\right] \right\rangle = \left\langle a_i^\dagger a_j^\dagger \right\rangle \left\langle \left[\hat{S}_-(\tau), S+\right] \right\rangle = 0 \tag{3.108}$$

We then have:

$$\begin{aligned} \frac{d}{dt} \left\langle \left[\hat{S}_+(t), S_+\right] \right\rangle = & -\left(f(t) + ig(t)\right) \left\langle \left[\hat{S}_+(t), S_+\right] \right\rangle \\ & + 2e^{-2iG(t)} \int_0^t d\tau \; e^{2iG(\tau)} \left(-\frac{1}{\hbar^2} \sum_i |g_i|^2 \langle n_i \rangle \, e^{i(\omega_i - 2E/\hbar)(t-\tau)} \left\langle \left[\hat{S}_+(\tau), S_+\right] \right\rangle \right) \end{aligned} \tag{3.109}$$

Next, in the second term above, we replace $g(t)$ with its value at $t = \infty$ which is nothing but Δ, giving us $G(t) = \Delta t$. We also make the substitution $f(t) + ig(t) = \int_0^t d\tau \; M(\tau)$ in the first term to get:

$$\begin{aligned} \frac{d}{dt} \left\langle \left[\hat{S}_+(t), S_+\right] \right\rangle = & -\left(\int_0^t d\tau \; \left(\frac{1}{\hbar^2} \sum_i |g_i|^2 e^{i(\omega_i - 2E/\hbar)\tau} \right) \right) \left\langle \left[\hat{S}_+(t), S_+\right] \right\rangle \\ & - 2 \int_0^t d\tau \; \left(\frac{1}{\hbar^2} \sum_i |g_i|^2 \langle n_i \rangle \, e^{i(\omega_i - 2E/\hbar - 2\Delta)\tau} \right) \left\langle \left[\hat{S}_+(t-\tau), S_+\right] \right\rangle \end{aligned} \tag{3.110}$$

Performing the Markov approximation on the second term and pulling $\left\langle \left[\hat{S}_+(t-\tau), S_+\right] \right\rangle$ out of the integral, we get:

$$\frac{d}{dt} \left\langle \left[\hat{S}_+(t), S_+\right] \right\rangle = -\left(\int_0^t d\tau \; \left(\frac{1}{\hbar^2} \sum_i |g_i|^2 e^{i(\omega_i - 2E/\hbar)\tau} \left(1 + 2 \langle n_i \rangle \, e^{-2i\Delta\tau}\right) \right) \right) \left\langle \left[\hat{S}_+(t), S_+\right] \right\rangle \tag{3.111}$$

$$= -U(t) \left\langle \left[\hat{S}_+(t), S_+\right] \right\rangle \tag{3.112}$$

where we defined a new function $U(t)$ to be equal to the time-dependent coefficient in parenthesis. We are finally done with the long sequence of manipulations, this is our final equation

of motion. It is a simple linear differential equation with a time-dependent coefficient. In this particular case, however, the solution is $\left\langle \left[\hat{S}_+(t), S_+ \right] \right\rangle = 0$ because the initial condition is $[S_+, S_+] = 0$. But $\left\langle \left[\hat{S}_+(t), S_- \right] \right\rangle$ also satisfies the exact same differential equation, with the initial condition being $\left\langle \left[\hat{S}_+(0), S_- \right] \right\rangle = \langle S_z \rangle$. Therefore, we have:

$$\left\langle \left[\hat{S}_+(t), S_- \right] \right\rangle = \langle S_z \rangle \, e^{-\int_0^t d\tau \; U(\tau)} \tag{3.113}$$

Next, it turns out that $\left\langle \left[\hat{S}_-(t), S_- \right] \right\rangle = 0$ too. The final commutator, $\left\langle \left[\hat{S}_-(t), S_+ \right] \right\rangle$, satisfies:

$$\frac{d}{dt} \left\langle \left[\hat{S}_-(t), S_+ \right] \right\rangle \tag{3.114}$$

$$= \left(\int_0^t d\tau \; \left(\frac{1}{\hbar^2} \sum_i |g_i|^2 e^{-i(\omega_i - 2E/\hbar)\tau} \left(1 - 2 \left(1 + \langle n_i \rangle \right) e^{-2i\Delta\tau} \right) \right) \right) \left\langle \left[\hat{S}_-(t), S_+ \right] \right\rangle \tag{3.115}$$

$$= -W(t) \left\langle \left[\hat{S}_-(t), S_+ \right] \right\rangle \tag{3.116}$$

where we introduced a new function $W(t)$ that is the negative of the time-dependent coefficient in parenthesis. For the initial condition, $\left\langle \left[\hat{S}_-(0), S_+ \right] \right\rangle = -\langle S_z \rangle$, the solution is:

$$\left\langle \left[\hat{S}_-(t), S_+ \right] \right\rangle = -\langle S_z \rangle \, e^{-\int_0^t d\tau \; W(\tau)} \tag{3.117}$$

In summary, the full dipole commutator autocorrelation is:

$$\langle [p(t), p] \rangle = |d|^2 \langle S_z \rangle \left(e^{2iEt/\hbar} e^{-\int_0^t d\tau \; U(\tau)} - e^{-2iEt/\hbar} e^{-\int_0^t d\tau \; W(\tau)} \right) \tag{3.118}$$

Susceptibility $\chi(\omega)$

In order to get an interpretable expression for the susceptibility, we set $\Delta = 0$ inside both $U(t)$ and $W(t)$. We define a new memory function $\hat{M}(t)$, its real part $\hat{r}(t)$, its imaginary part $\hat{s}(t)$, and their time integrals $\theta(t)$ and $\pi(t)$ in the following manner:

$$U(t) = \int_0^t d\tau \; \left(\frac{1}{\hbar^2} \sum_i |g_i|^2 e^{i(\omega_i - 2E/\hbar)\tau} \left(1 + 2 \langle n_i \rangle \right) \right) \tag{3.119}$$

$$= \int_0^t d\tau \; \hat{M}(\tau) = \int_0^t d\tau \; \left(\hat{r}(\tau) - i\hat{s}(\tau) \right) = \theta(t) - i\pi(t) \tag{3.120}$$

$\hat{M}(t)$ can be decomposed into a temperature-dependent and temperature-independent parts:

$$\hat{M}(t) = \hat{M}_{\text{temp}}(t) + \hat{M}_{\text{notemp}}(t) \tag{3.121}$$

$$= \frac{1}{\hbar^2} \sum_i 2|g_i|^2 e^{i(\omega_i - 2E/\hbar)\tau} \langle n_i \rangle + \frac{1}{\hbar^2} \sum_i |g_i|^2 e^{i(\omega_i - 2E/\hbar)\tau} \tag{3.122}$$

indicating that $\theta(t)$ can also be decomposed the same way. In the same notation, $W(t)$ turns out to be:

$$W(t) = \int_0^t d\tau \ \left(\frac{1}{\hbar^2} \sum_i |g_i|^2 e^{-i(\omega_i - 2E/\hbar)\tau} \left(1 + 2\left\langle n_i \right\rangle\right) \right) = \theta(t) + i\pi(t) \tag{3.123}$$

Defining $\Theta(t) = \int_0^t d\tau \ \theta(\tau)$ and $\Pi(t) = \int_0^t d\tau \ \pi(\tau)$, we have:

$$\langle [p(t), p] \rangle = 2i|d|^2 \left\langle S_z \right\rangle e^{-\Theta(t)} \sin\left(\Pi(t) + 2Et/\hbar\right) \tag{3.124}$$

In the above, $\Theta(t)$ is a sum of $\Theta_{\text{temp}}(t)$ and $\Theta_{\text{notemp}}(t)$ implying that $e^{-\Theta(t)}$ is a product of a temperature-dependent piece and a temperature-independent piece. **This temperature-dependence is non-trivial because it involves phonon populations, not electron populations.**

The final expression for $\chi''(\omega)$ is:

$$\chi''(\omega) = \frac{|d|^2 \left\langle S_z \right\rangle}{\hbar} \int_{-\infty}^{\infty} dt \ ie^{-\Theta(t)} \sin\left(\Pi(t) + 2Et/\hbar\right) e^{-i\omega t} \tag{3.125}$$

Again we assume that the 'phase' function $\Pi(t)$, just like $G(t)$ earlier, is a simple linear function $\Pi(t) = \zeta t$. Referring to the Fourier transform of $e^{-\Theta(t)}$ as $L(\omega)$ we have:

$$\chi''(\omega) = \frac{|d|^2 \left\langle S_z \right\rangle}{2\hbar} \left(L(\omega - \zeta - 2E/\hbar) - L(\omega + \zeta + 2E/\hbar) \right) \tag{3.126}$$

Since $e^{-\Theta(t)}$ is a product of a temperature-dependent piece and a temperature-independent piece, $L(\cdot)$ is the convolution of a temperature-dependent piece and a temperature-independent piece. This lineshape is interpreted in the next subsection.

Discussion

We draw conclusions for this case in exactly the same spirit as that of our discussion in Chapter 2, Section 2.7. There, we had noted that an 'appropriate' choice of the wire density of states $\rho(E)$ and the dot-wire interaction strength $|V(E)|^2$ would lead to rapidly decaying lineshapes.

As in that case, $\theta(t)$ and $\pi(t)$ here arise from the time integral of the memory function $\hat{M}(t)$ in the integrand of Eq. (3.120). If $\hat{M}(t)$ is infinitely differentiable and has an even real part $\hat{r}(t)$ and an odd imaginary part $\hat{s}(t)$, then both $\theta(t)$ and $\pi(t)$ will be infinitely differentiable with $\theta(t)$ containing only odd powers of t in its Taylor series and $\pi(t)$ containing only even powers of t. This will imply that $\Theta(t)$ also contains only even powers of t in its Taylor series, which in turn tells us that $e^{-\Theta(t)}$ is in the Schwartz space. Hence, $L(\omega)$ also belongs to the Schwartz space and is rapidly decaying in ω. Moreover, since $\theta(t)$ contains temperature dependence in the form of the phonon thermal occupations, $L(\omega)$ too has a T dependence!

This is finally a breakthrough from the previous models where we repeatedly obtained T-independent absorption lineshapes (setting aside the simple temperature dependence of the electron Fermi occupations).

Before ending this section, we rewrite the memory function $\hat{M}(t)$ as follows:

$$\hat{M}(t) = \frac{1}{\hbar^2} \int d\omega \ \rho(\omega)|g(\omega)|^2 \left(1 + 2\left\langle n(\omega)\right\rangle\right) e^{i(\omega - 2E/\hbar)t} \tag{3.127}$$

Since $\hat{M}(t)$ is the Fourier transform of $\rho(\omega)|g(\omega)|^2 \left(1 + 2\left\langle n(\omega)\right\rangle\right)$, we suspect Fourier theory demands that this product has to be continuous and rapidly decaying in ω in order to produce an infinitely differentiable $\hat{M}(t)$, and consequently, a rapidly decaying lineshape $L(\omega)$ by the reasoning in the previous paragraph. In summary, we have successfully derived a rapidly decaying spectral lineshape $L(\omega)$, but unfortunately, we are still unable to deduce the exact asymptotic formula for the rapidly decaying tails of $L(\omega)$ at our level of analysis.

3.8 Optical absorption in a general crystal

Essence of this section: In the penultimate section of this chapter, we write down the Heisenberg equations for the general problem of optical absorption by an electron in a periodic crystal. We do this simply as a means of depicting the immense complexity of the equations, and to emphasize the need for clever simplifications and/or reformulations in terms of 2-level systems in the future in order to obtain a general derivation of the Urbach tail.

Hamiltonian

We now shift our attention to the case of several electrons in a general crystal. The Hamiltonian is given by:

$$H_e = \sum_{kn} E_{kn} c_{kn}^\dagger c_{kn} \tag{3.128}$$

$$H_{ph} = \sum_{q\alpha} \hbar\omega_{q\alpha} \left(a_{q\alpha}^\dagger a_{q\alpha} + \frac{1}{2}\right) \tag{3.129}$$

$$H_{eph} = \sum_{q\alpha k'n'kn} M(q\alpha k'n'kn) c_{k'n'}^\dagger c_{kn} \left(a_{q\alpha} + a_{-q\alpha}^\dagger\right) \tag{3.130}$$

where E_{kn} is the energy of the electron state with crystal momentum k in band n, $\omega_{q\alpha}$ is the frequency of the phonon state with crystal momentum q in band α, and the operators have their usual meanings. The matrix element $M(q\alpha k'n'kn)$ is the strength of the interaction that causes electrons to be scattered between the kn and $k'n'$ states by phonons in the $q\alpha$ state. The exact form of this matrix element can be found in the solid-state physics textbook by Cohen and Louie [91].

Heisenberg equations

$$\frac{dc_{kn}(t)}{dt} = -\frac{i}{\hbar}E_{kn}c_{kn}(t) - \frac{i}{\hbar}\sum_{q\alpha k'n'} M(q\alpha knk'n')c_{k'n'}(t)\left(a_{q\alpha}(t) + a^{\dagger}_{-q\alpha}(t)\right) \tag{3.131}$$

$$\frac{da_{q\alpha}(t)}{dt} = -i\omega_{q\alpha}a_{q\alpha}(t) - \frac{i}{\hbar}\sum_{k'n'kn} M(-q\alpha k'n'kn)c^{\dagger}_{k'n'}(t)c_{kn}(t) \tag{3.132}$$

Define:

$$c_{kn}(t) = \hat{c}_{kn}(t)e^{-i\omega_{kn}t} \tag{3.133}$$

$$a_{q\alpha}(t) = \hat{a}_{q\alpha}(t)e^{-i\omega_{q\alpha}t} \tag{3.134}$$

Rewriting the equations, we get:

$$\frac{d\hat{c}_{kn}(t)}{dt} = -\frac{i}{\hbar}\sum_{q\alpha k'n'} M(q\alpha knk'n')\hat{c}_{k'n'}(t)\left(\hat{a}_{q\alpha}(t)e^{i(\omega_{kn}-\omega_{k'n'}-\omega_{q\alpha})t} + \hat{a}^{\dagger}_{-q\alpha}(t)e^{i(\omega_{kn}-\omega_{k'n'}+\omega_{-q\alpha})t}\right) \tag{3.135}$$

$$\frac{d\hat{a}_{q\alpha}(t)}{dt} = -\frac{i}{\hbar}\sum_{k'n'kn} M(-q\alpha k'n'kn)\hat{c}^{\dagger}_{k'n'}(t)\hat{c}_{kn}(t)e^{i(\omega_{q\alpha}+\omega_{k'n'}-\omega_{kn})t} \tag{3.136}$$

$$\frac{d\hat{a}^{\dagger}_{-q\alpha}(t)}{dt} = \frac{i}{\hbar}\sum_{k'n'kn} M(-q\alpha knk'n')\hat{c}^{\dagger}_{kn}(t)\hat{c}_{k'n'}(t)e^{i(-\omega_{-q\alpha}-\omega_{k'n'}+\omega_{kn})t} \tag{3.137}$$

Elimination of phonon variables

Eliminating the phonon variables, we get:

$$\begin{aligned}\frac{d\hat{c}_{kn}(t)}{dt} = &-\frac{1}{\hbar^2}\sum_{q\alpha k'n'k'''n'''k''n''} M(q\alpha knk'n')M(-q\alpha k'''n'''k''n'')\hat{c}_{k'n'}(t)e^{i(\omega_{kn}-\omega_{k'n'}-\omega_{q\alpha})t}\\ &\int_0^t d\tau\hat{c}^{\dagger}_{k'''n'''}(\tau)\hat{c}_{k''n''}(\tau)e^{i(\omega_{q\alpha}+\omega_{k'''n'''}-\omega_{k''n''})\tau}\\ &+\frac{1}{\hbar^2}\sum_{q\alpha k'n'k'''n'''k''n''} M(q\alpha knk'n')M(-q\alpha k''n''k'''n''')\hat{c}_{k'n'}(t)e^{i(\omega_{kn}-\omega_{k'n'}-\omega_{q\alpha})t}\\ &\int_0^t d\tau\hat{c}^{\dagger}_{k'''n'''}(\tau)\hat{c}_{k''n''}(\tau)e^{i(\omega_{q\alpha}+\omega_{k'''n'''}-\omega_{k''n''})\tau}\\ &-\frac{i}{\hbar}\sum_{q\alpha k'n'} M(q\alpha knk'n')\hat{c}_{k'n'}(t)\left(\hat{a}_{q\alpha}(0)e^{i(\omega_{kn}-\omega_{k'n'}-\omega_{q\alpha})t} + \hat{a}^{\dagger}_{-q\alpha}(0)e^{i(\omega_{kn}-\omega_{k'n'}+\omega_{-q\alpha})t}\right)\end{aligned} \tag{3.138}$$

We note the tremendous complexity of this differential equation and admit that we are not aware of a method of solving it. Given the near universality of the Urbach tail, we believe

it is not necessary to actually solve this problem in this brute-force fashion, and that an elegant derivation exists through simpler model systems. It should be possible to combine the spectra of the individual 2-level systems that make up a semiconductor in a certain way to derive the final Urbach tail of the entire material. However, we do not yet know the exact way to do this.

3.9 Conclusion

The summary of this chapter is:

1. The absorption coefficient $\alpha(\omega)$ is directly proportional to the imaginary part of the susceptibility $\chi''(\omega)$.

2. The imaginary susceptibility $\chi''(\omega)$ is obtained by taking the Fourier transform of the system's dipole moment autocorrelation function. This is the Green-Kubo relation.

3. The traditional Lorentz-Lorenz model and Kubo's statistical model fail to give us Urbach tails.

4. The absorption by an electron in a harmonic well that is coupled to a bath of phonons, computed via the Green-Kubo relation, fails to display temperature-dependent exponential spectral tails. This might be due to the simple linear nature of the dipole moment dynamics.

5. The exact description of the absorption in a semiconductor, as presented in the previous section, is highly complex, and we think such a brute-force approach is likely to not yield a derivation of the Urbach tail.

6. Our final complete calculation, the absorption by an electron in a 2-level system that is coupled to a bath of phonons, does succeed in producing temperature-dependent exponential spectral tails. The temperature dependence arises from the thermal occupations of the various modes in the phonon bath. We believe the nonlinear coupling of the physical variables involved—the spin operators and the phonon operators—gave rise to this temperature dependence. We could not, however, isolate the exact formula of the spectral tail in terms of the temperature at the point of this writing. This direction seems quite promising and some more work along these lines could lead to a derivation of the Urbach tail.

7. We conjecture that a good approximation of the semiconductor absorption can be obtained by interpreting it as a collection of many 2-level systems and carefully convolving the susceptibility responses of all of them. Interference of 2-level system spectra with complex coefficients could arise and the spectral tail of the result could be exponential. We recognize this approach as a promising direction for future work on this problem.

Part II

Physics-based Optimization

Chapter 4

Physics principles as Optimization tools

In this part of the thesis, we ask whether physics can be exploited to address optimization and contribute to large-scale industrial problems. As we mentioned in Chapter 1, nature is so structured that almost every fundamental physics principle can be thought of in terms of optimization. We have principles like the 'Principle of Least Action' and the 'Principle of Minimum Power Dissipation' (also called Minimum Entropy Generation) and physics-based algorithms such as annealing and the adiabatic method (also referred to as 'quantum annealing' in the quantum computing community).

In due course, the research community may learn how to use each of these principles to build industrial-scale non-digital optimization accelerators and solvers that offer time and energy benefits in problem solving. We discuss each principle briefly in this chapter but the main focus will be on minimum power dissipation for the remainder of this thesis.

Optimization is generally accompanied by constraints. For example, perhaps the constraint is that the final answers must be restricted to be ± 1. Such a digitally constrained optimization produces answers compatible with any digital computer. So, if we could perhaps construct a physical system that performs optimization subject to this binary constraint, we could then directly integrate it with the digital pipeline of mainstream computation. As a step in this direction, a series of physics-based machines have been created in the physics and engineering community in the past decade to approximately solve the **NP**-hard **Ising problem**. The main insights in our work are that most of these Ising solvers use hardware based on the Principle of Minimum Power Dissipation and that almost all of them implement the well-known **Lagrange Multipliers method for constrained optimization**.

An early work was by Yamamoto et al. in [92] and this was followed by further work from their group [6, 93–95], and other groups [96–103]. These entropy generating machines range from coupled optical parametric oscillators, to RLC electrical circuits, to coupled exciton-polaritons, and silicon photonic coupler arrays. These machines have the advantage that they solve digital problems orders-of-magnitude faster, and in a more energy-efficient manner, than conventional digital chips that are limited by latency and the energy cost [6].

Within the framework of these dissipative machines, constraints are readily included. These machines in fact perform constrained optimization equivalent to the technique of Lagrange multipliers. We illustrate this connection by surveying 5 published physically distinct machines and showing that each minimizes power dissipation in its own way, subject to constraints; in fact, they perform Lagrange multiplier optimization. Digging further into their dynamics, one notices that these physical machines perform local steepest descent in the power dissipation rate. For this reason, they can become stuck in local optima. However, recent work [6] has demonstrated that these systems are also adaptable toward more advanced techniques for approaching a global optimum. Our viewpoint is that, these machines can, at the very least, help us perform rapid searches for local optima, and thus narrow down on the global optimum.

At this point, we note that there are several other streams of work on physical optimization in the literature that we shall not be dealing with in detail in the following chapters. These works include a variety of Lagrange-like continuous-time solvers [104, 105], Memcomputing methods [106], Reservoir Computing [107, 108], adiabatic solvers using Kerr nonlinear oscillators [109], and probabilistic bit logic [110].

In the remainder of the chapter, we discuss all the aforementioned optimization principles of physics in some detail and then move on to the principle of minimum power dissipation. In the next chapter, we discuss how **coupled nonlinear oscillator networks**, the underlying physical prototype in most of the publications we have mentioned, can be used to solve constrained optimization problems through the principle of minimum power dissipation.

4.1 Optimization in Physics: Principles and Algorithms

We survey the minimization principles of physics and the important optimization algorithms derived from them. The aim is to design physical optimization machines that are 'initialization agnostic', that is, machines that converge to the global optimum, or a good local optimum, irrespective of the initial point for the search.

The principle of Least Action

The principle of Least Action is the most fundamental principle in physics. Newton's Laws of Mechanics, Maxwell's Equations of Electromagnetism, Schrödinger's Equation in Quantum Mechanics, and Quantum Field Theory can all be interpreted as minimizing a quantity called Action. For the special case of light propagation, this reduces to the principle of Least Time, as shown in Fig. 4.1.

A conservative system without friction or losses evolves according to the principle of Least Action. The fundamental equations of physics are reversible. A consequence of this reversibility is the Liouville Theorem which states that volumes in the phase space of the system are left unchanged as the system evolves. Since phase space volumes don't change,

it means each initial condition of the system will lead to a different final state. There is no contraction of a set of initial conditions to a smaller set or a single point in the end. Therefore, one has to cleverly design both an initial state and a set of dynamics to take the reversible system from that initial state to the final answer. In other words, we are talking about algorithm design. An algorithm has to be designed and a reversible system that implements that algorithm has to constructed to solve the optimization problem. While many of the problems we encounter already have fast algorithms, we look forward to future computer science breakthroughs in algorithms and quantum computing that would allow the Principle of Least Action to address **NP**-hard problems like the Ising problem.

One needs to keep in mind that, even in reversible computing, there is a final energy cost that arises from recording the result of the computation by overwriting the previous answer bits. This energy cost, first discussed by Landauer [111] and Bennett, depends on the number of bits in the answer N and the temperature T as $kTN \ln 2$. This is only an energy cost, not continuous power dissipation.

An alternative approach to computing would involve physical systems that continuously dissipate power, aiding in the contraction of phase space toward a final solution. This brings us to the principle of Least Power Dissipation.

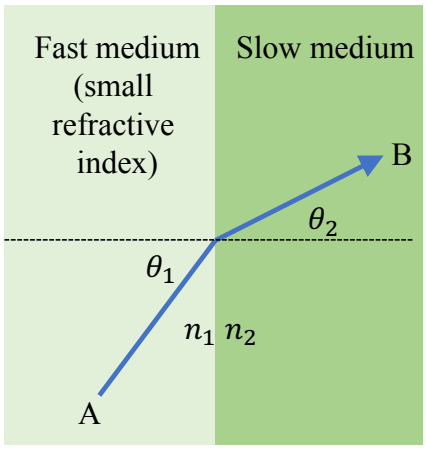

Figure 4.1: The principle of Least Time, a subset of the principle of Least Action: The actual path that light takes to travel from point A to point B is the one that takes the least time to traverse. Recording the correct path entails a small energy cost consistent with the Landauer Limit.

The principle of Least Power Dissipation

If we consider systems that continuously dissipate power, we are led to a second optimization principle in physics, the principle of Least Entropy Generation or Minimum Power Dissipation.

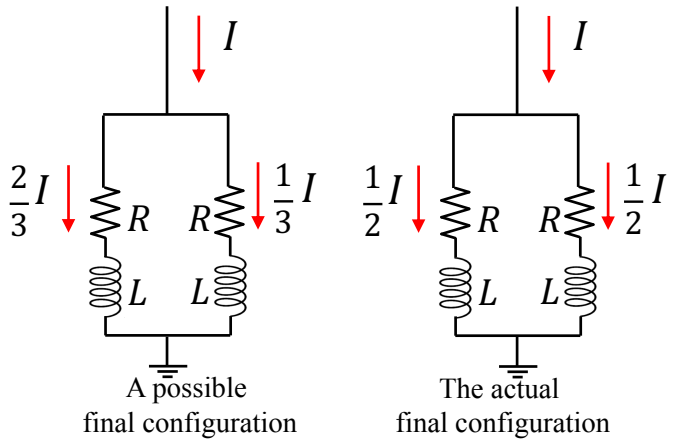

Figure 4.2: The principle of Least Power Dissipation: In a parallel connection, the current distributes itself in a manner that minimizes the power dissipation, subject to the constraint of fixed input current I.

This principle states that any lossy physical system will evolve into a steady-state configuration of currents or fluxes that minimizes the rate of power dissipation subject to the input driving constraints (such as fixed thermodynamic forces, voltage sources, or input power) that are imposed on the system. An early version of this statement is provided by Onsager in his celebrated papers on the reciprocal relations [4]. This was followed by further foundational work on this principle by Prigogine, [112], and de Groot, [113]. This principle is readily seen in action in electrical circuits and is illustrated in Fig. 4.2. When the input driving current into the two branches is fixed at I, the currents rearrange themselves to $I/2$ in each branch in steady state and that configuration of currents minimizes the power dissipation. The general lesson is that, if we can construct a system that has the same power dissipation function as the objective function in our optimization problem, the dynamical variables in that system will settle down in steady state to a configuration that minimizes the objective function, that is, to a solution of our problem. This will be the focus of the next chapter. It should be kept in mind that minimum power dissipation gives significant information only when the system is driven. Else, the system will trivially minimize power dissipation by evolving towards the steady-state configuration where all the dynamical variables are zero. We provide a detailed discussion of this principle in Appendix C.

Since Minimum power dissipation is inherently lossy, volumes in the phase space of the system shrink as the system evolves. This means several different initial conditions of the system all lead to the same final state or small set of final states. There is a contraction of a set of initial conditions to a smaller set or a single point in the end. However, it is still essential to design a *good* contractive dynamics that will take the lossy system to the final answer quickly and in a reliable manner. We study the dynamics of a variety of lossy physical Ising solvers in the literature, and show that almost all of them implement the method of Lagrange

Multipliers in some manner in this thesis. It remains an open question as to whether we can design irreversible dynamics that can solve **NP**-hard problems. However, the natural way in which these solvers perform Lagrange multiplier optimization, and the possibility of time and energy savings, indicates that the primary application of these machines should be as accelerators for the less fundamentally daunting but more power hungry optimization problems that may arise in machine learning.

Physical Annealing; Energy Minimization

Physical annealing is widely used in materials science and metallurgy to correct material defects, and involves the slow cooling of a material starting from a high-temperature. As the cooling proceeds, the system tries to maintain thermodynamic equilibrium by reorganizing itself into the lowest energy minimum in its phase space. The Boltzmann factor $e^{-E/kT}$ tells us that higher energy configurations become less likely at low temperatures. Physically, annealing corresponds to the making and breaking of bonds and the rearrangement of atoms towards configurations of progressively lower potential energy as the temperature is reduced. Due to thermal noise, however, the material never settles down into one fixed configuration but keeps jumping between configurations with similar energies.

This physical phenomenon was adapted into an optimization algorithm called Simulated Annealing several decades ago [114]. The optimization starts out at an initial point and performs gradient descent on the objective function, but with artificial 'thermal' noise added to its dynamics. This noise is chosen according to a 'temperature' parameter that is gradually reduced as the optimization proceeds. Fluctuations due to finite temperature help the algorithm escape from local optima as shown in Fig. 4.3. This procedure leads to global optima as the temperature tends to zero in theory, but for some difficult problems like the **NP**-hard ones, the temperature has to be lowered prohibitively slowly for this to happen.

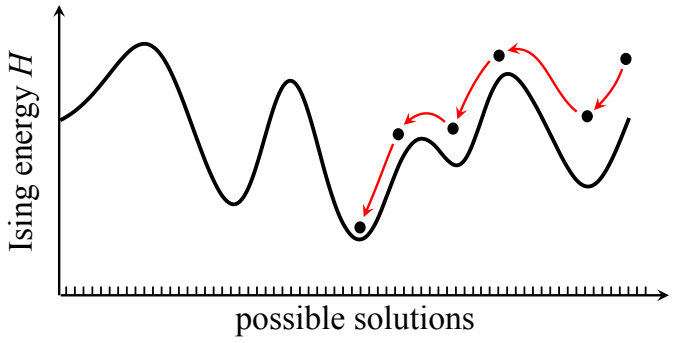

Figure 4.3: Physical annealing involves the slow cooling down of a system. The system performs gradient descent in configuration space with occasional jumps activated by finite temperature. If the cooling is done slowly enough, the system ends up in the ground state of configuration space.

Adiabatic Method

The Adiabatic Method, illustrated in Fig. 4.4, involves the slow transformation of a system from initial conditions that are easily constructed to final conditions that capture the difficult problem at hand.

More specifically, to solve the Ising problem, one initializes the system of spins in the ground state of a simple Hamiltonian and then transforms this Hamiltonian into the Ising problem by slowly varying some system parameters. If the parameters are varied slowly enough, the adiabatic theorem guarantees that the system stays in the instantaneous ground state throughout, and the problem gets solved. In a quantum mechanical system, this is sometimes referred to as 'quantum annealing'. Several proposals and demonstrations, including the well-known D-Wave machine [115], utilize this algorithm.

The slow rate of variation of the Hamiltonian parameters is determined by the minimum energy spacing between the instantaneous ground state and first excited state that occurs as we move from the initial Hamiltonian to the final one. The smaller the gap is, the slower the rate at which we need to perform the variation to successfully solve the problem. It has been shown that the gap can become exponentially small in the problem size for **NP**-hard problems in the worst case, implying that this algorithm takes exponential time in the worst case.

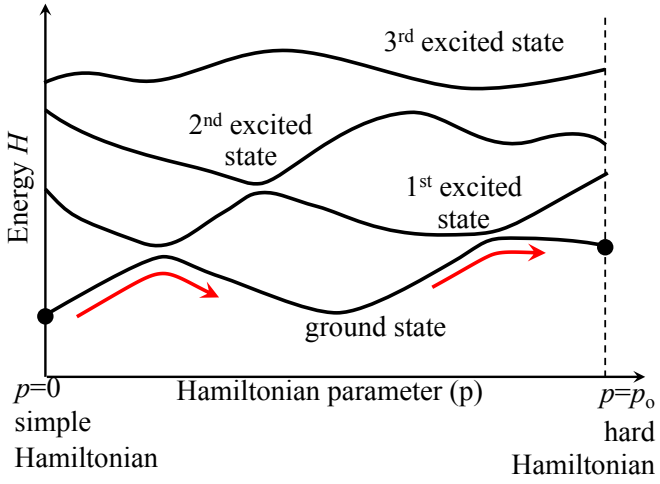

Figure 4.4: A system initialized in the ground state of a simple Hamiltonian continues to stay in the ground state as long as the Hamiltonian is changed slowly enough.

Laser physics

A laser is an optoelectronic device where light bounces around in an optical cavity that is composed of a gain medium. The gain medium, upon being pumped with external energy,

becomes an amplifier of light, instead of absorbing it like most materials. The spontaneously emitted light from the laser gain medium couples into the various cavity modes which have different loss coefficients due to their differing spatial profiles. The interplay between the gain provided to a mode by the gain medium, and the losses experienced by the mode, leads to each mode settling down to a steady-state amplitude. As intuition suggests, the optical mode with the least loss has the highest steady-state intensity for a given gain.

The idea introduced by [92] and [93] was to essentially use these concepts, but applied to networks of lasers, to solve optimization problems. Each candidate solution of the optimization problem is mapped onto a joint oscillatory mode of the laser network. The connectivity of the lasers is designed such that the objective function value at the candidate solution gets mapped to the loss coefficients of the corresponding joint network oscillatory mode. Then, when the entire laser network is pumped with gain, the joint network oscillatory mode with the least loss (that is, least objective function value) dominates and reaches the highest steady-state amplitude. By raising the gain value from 0 to a value that is very close to the value of the least-loss, one can ensure that only the least-loss mode has almost all the intensity. One can then read out the solution to the optimization problem from the steady-state joint network oscillatory mode.

We consider only classical coupled laser networks in this thesis. It is not yet clear how one would employ quantum coupled laser networks in this context or whether they offer any benefits at all. Since the N-spin Ising problem has 2^N candidate solutions, and a classical coupled laser network composed of N lasers supports only N joint oscillatory modes, it is immediately clear that the solution mechanism from the previous paragraph does not directly carry over to the Ising problem. We discuss the application of a version of this idea to the Ising problem in detail in the next chapter and illustrate its connection to the principle of minimum power dissipation.

4.2 The Ising problem

The Ising problem is a famously difficult problem in physics that asks for the ground state of a set of interacting magnets or spins. More precisely, let us say we have N electronic spins that can either point in the $+z$ or $-z$ direction as shown in Fig. 4.5.

Since this is an entirely classical problem, the electrons are not allowed to be in superpositions. The orientation of the i-th spin is represented by a ± 1 binary variable x_i which is $+1$ if the spin is pointed along $+z$ and -1 otherwise. The interaction energy between the i-th and j-th spins is specified by the quantity J_{ij} and their respective orientations x_i and x_j. Further, the i-th spin could also interact with a local magnetic field specified by h_i. The total interaction energy of the N spins is the sum of all the pairwise interaction energies and the interaction energies with the local magnetic fields and is given by:

$$H = -\sum_{i=1}^{N} h_i x_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij} x_i x_j. \tag{4.1}$$

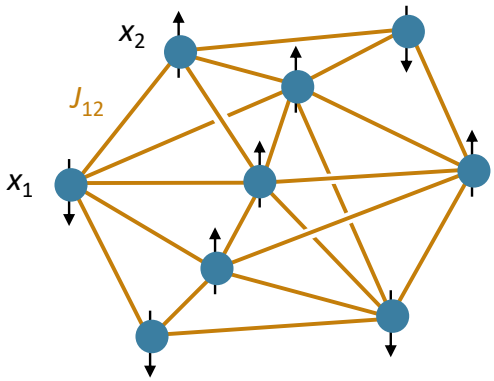

Figure 4.5: The Ising problem setup. A collection of N spins, which can each individually only point up ($x_i = 1$ if the i-th spin is up) or down ($x_i = -1$), interact via pairwise interaction energies $-J_{ij}x_ix_j$. The problem is to find the vector of orientations of the N spins, $\boldsymbol{x}$, that minimizes the sum of all the pairwise energies.

This expression is called the Ising Hamiltonian. We will assume that $J_{ii} = 0$ for all i. The Ising problem asks us to find the assignment of binary spin directions x_i^* that minimizes the Ising Hamiltonian, or equivalently, to find $\boldsymbol{x}^*$:

$$\boldsymbol{x}^* := \underset{\boldsymbol{x}:\ x_i=\pm 1\ \forall\ i}{\arg\max} \left(\sum_{i=1}^{N} h_i x_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} x_i x_j \right) = \underset{\boldsymbol{x}:\ x_i=\pm 1\ \forall\ i}{\arg\max} \left(\boldsymbol{h}^{\boldsymbol{T}} \boldsymbol{x} + \frac{1}{2} \boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{J} \boldsymbol{x} \right). \quad (4.2)$$

The second line is written in matrix notation. The Ising problem is therefore a discrete constrained optimization problem. It is **NP**-hard [5], and consequently, there is no known polynomial-time algorithm that solves it exactly.

Chapter 5

Coupled LC Oscillator Ising Solver and Lagrange Multipliers

The fact that the least loss-mode in a laser builds up the most intensity for a given gain was first used by Yamamoto et al. in [92] to construct coupled laser pulse solvers for the Ising problem. Each laser pulse is rendered bistable through the use of nonlinear physics. The number of laser pulses is equal to the number of spins in the specified Ising problem. The up and down spins of the Ising problem are mapped to the two bistable oscillatory states of each pulse. We note that using phases to represent digital bits is not an entirely new idea—it was first proposed by von Neumann [116] and Goto [117]. Next, the interactions between the laser pulses are designed so as to mimic the dynamics of the Ising model in a certain way.

Further work in various platforms was done by other groups [97, 99–101, 103], including a prominent machine in the electrical oscillator domain designed by Wang et al. [98]. A simple coupled LC electrical oscillator system that incorporated elements of both Yamamoto et al.'s optical system and Wang et al.'s electrical system was designed by Xiao [96]. We study Xiao's machine in detail in this chapter because it brings out the essential features of all the other machines.

5.1 Coupled LC oscillator Ising solver

Essence of this section: In this section, we describe (1) the implementation of spins of the Ising problem using bistable LC oscillators, and (2) the implementation of the Ising couplings J_{ij} using resistive connections between the LC oscillators. Further, we derive the equations of motion of the circuit voltages and simplify them using the slowly-varying amplitude approximation.

The coupled LC oscillator solver, as its name suggests, is composed of several nonlinear LC oscillators that are resistively coupled to one another. The number of oscillators in the

network is equal to the number of spins N in the specified Ising problem.

Parametric amplification

A simple LC cavity with linear components supports sinusoidal oscillations of arbitrary amplitude and phase. The voltage across the capacitor $V_c(t)$ can be of the form $A\cos(\omega_0 t + \phi)$ for any arbitrary A and ϕ. Here, ω_0 is the natural frequency of the LC cavity, $\omega_0 = 1/\sqrt{LC_0}$. When one incorporates nonlinear components in the LC cavity, it turns out that one can force the circuit to choose a specific set of phases. What we mean by nonlinear components is very simple—the charge and voltage of a nonlinear capacitor are related to one another by a nonlinear equation. For instance, a capacitor is said to have a second-order nonlinearity if its charge is proportional to the square of its voltage. The characteristic equation of such a capacitor is:

$$Q = C_0 V_c + C^{(2)} V_c^2 \tag{5.1}$$

where C_0 is the linear capacitance and $C^{(2)}$ is the second-order capacitance that connects the squared voltage to the charge.

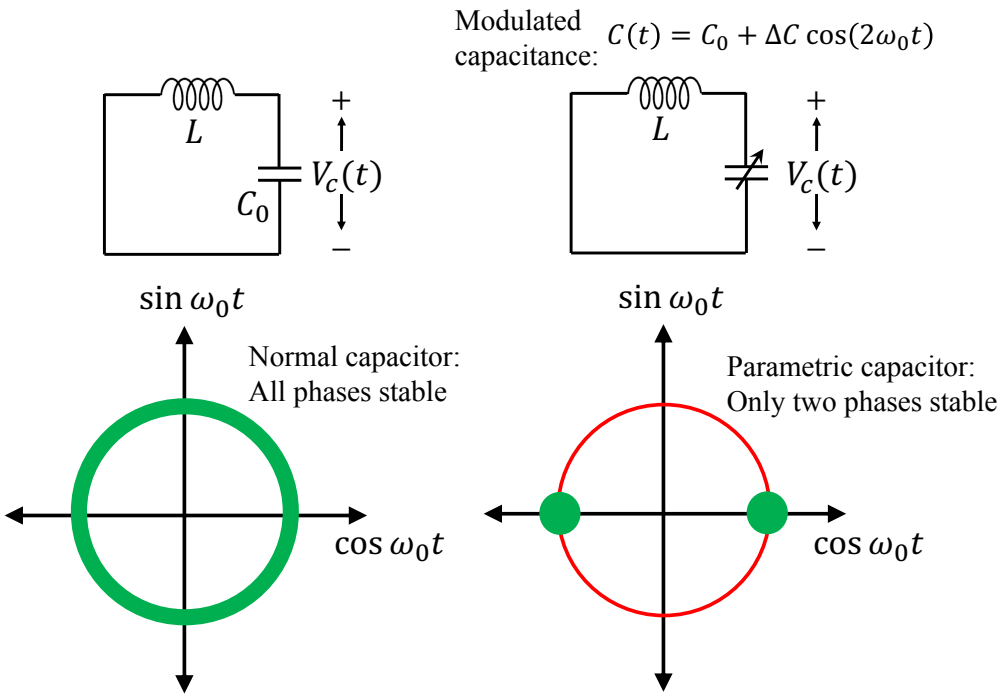

Figure 5.1: Parametric pumping induces bistability in phase in LC oscillators. A normal LC oscillator, shown on the left, can support oscillations of any phase. A parametric LC oscillator, shown on the right, can support oscillations only at two phases, ϕ_0 and $\phi_0 + \pi$, for some ϕ_0.

We now describe the concept of parametric bistability in a nutshell. The same material is covered more elaborately as we progress. If the LC cavity has a capacitor with a second-order nonlinearity, it is possible to force the phase ϕ to take on only two specific values, ϕ_0 and $\phi_0 + \pi$, through a process called parametric amplification. Let us say the voltage of the nonlinear capacitor takes on the form $V_c(t) = V_s + V_p$ where V_s is the 'signal' voltage, or the voltage we are interested in, and V_p is the so-called 'pump' voltage. V_s oscillates at the natural frequency ω_0 while V_p oscillates at $2\omega_0$. The pump voltage is so called because it provides gain to the signal voltage through their nonlinear interaction. If the phase of the signal voltage is in two particular alignments with respect to the pump's phase, the signal is able to receive energy gain from the pump. In other phase configurations, the signal loses its energy to the pump. That is, only two signal phases are 'encouraged' to exist in the circuit in steady-state by the pump. All other phases decay to zero in steady-state by losing their energy to the pump. The signal voltage, therefore, experiences phase bistability in the presence of a second-harmonic pump and a second-order nonlinearity. We will be mapping pulses of specific fixed amplitudes along these phases to the downward and upward pointing spins in the Ising problem. In our specific study, the two spins of the Ising problem are mapped to cosine oscillations of amplitude +1 in the 0 phase or amplitude +1 in the π phase. Phase bistability by itself is not enough; we also need the voltages to settle down to fixed amplitudes along those phases. For this reason, we also call the LC oscillator system 'coupled oscillators on the real axis'. This point will be elaborated on when we discuss the equations of motion of the system.

Before moving on, we point out that a second-order nonlinear capacitor with a pump voltage at $2\omega_0$ can be *simulated by a linear capacitor* whose capacitance value is modulated at frequency $2\omega_0$. To see this, we expand Eq. (5.1) with $V_c = V_s + V_p$:

$$Q = C_0 \left(V_s + V_p\right) + C^{(2)} \left(V_s^2 + 2V_sV_p + V_p^2\right). \tag{5.2}$$

If we are only interested in capturing the behavior of the capacitor accurately at the frequency ω_0, we are allowed to drop the second term in the first parenthesis and the first and third terms in the second parenthesis since they do not contribute to ω_0. We then have:

$$Q = \left(C_0 + 2C^{(2)}V_p\right) V_s = \left(C_0 + \Delta C(t)\right) V_s \tag{5.3}$$

where we introduced the capacitance modulation $\Delta C(t) = 2C^{(2)}V_p$. Since V_p oscillates at $2\omega_0$, the modulation $\Delta C(t)$ oscillates at $2\omega_0$ too.

Coupling

We have seen that each spin in the Ising problem will be represented by a phase-bistable nonlinear LC oscillator. We now specify the coupling scheme for the oscillators that implements the spin-spin interactions J_{ij}. This scheme is well-known and was used earlier in Wang et al.'s Ising solver [98].

The scheme is very simple to state for the case where all the J_{ij} in the specified Ising problem take values ± 1. $J_{ij} = 1$ means that the i-th and the j-th spins have a force that pushing them to align their spins with each other. Since spin is oscillator phase in our circuit, we want to couple the i-th and j-th oscillators in a fashion that induces them to oscillate together in the same phase. This can be achieved by linking their terminals up via a 'straight-linking' pair of resistors as shown in Fig. 5.2. If the i-th and j-th oscillators were oscillating in opposite phases, the ends of the resistors would experience large voltage drops, causing large currents to flow from one oscillator to the other and potentially resulting in phase flips. If they were oscillating in the same phase, little current would flow and the status quo would be maintained. Using similar reasoning, one can see that we would need to use a 'cross-linking' pair of resistors to represent $J_{ij} = -1$.

With this, we have completed the specification of the coupled LC oscillator Ising solver. Bistable oscillators represent the spins, and 'straight-linking' or 'cross-linking' resistive connections between them represent the J_{ij} in the specified Ising problem. However, what exactly does the circuit implement? This is answered only by looking at its equations of motion which were derived by Xiao [96]. The derivation of the circuit equations and the slowly-varying amplitude equations presented in the next subsection and in Appendix D is a simplified and slightly more general version of the original derivation by Xiao [96].

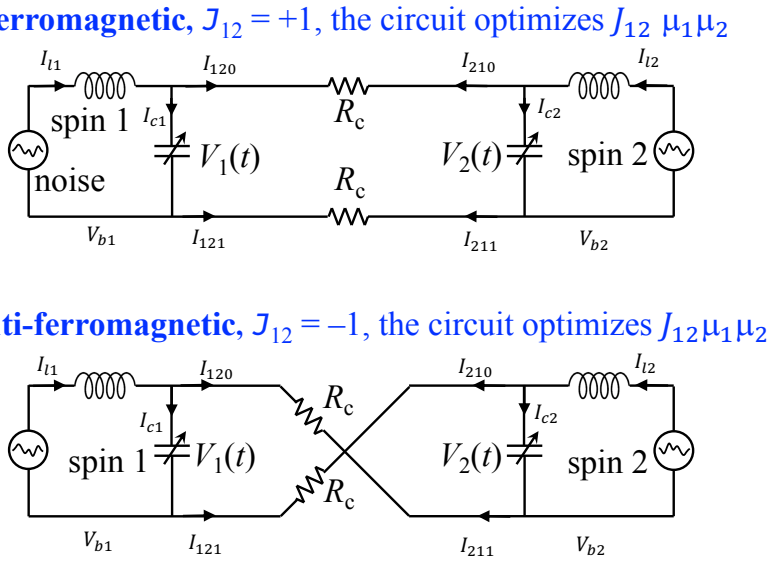

Figure 5.2: Coupled LC oscillator circuit for two coupled magnets. The oscillation of the LC oscillators represents the magnetic moments, while the parallel or antiparallel cross-connections represent ferromagnetic $J_{ij} = 1$ or antiferromagnetic $J_{ij} = -1$ coupling, respectively. The nonlinear capacitors are pumped by $V(2\omega_0)$ at frequency $2\omega_0$, providing parametric gain at ω_0.

Equations of motion

Deriving the equations of motion involves the application of the Kirchhoff voltage law (KVL) and current law (KCL) to the entire circuit. This requires a little care as the circuit is slightly complicated. The notation we will use is indicated in Fig. 5.2. One of the terminals of the capacitor in the oscillator labelled $i = 1$ is arbitrarily chosen as its 'bottom' terminal, and its other terminal is labelled its 'top' terminal. For each oscillator that is connected to $i = 1$ through a +1 connection, the terminal in that oscillator that is directly connected to the bottom terminal of $i = 1$ is labelled its 'bottom' terminal. Similarly, for each oscillator that is connected to $i = 1$ through a -1 connection, the terminal in that oscillator that is directly connected to the bottom terminal of $i = 1$ is labelled its *top terminal*. We continue this process recursively. If two terminals are connected by a + connection and one of them is the bottom terminal of its host oscillator, the other terminal is labelled the bottom terminal of its own host oscillator. If two terminals are connected by a - connection and one of them is the bottom terminal of its host oscillator, the other terminal is labelled the top terminal of its own host oscillator. Through this process, we can identify the bottom terminals of all the oscillators. The 'bottom' labelling is shown in Fig. 5.2 for the 2-spin case.

Now that this is settled, let the potential at the 'bottom' terminal of oscillator i be V_{bi}. The current that flows out from the bottom terminal of the i-th oscillator into the resistor that connects it to the j-th oscillator is i_{ij1}. Similarly, the current that flows out from the top terminal of the i-th oscillator into the resistor that connects it to the j-th oscillator is i_{ij0}. In the i-th oscillator, the voltage difference between the top and the bottom terminals of the capacitor is denoted by V_i, the current passing through the inductor towards the top terminal is i_{li}, and the current passing through the capacitor from the top to the bottom terminals is i_{ci}. All of this notation is again indicated on Fig. 5.2 for the 2-spin case.

KCL and KVL for the circuit can be written as:

$$i_{ci} = f_i(v_i, \dot{v_i}), \text{ for all } i \in \{1, 2, \ldots, N\} \tag{5.4}$$

$$v_i = -L\dot{i_{li}}, \text{ for all } i \in \{1, 2, \ldots, N\} \tag{5.5}$$

$$\sum_j i_{ij0} = i_{li} - i_{ci}, \text{ for all } i \in \{1, 2, \ldots, N\} \tag{5.6}$$

$$\sum_j i_{ij0} + \sum_j i_{ij1} = 0, \text{ for all } i \in \{1, 2, \ldots, N\} \tag{5.7}$$

$$v_{b1} = 0 \tag{5.8}$$

$$\begin{aligned} i_{ijk} = {} & k\left[\left(\frac{v_{bi} - v_{bj}}{R}\right)\left(\frac{1 + J_{ij}}{2}\right) + \left(\frac{v_{bi} - v_{bj} - v_j}{R}\right)\left(\frac{1 - J_{ij}}{2}\right)\right] \\ & + (1 - k)\left[\left(\frac{v_{bi} + v_i - v_{bj} - v_j}{R}\right)\left(\frac{1 + J_{ij}}{2}\right) + \left(\frac{v_{bi} + v_i - v_{bj}}{R}\right)\left(\frac{1 - J_{ij}}{2}\right)\right], \\ & \text{for all } i \in \{1, 2, \ldots, N\}, \text{ for all } j \in \{1, 2, \ldots, N\}, \ j \neq i, \text{ for all } k \in \{0, 1\} \end{aligned} \tag{5.9}$$

The first equation is the current characteristic of the i-th capacitor. Since the capacitor

is time-varying, we have chosen to simply represent its current by some function f_i of its voltage and voltage velocity. f_i will be specified later on. The second equation is simply the inductor's characteristic, the third one is the current law at the top terminal, the fourth one is a combination of the current laws at the top and the bottom terminals, the fifth one fixes the voltage reference by setting the potential of the bottom terminal of the first oscillator to 0, and the final equation writes out the characteristic of all the straight- and cross-linking resistors in the system.

Some simple algebra on the above equations, worked out in Appendix D, leads to the following vector equation for the capacitor voltages in all the oscillators:

$$(\boldsymbol{J} - (N-1)\,\boldsymbol{I})\,\dot{\boldsymbol{v}} = 2R_c\left(\dot{\boldsymbol{f}} + \frac{\boldsymbol{v}}{L}\right) \tag{5.10}$$

where $\dot{\boldsymbol{v}} = \frac{d\boldsymbol{v}}{dt}$, $\dot{\boldsymbol{f}} = \frac{d\boldsymbol{f}}{dt}$, $\boldsymbol{J}$ is the specified Ising matrix, $\boldsymbol{I}$ is the $N \times N$ identity matrix, R_c is the coupling resistance, and L is the value of the inductance inside each LC oscillator. All-to-all coupling is assumed in the derivation of this equation.

Slowly varying amplitude approximation

Next, we introduce the time-dependence of the capacitor through this following explicit form for $f_i(v_i, \dot{v}_i)$:

$$Q_i = (C_0 + \Delta C_i \cos(2\omega_0 t))\, v_i \tag{5.11}$$

$$\implies i_{ci} = (C_0 + \Delta C_i \cos(2\omega_0 t))\, \dot{v}_i - 2\omega_0 \Delta C_i \sin(2\omega_0 t) v_i = f_i(v_i, \dot{v}_i) \tag{5.12}$$

For now, let us assume that the capacitance modulation is the same for all the oscillators, that is, $\Delta C_i = \Delta C$ for all i (all the functions f_i are equal to the same function f). Rewriting Eq. (5.10) for this form of f, we get:

$$\begin{aligned}\left[\frac{\boldsymbol{J} - (N-1)\,\boldsymbol{I}}{2R_c C_0} + 4\omega_0 \frac{\Delta C}{C_0}\sin(2\omega_0 t)\right]\dot{\boldsymbol{v}} = &\left(1 + \frac{\Delta C}{C_0}\cos(2\omega_0 t)\right)\ddot{\boldsymbol{v}} \\ &+ \left(\frac{1}{LC_0} - 4\omega_0^2 \frac{\Delta C}{C_0}\cos(2\omega_0 t)\right)\boldsymbol{v}\end{aligned} \tag{5.13}$$

Our next step will be to apply the slowly-varying amplitude approximation to this equation. We are interested in voltage solutions to this equation that have frequencies near ω_0. For this purpose, let us decompose the capacitor voltage as $\boldsymbol{v}(t) = \boldsymbol{A}(t)\cos(\omega_0 t + \phi) + \boldsymbol{B}(t)\sin(\omega_0 t + \phi)$ where ϕ is some arbitrary phase and $\boldsymbol{A}(t)$ and $\boldsymbol{B}(t)$ are the amplitudes of the cosine and sine components respectively. We have the following for $\dot{\boldsymbol{v}}$ and $\ddot{\boldsymbol{v}}$:

$$\dot{\boldsymbol{v}} = \dot{\boldsymbol{A}}\cos(\omega_0 t + \phi) - \boldsymbol{A}\omega_0 \sin(\omega_0 t + \phi) + \dot{\boldsymbol{B}}\sin(\omega_0 t + \phi) + \boldsymbol{B}\omega_0 \cos(\omega_0 t + \phi) \tag{5.14}$$

$$\begin{aligned}\ddot{\boldsymbol{v}} = &\ddot{\boldsymbol{A}}\cos(\omega_0 t + \phi) - 2\dot{\boldsymbol{A}}\omega_0 \sin(\omega_0 t + \phi) - \boldsymbol{A}\omega_0^2 \cos(\omega_0 t + \phi) \\ &+ \ddot{\boldsymbol{B}}\sin(\omega_0 t + \phi) + 2\dot{\boldsymbol{B}}\omega_0 \cos(\omega_0 t + \phi) - \boldsymbol{B}\omega_0^2 \sin(\omega_0 t + \phi)\end{aligned} \tag{5.15}$$

The slowly-varying amplitude approximation is applicable when the amplitudes $\boldsymbol{A}(t)$ and $\boldsymbol{B}(t)$ vary at a rate that is much slower than the frequency ω_0, that is, when $\ddot{\boldsymbol{A}} \ll \dot{\boldsymbol{A}}\omega_0$ holds element-wise. We shall design our system to have this property, so the approximation becomes applicable. The approximation simply involves dropping $\ddot{\boldsymbol{A}}$ in the presence of $\dot{\boldsymbol{A}}\omega_0$. On the application of the approximation, the first and fourth terms on the right side of Eq. (5.15) get dropped. Plugging these expressions into Eq. (5.13), we get:

$$\begin{aligned}
&\left[\frac{\boldsymbol{J}-(N-1)\,\boldsymbol{I}}{2R_cC_0}+4\omega_0\frac{\Delta C}{C_0}\sin\left(2\omega_0 t\right)\right]\left(\left(\dot{\boldsymbol{A}}+\boldsymbol{B}\omega_0\right)\cos(\omega_0 t+\phi)+\left(\dot{\boldsymbol{B}}-\boldsymbol{A}\omega_0\right)\sin(\omega_0 t+\phi)\right)\\
&=\left(1+\frac{\Delta C}{C_0}\cos\left(2\omega_0 t\right)\right)\left(\left(-2\dot{\boldsymbol{A}}\omega_0-\boldsymbol{B}\omega_0^2\right)\sin(\omega_0 t+\phi)+\left(2\dot{\boldsymbol{B}}\omega_0-\boldsymbol{A}\omega_0^2\right)\cos(\omega_0 t+\phi)\right)\\
&\quad+\left(\frac{1}{LC_0}-4\omega_0^2\frac{\Delta C}{C_0}\cos\left(2\omega_0 t\right)\right)\left(\boldsymbol{A}\cos(\omega_0 t+\phi)+\boldsymbol{B}\sin(\omega_0 t+\phi)\right)
\end{aligned} \tag{5.16}$$

Since our quadrature basis, $\cos\left(\omega_0 t+\phi\right)$ and $\sin\left(\omega_0 t+\phi\right)$, was very general, we are allowed to choose any particular basis we are interested in by picking ϕ. It is the case that $\phi = 3\pi/4$ is the most convenient. In Appendix D, we multiply everything out, compare the coefficients of $\cos\left(\omega_0 t\right)$ and $\sin\left(\omega_0 t\right)$ on both sides, perform some more approximations, and set $\phi = 3\pi/4$ to get the following two equations:

$$\dot{\boldsymbol{A}}=\left(\frac{\boldsymbol{J}-(N-1)\,\boldsymbol{I}}{4R_cC_0}+\frac{\omega_0\Delta C}{4C_0}\right)\boldsymbol{A} \tag{5.17}$$

$$\dot{\boldsymbol{B}}=\left(\frac{\boldsymbol{J}-(N-1)\,\boldsymbol{I}}{4R_cC_0}-\frac{\omega_0\Delta C}{4C_0}\right)\boldsymbol{B} \tag{5.18}$$

Both the cosine and the sine components evolve exponentially from their starting conditions. Since all the eigenvalues of the loss matrix $\left(\boldsymbol{J}-(N-1)\,\boldsymbol{I}\right)/\left(4R_cC_0\right)$ are negative, the cosine component $\boldsymbol{A}$ grows if there is enough gain $g=\omega_0\Delta C/\left(4C_0\right)$ while the sine component decays always, even at zero gain. Since both $\boldsymbol{A}$ and $\boldsymbol{B}$ start out at the noise level in the circuit, and $\boldsymbol{A}$ quickly rises above the noise, we are justified in dropping the $\boldsymbol{B}$ equation altogether.

Our final evolution equation is:

$$\dot{\boldsymbol{A}}=\left(\frac{\boldsymbol{J}-(N-1)\,\boldsymbol{I}}{4R_cC_0}+\frac{\omega_0\Delta C}{4C_0}\right)\boldsymbol{A} \tag{5.19}$$

This is the evolution equation for the slowly-varying amplitude $\boldsymbol{A}(t)$ of the cosine component of the capacitor voltages in our coupled network. In this form, it looks like $\boldsymbol{A}(t)$ can grow indefinitely if the gain is large enough, but that is because we have approximated the behavior of the gain and totally ignored the dynamics of the pump circuit that drives each oscillator. In reality, the gain $\omega_0\Delta C/4C_0$ itself varies with time too. If the amplitude $\boldsymbol{A}(t)$ grows too much, the gain gets depleted and $\boldsymbol{A}(t)$ falls. We will return to this natural control mechanism

in a couple of sections. One final thing we note before moving to the next subsection is the fact the right-hand side of Eq. (5.19) can be written as the gradient of a quadratic form:

$$\left(\frac{\boldsymbol{J}-(N-1)\,\boldsymbol{I}}{4R_cC_0}+\frac{\omega_0\Delta C}{4C_0}\right)\boldsymbol{A}=\boldsymbol{\nabla}_{\boldsymbol{A}}\left[\frac{1}{2}\boldsymbol{A}^{\boldsymbol{T}}\left(\frac{\boldsymbol{J}-(N-1)\,\boldsymbol{I}}{4R_cC_0}+\frac{\omega_0\Delta C}{4C_0}\right)\boldsymbol{A}\right] \tag{5.20}$$

If we define a new function $h(\boldsymbol{A})$ as follows:

$$h(\boldsymbol{A})=\boldsymbol{A}^{\boldsymbol{T}}\left(\frac{(N-1)\,\boldsymbol{I}-\boldsymbol{J}}{4R_c}-\frac{\omega_0\Delta C}{4}\right)\boldsymbol{A}, \tag{5.21}$$

we can rewrite Eq. (5.19) as:

$$\dot{\boldsymbol{A}}=-\frac{1}{2C_0}\boldsymbol{\nabla}_{\boldsymbol{A}}h(\boldsymbol{A}) \tag{5.22}$$

Therefore, as time passes, Eq. (5.19) actually performs gradient descent on $h(\boldsymbol{A})$! We will now show that $h(\boldsymbol{A})$ is directly related to the **power dissipation of the circuit** in Fig. 5.2.

5.2 The circuit follows the principle of minimum power dissipation

Essence of this section: In this section, we show how the slowly-varying amplitude circuit equations perform gradient on a natural 'power dissipation function' that can be defined for the circuit. We show that this power dissipation function is identical to the one discussed by Onsager.

In this section, we will tie the slowly-varying amplitude dynamics of the circuit derived in the previous section to the power dissipation in the circuit. It will turn out that the dynamics Eq. (5.19) performs gradient descent on the power dissipation as time proceeds and finally ends up in a local minimum of the dissipation function in steady state.

Power is dissipated in the circuit in Fig. 5.2 only in the coupling resistors; the oscillators themselves have been assumed to be lossless and that shall continue to be the case. The power dissipated in each coupling resistor is simply the squared voltage difference across the resistor divided by the resistance value. We need access to the potentials of the top and bottom terminals of all the oscillators to be able to compute the voltage differences across all the coupling resistors. This can be done elegantly through a little linear algebra as shown in Appendix D. The formula for the dissipated power as a function of the voltages across the capacitors $\boldsymbol{v}(t)$ is derived there and is expressed as follows in matrix form:

$$P(t)=\boldsymbol{v}^{\boldsymbol{T}}(t)\left(\frac{(N-1)\,\boldsymbol{I}-\boldsymbol{J}}{2R_c}\right)\boldsymbol{v}(t) \tag{5.23}$$

Plugging in $\boldsymbol{v}(t) = \boldsymbol{A}(t)\cos\left(\omega_0 t + \frac{3\pi}{4}\right)$ and averaging over the cycle time period of $2\pi/\omega_0$ on both sides, we see that the time-averaged power can be expressed in terms of the slowly-varying amplitude $\boldsymbol{A}$ as follows:

$$\langle P(\boldsymbol{A})\rangle = \boldsymbol{A}^{\boldsymbol{T}}\left(\frac{(N-1)\,\boldsymbol{I}-\boldsymbol{J}}{4R_c}\right)\boldsymbol{A} \tag{5.24}$$

This expression is exactly the first term in $h(\boldsymbol{A})$! What about the second term? The term $\omega_0\Delta C/4C_0$ acts like a $1/RC_0$ time constant in Eq. (5.19), the only difference being that it gives energy to the signal rather than taking away energy as a normal resistor would. Therefore, $\omega_0\Delta C/4$ is a kind of **effective negative conductance** that the cosine component experiences due to the parametric pumping. Some care should be exercised here since Eq. (5.19) is a slowly-varying amplitude equation. Resistances appear in slowly-varying equations as $1/2RC$ instead of as $1/RC$. Therefore, the actual negative conductance is $\omega_0\Delta C/2$. The time-averaged power supplied by this negative resistor is:

$$\langle P_{\text{gain}}(t)\rangle = \left\langle \boldsymbol{v}^{\boldsymbol{T}}(t)\frac{\omega_0\Delta C}{2}\boldsymbol{v}(t)\right\rangle \implies \langle P(\boldsymbol{A})\rangle = \boldsymbol{A}^{\boldsymbol{T}}\frac{\omega_0\Delta C}{4}\boldsymbol{A} \tag{5.25}$$

We recover the second term of $h(\boldsymbol{A})$ too. Since the parametric terms acts as a negative resistance, the power 'dissipation' that happens in it is negative. The net power dissipation is therefore:

$$\langle P(\boldsymbol{A})\rangle = \boldsymbol{A}^{\boldsymbol{T}}\left(\frac{(N-1)\,\boldsymbol{I}-\boldsymbol{J}}{4R_c} - \frac{\omega_0\Delta C}{4}\right)\boldsymbol{A} \tag{5.26}$$

and we get the result $h(\boldsymbol{A}) = \langle P(\boldsymbol{A})\rangle$; the function $h(\boldsymbol{A})$ and the net power dissipation $\langle P(\boldsymbol{A})\rangle$ are exactly the same! Rewriting the slowly-varying amplitude equation, we get:

$$\dot{\boldsymbol{A}}(t) = -\frac{1}{2C_0}\boldsymbol{\nabla}_{\boldsymbol{A}}\langle P(\boldsymbol{A}(t))\rangle \tag{5.27}$$

The circuit performs gradient descent on the total instantaneous cycle-averaged power dissipation! This power dissipation includes both the positive dissipation in the coupling resistors and the negative 'dissipation' through the parametric capacitors. Eventually, the capacitor voltage amplitudes will settle down to a local minimum of the power dissipation in steady state.

Connection to Onsager's power dissipation

The fact that the circuit operates by the principle of minimum power dissipation is not surprising due to the linear nature of the signal circuit at the frequency ω_0. At our level of approximation so far, the time-varying capacitor, which provides the gain, only acts as a constant phase-dependent linear negative resistance. In 1930, Onsager noted in his influential reciprocity relations papers [4] that the equations of linear response could be recast as a 'minimum power dissipation' principle. Consider a general thermodynamic system with

thermodynamic forces given by the vector $\boldsymbol{V}$ and the fluxes given by the vector $\boldsymbol{J}$. In the linear response regime, they are connected by a symmetric matrix $\boldsymbol{R}$:

$$\boldsymbol{V} = \boldsymbol{R}\boldsymbol{J} \tag{5.28}$$

The fact that $\boldsymbol{R}$ is symmetric when time-reversal symmetry holds, the famous Onsager reciprocity relation, was the main subject of his papers. However, he did note that the linear response equation could equivalently be expressed in a variational form as:

$$\boldsymbol{V} = \boldsymbol{R}\boldsymbol{J} \iff \boldsymbol{\nabla}_{\boldsymbol{J}} \left(\frac{1}{2} \boldsymbol{J}^{\boldsymbol{T}} \boldsymbol{R} \boldsymbol{J} - \boldsymbol{V}^{\boldsymbol{T}} \boldsymbol{J} \right) = \boldsymbol{\nabla}_{\boldsymbol{J}} P(\boldsymbol{J}) = 0 \tag{5.29}$$

where we defined $P(\boldsymbol{J}) = \frac{1}{2} \boldsymbol{J}^{\boldsymbol{T}} \boldsymbol{R} \boldsymbol{J} - \boldsymbol{V}^{\boldsymbol{T}} \boldsymbol{J}$, the 'net power dissipation' function. For a given driving force $\boldsymbol{V}$, the currents $\boldsymbol{J}$ will adjust themselves such that, at steady-state, they render the net dissipation $P(\boldsymbol{J})$ stationary. This is the essence of Eq. (5.26) too. We had a positive dissipation term and a negative dissipation driving term, and, for fixed driving/gain ($\omega_0 \Delta C/2$) the equations move toward a local minimum of the net dissipation function. This variational formulation of the linear response equations makes it clear why it is critical to take into account 'negative dissipation' or the driving in the net dissipation function. Without that, one simply ends up with zero currents everywhere in steady-state. A driving force or gain or negative dissipation is necessary for non-zero currents and voltages to be maintained in any system.

Eq. (5.19) contains more information; it says that our particular coupled oscillator system approaches the stationary steady-state point via gradient descent of the dissipation function in time. This is an interesting curiosity of our particular system. It need not always be true; systems could also 'spiral' into the same local minimum via a 'spiralling gradient descent' in time.

We conclude this section by saying that our slowly-varying equations of motion and their implications are consistent with the minimum power dissipation principle as put forth by Onsager. We comment on the principle of minimum power dissipation for nonlinear systems in Appendix C.

5.3 Lagrange multipliers

Essence of this section: In this section, we point out a similarity between the power dissipation and the method of Lagrange multipliers. Next, we present a quick conceptual recap of Lagrange multipliers, and end with a discussion of the iterative 'method of multipliers' algorithm that helps us find local optima in constrained optimization problems.

The function $h(\boldsymbol{A})$ that is minimized by the slowly-varying amplitude dynamics has a nice physical interpretation in terms of power dissipation that we just discussed. It has an equally elegant *mathematical* interpretation—this will be the subject of the current section.

Since $h(\boldsymbol{A})$ is so important, we repeat it here for our present discussion:

$$h(\boldsymbol{A}) = \langle P(\boldsymbol{A}) \rangle = \boldsymbol{A}^{\boldsymbol{T}} \left(\frac{(N-1)\,\boldsymbol{I} - \boldsymbol{J}}{4R_c} - \frac{\omega_0 \Delta C}{4} \right) \boldsymbol{A} \tag{5.30}$$

Let us say we were to supply different gains to each of the oscillators. That is, let the capacitor modulation of the i-th capacitor be ΔC_i. The function $h(\boldsymbol{A})$ would then be:

$$h(\boldsymbol{A}) = \boldsymbol{A}^{\boldsymbol{T}} \left(\frac{(N-1)\,\boldsymbol{I} - \boldsymbol{J}}{4R_c} \right) \boldsymbol{A} - \sum_{i=1}^{N} \frac{\omega_0 \Delta C_i}{4} A_i^2 \tag{5.31}$$

The first term is essentially the Ising objective function $\boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{J} \boldsymbol{x}$ with some additional decoration. The second term, the sum, looks like a set of penalties that were added to the original objective function. In fact, they are the Ising constraints added to the objective function as penalty terms in exactly the same fashion as in the method of Lagrange multipliers! We now present a brief review of Lagrange multipliers and then show the explicit isomorphism between $h(\boldsymbol{A})$ and the Lagrange objective function. Much of the material that is presented here—the KKT conditions, the method of multipliers, and the Augmented Lagrange function—is drawn from the excellent optimization textbook by Bertsekas [118].

Review of Lagrange multipliers

The method of Lagrange multipliers is a very well-known procedure for solving constrained optimization problems. Let us say we wish to minimize a merit function of n variables, $f(\boldsymbol{x})$, within the set of all the $\boldsymbol{x}$ that satisfy the constraint $g(\boldsymbol{x}) = 0$. Let $\boldsymbol{x^*}$ be a local minimum of the merit function subject to the constraint. The local minimum in a general unconstrained problem has the property that the slope of the merit function is zero as infinitesimal steps are taken away from $\boldsymbol{x^*}$ in any direction. In constrained optimization, however, we only ask that the merit function remain unchanged for small deviations that respect the constraints. In our example, the infinitesimal steps are restricted to the constraint curve $g(\boldsymbol{x}) = 0$. The slope of $f(\boldsymbol{x})$ will be zero in the direction of small displacements along $g(\boldsymbol{x}) = 0$ if and only if the gradient of $f(\boldsymbol{x})$ is normal to $g(\boldsymbol{x}) = 0$. This means $\boldsymbol{\nabla} f(\boldsymbol{x^*})$ and $\boldsymbol{\nabla} g(\boldsymbol{x^*})$ should be parallel to each other:

$$\boldsymbol{\nabla} f(\boldsymbol{x^*}) = -\lambda^* \boldsymbol{\nabla} g(\boldsymbol{x^*}). \tag{5.32}$$

The proportionality constant λ^* is called the Lagrange multiplier corresponding to the constraint $g(\boldsymbol{x}) = 0$. A 2-dimensional example for maximization is shown in Fig. 5.3. The Lagrange method can be used both for minimization and maximization, and a gradient proportionality relation holds at all stationary points (local minima, local maxima, and saddle points). In this example, the isocontour lines of the function $f(\boldsymbol{x})$ increase until they are limited by, and just touch, the constraint curve $g(\boldsymbol{x}) = 0$ at the point $\boldsymbol{x^*}$.

When there are multiple constraints $g_1 = 0, \ldots, g_p = 0$, (5.32) is generalized as follows:

$$\boldsymbol{\nabla} f(\boldsymbol{x^*}) = -\sum_{i=1}^{p} \lambda_i^* \boldsymbol{\nabla} g_i(\boldsymbol{x^*}), \tag{5.33}$$

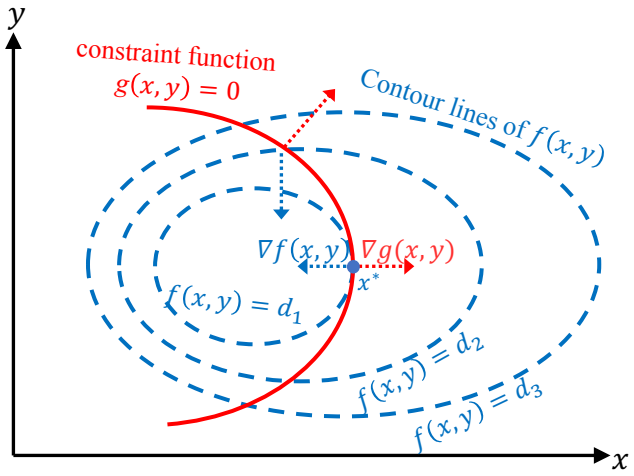

Figure 5.3: Maximization of function $f(x, y)$ subject to the constraint $g(x, y) = 0$. At the constrained local optimum, the gradients of f and g, namely $\nabla f(x, y)$ and $\nabla g(x, y)$, are parallel.

The gradient vector $\boldsymbol{\nabla}$ represents n equations, accompanied by the p constraint equations $g_i(\boldsymbol{x}) = 0$, resulting in $n + p$ equations. These equations solve for the n components in the vector $\boldsymbol{x}^*$, and the p unknown Lagrange Multipliers λ_i^*. That would be $n + p$ equations for $n + p$ unknowns. Every point $\boldsymbol{x}^*$ that locally minimizes or maximizes $f(\boldsymbol{x})$ subject to the constraints satisfies Eq. (5.33) for some $\boldsymbol{\lambda}^*$.

Eq. (5.33) suggests that we introduce a Lagrange function $L(\boldsymbol{x}, \boldsymbol{\lambda})$ defined as follows:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}), \tag{5.34}$$

This function has the property that the local minimum $\boldsymbol{x}^*$ and its associated multipliers $\boldsymbol{\lambda}^*$ satisfy:

$$\boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = 0, \;\; \boldsymbol{\nabla}_{\boldsymbol{\lambda}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = 0. \tag{5.35}$$

These are $n + p$ equations in all. These conditions are widely known as the Karush-Kuhn-Tucker (KKT) necessary and sufficient conditions for local stationarity. If a candidate point $(\boldsymbol{x'}, \boldsymbol{\lambda'})$ satisfies these conditions, then $\boldsymbol{x'}$ is a stationary point (local minimum, local maximum, or saddle point) of $f(\boldsymbol{x})$ subject to the constraints. As in normal calculus, one has to look at the second derivatives of $L(\boldsymbol{x}, \boldsymbol{\lambda})$ at $(\boldsymbol{x'}, \boldsymbol{\lambda'})$ to decide if the stationary point $\boldsymbol{x'}$ is specifically a local minimum , a local maximum, or a saddle point of $f(\boldsymbol{x})$ subject to the constraints.

Duality and the saddle point nature of $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$

Let us say we are searching for **constrained global minima** instead of **constrained local minima**. The problem we are trying to solve is:

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && g_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, p. \end{aligned}$$

Standard optimization textbooks show that this problem can be rewritten as:

$$\min_{\boldsymbol{x}:\ g_i(x)=0\ \forall\ i} f(\boldsymbol{x}) = \min_{\boldsymbol{x}} \left(\max_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right) \tag{5.36}$$

where $L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x})$ is the Lagrange function. We have converted a constrained optimization problem into an **unconstrained** nested min-max optimization problem. The well-known min-max inequality that is true for arbitrary functions tells us that:

$$\min_{\boldsymbol{x}} \left(\max_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right) \geq \max_{\boldsymbol{\lambda}} \left(\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right) \tag{5.37}$$

This relation holds for any optimization problem and is also called 'weak duality'. For some special optimization problems—which includes many common convex optimization problems—we actually have equality:

$$\min_{\boldsymbol{x}} \left(\max_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right) = \max_{\boldsymbol{\lambda}} \left(\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right) \tag{5.38}$$

The above relation says that the constrained global minimum $\boldsymbol{x}^*$ of $f(\boldsymbol{x})$ and its associated multiplier $\boldsymbol{\lambda}^*$ form a saddle point of $L(\boldsymbol{x}, \boldsymbol{\lambda})$. To see why they form a saddle point of $L(\boldsymbol{x}, \boldsymbol{\lambda})$, note that $(\boldsymbol{x}, \arg\max_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}))$ on the left-hand side represents a '1D' curved slice of the full space that passes through $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$. Moreover $L(\boldsymbol{x}, \boldsymbol{\lambda})$ is minimized over this slice at $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$. Therefore, moving away from $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ along the tangent to this slice increases L. Similarly, the right-hand side says that, over the '1D' curved slice represented by $(\arg\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}), \boldsymbol{\lambda})$, $L(\boldsymbol{x}, \boldsymbol{\lambda})$ is maximized at $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$. Therefore, moving away from $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ along the tangent to this slice decreases L.

Algorithm to find constrained global minimum $\boldsymbol{x}^*$ and its multipliers $\boldsymbol{\lambda}^*$

One of the standard ways to find the constrained global minimum $\boldsymbol{x}^*$ and its multipliers $\boldsymbol{\lambda}^*$ is to solve the nested max-min optimization problem on the right-hand side of Eq. (5.38) iteratively. For a fixed $\boldsymbol{\lambda}$, we minimize $L(\boldsymbol{x}, \boldsymbol{\lambda})$ in the $\boldsymbol{x}$ directions. Then, we take take one ascent step in the $\boldsymbol{\lambda}$ directions. We again solve the inner optimization problem for this slightly different $\boldsymbol{\lambda}$. The nested problem is solved iteratively through an inner loop of gradient descent in the $\boldsymbol{x}$ directions and an outer loop of gradient ascent in the $\boldsymbol{\lambda}$ directions.

More precisely, we choose an initial $\left(\boldsymbol{x}^{(\mathbf{0})}, \boldsymbol{\lambda}^{(\mathbf{0})}\right)$, run gradient descent on $L(\boldsymbol{x}, \boldsymbol{\lambda}^{(\mathbf{0})})$ in the $\boldsymbol{x}$ directions till we reach a local optimum $\left(\boldsymbol{x}^{(\mathbf{1})}, \boldsymbol{\lambda}^{(\mathbf{0})}\right)$, then take one gradient ascent step on $L(\boldsymbol{x}^{(\mathbf{1})}, \boldsymbol{\lambda})$ in the $\boldsymbol{\lambda}$ directions to reach $L(\boldsymbol{x}^{(\mathbf{1})}, \boldsymbol{\lambda}^{(\mathbf{1})})$, and repeat. The $i+1$-th iteration looks like the following:

1. Starting from $(\boldsymbol{x}^{(\boldsymbol{i})}, \boldsymbol{\lambda}^{(\boldsymbol{i})})$, perform **gradient descent** on $L(\boldsymbol{x}, \boldsymbol{\lambda}^{(\boldsymbol{i})})$ in the $\boldsymbol{x}$ directions till a local minimum is reached. Label that local minimum $(\boldsymbol{x}^{(\boldsymbol{i+1})}, \boldsymbol{\lambda}^{(\boldsymbol{i})})$.

2. Perform **one step of gradient ascent** on $L(\boldsymbol{x}^{(\boldsymbol{i+1})}, \boldsymbol{\lambda})$ in the $\boldsymbol{\lambda}$ directions to reach $(\boldsymbol{x}^{(\boldsymbol{i+1})}, \boldsymbol{\lambda}^{(\boldsymbol{i+1})})$.

This is the standard 'Method of Multipliers'. This algorithm outputs the **global optimal** $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ for a minimization problem which satisfies strong duality. Unfortunately, most difficult problems are highly non-convex and do not satisfy strong duality. For the time being, however, we will stick to this algorithm and see whether our coupled LC oscillator system implements some dynamics of this kind. A similar alternating procedure also exists to find **local minima** in problems with only weak duality—the 'Augmented Lagrangian Method of Multipliers'—and we shall take a look at it in the final section of this chapter.

5.4 Exact equivalence between coupled LC oscillators and Lagrange multipliers

Essence of this section: In this section, we show that the coupled LC oscillator circuit **exactly** implements the method of multipliers. The **net gain** of each oscillator plays the role of the **Lagrange multiplier** of the associated spin. The signal circuit equations perform gradient descent on $L(\boldsymbol{x}, \boldsymbol{\lambda})$ in the $\boldsymbol{x}$ directions whereas the pump circuit implements gradient ascent on $L(\boldsymbol{x}, \boldsymbol{\lambda})$ in the $\boldsymbol{\lambda}$ directions. To show the latter, a common pump circuit is presented and its equations of motion are derived.

The signal circuit performs gradient descent on the Lagrange function

In this subsection, we show that the dynamics of the coupled LC oscillator circuit performs step 1 in the 'Method of Multipliers'.

For an Ising problem with N spins, the merit function to be minimized is $-\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}x_i x_j$. There are $p = N$ constraints, one for each of the spins i, given by $g_i(\boldsymbol{x}) = 1 - x_i^2 = 0$. Therefore, the Lagrange function for the Ising problem, with these digital constraints, is given by:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = -\sum_{i=1}^{n}\sum_{j=1}^{n} J_{ij}x_i x_j + \sum_{i=1}^{n} \lambda_i \left(1 - x_i^2\right) \tag{5.39}$$

where λ_i is the Lagrange Multiplier associated with the constraint on the i-th spin. Step 1 in the method of multipliers performs gradient descent on $L(\boldsymbol{x}, \boldsymbol{\lambda})$ in $\boldsymbol{x}$. In iteration form, step 1 looks like:

$$\boldsymbol{x}[k+1] = \boldsymbol{x}[k] - \epsilon \boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{5.40}$$

where ϵ is a gradient descent step size that converts the units of the gradient to those of $\boldsymbol{x}$. This discrete time equation can be written in continuous time as below:

$$\dot{\boldsymbol{x}}(t) = -\kappa \boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = 2\kappa \left(\boldsymbol{J}\boldsymbol{x}(t) + \boldsymbol{\Lambda}\boldsymbol{x}(t) \right) \tag{5.41}$$

where κ is a gradient descent step size that converts the units of the gradient to those of $\dot{\boldsymbol{x}}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix that contains the λ_is on its diagonal.

Let us compare this with $h(\boldsymbol{A})$ repeated below:

$$h(\boldsymbol{A}) = -\frac{1}{4R_c} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} A_i A_j + \sum_{i=1}^{N} \left(\frac{\omega_0 \Delta C_i}{4} - \frac{N-1}{4R_c} \right) \left(-A_i^2 \right) \tag{5.42}$$

$h(\boldsymbol{A})$ is essentially the same as Eq. (5.39), but with a missing '1' adjacent to the $-A_i^2$ in the brackets at the end. If that is put in, we exactly retrieve the Lagrange function. The **net gain** of each oscillator **acts like the Lagrange multiplier** for the corresponding spin. For the purpose of taking gradients with respect to $\boldsymbol{A}$, the missing '1' in $h(\boldsymbol{A})$ doesn't matter. The Lagrange function for the circuit is:

$$L(A_1, \ldots, A_N, \Delta C_1, \ldots, \Delta C_N) = -\frac{1}{4R_c} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} A_i A_j + \sum_{i=1}^{N} \left(\frac{\omega_0 \Delta C_i}{4} - \frac{N-1}{4R_c} \right) \left(1 - A_i^2 \right) \tag{5.43}$$

It is clear from the form of the Lagrange function of the circuit that the two spins of the Ising problem are being mapped to cosine oscillations of amplitude +1 in the 0 phase or +1 in the π phase. Phase bistability by itself is not enough; we also need the voltages to settle down to fixed amplitudes along those phases. Since $h(\boldsymbol{A})$ and $L(A_1, \ldots, A_N, \Delta C_1, \ldots, \Delta C_N)$ both have the same partial derivatives with respect to A_i, and we already saw that the circuit was performing gradient descent on $h(\boldsymbol{A})$, we conclude that it equivalently performs gradient descent on $L(A_1, \ldots, A_N, \Delta C_1, \ldots, \Delta C_N)$. The circuit evolution equation is:

$$\dot{\boldsymbol{A}}(t) = -\frac{1}{2C_0} \boldsymbol{\nabla}_{\boldsymbol{A}} h(\boldsymbol{A}(t)) = \frac{1}{4R_c C_0} \left(\boldsymbol{J}\boldsymbol{A}(t) + \boldsymbol{\Gamma}\boldsymbol{A}(t) \right) \tag{5.44}$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix whose i-th diagonal element is $\Gamma_{ii} = \omega_0 \Delta C_i R_c - (N-1)$, a measure of the net gain received by the i-th oscillator. It can be verified that the right-hand side of the above dynamics is indeed the gradient of the aforementioned Lagrange function.

We have shown that the signal circuit's natural slowly-varying amplitude dynamics exactly implements **gradient descent on the Lagrange function in the $\boldsymbol{A}$ directions**.

The pump circuit performs gradient ascent on the Lagrange function

Step 2 in the method of multipliers performs gradient ascent on $L(\boldsymbol{x}, \boldsymbol{\lambda})$ in $\boldsymbol{\lambda}$. In iteration form, step 2 looks like:

$$\boldsymbol{\lambda}[k+1] = \boldsymbol{\lambda}[k] + \epsilon' \boldsymbol{\nabla}_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{5.45}$$

where ϵ' is a gradient descent step size that converts the units of the gradient to those of $\boldsymbol{\lambda}$. This discrete time equation can be written in continuous time as below:

$$\dot{\boldsymbol{\lambda}}(t) = \kappa' \boldsymbol{\nabla}_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = \kappa' \left(1 - \boldsymbol{x}^{\boldsymbol{2}}(t)\right) \tag{5.46}$$

where κ' is a gradient descent step size that converts the units of the gradient to those of $\dot{\boldsymbol{\lambda}}$, and $\boldsymbol{x}^{\boldsymbol{2}}(t)$ is a vector whose individual elements are the squares of the corresponding elements in the vector $\boldsymbol{x}(t)$. Does the coupled LC oscillator circuit implement something similar? The answer is **yes**. To see how this gradient ascent takes place in the circuit, we have to study the gain that is provided by the pump circuit. We already identified the gain of an oscillator with the Lagrange multiplier of the corresponding Ising spin. Therefore, it is natural that the time dynamics of $\boldsymbol{\lambda}$ the Lagrange multipliers is implemented by the pump circuit which provides the gain.

Pump circuit

In Fig. 5.4, we depict one of the standard parametric pump-signal coupling circuits [119]. Our analysis of this circuit will lead directly to the gradient ascent equations for the gains/Lagrange multipliers. The circuit equations are:

$$C_1 \dot{V}_s = I_3 - I_4, \qquad C_2 \dot{V}_p = I_2 - I_3, \tag{5.47}$$

$$V_s = L_1 \left(\dot{I}_4 - \dot{I}_5\right), \qquad V_p = L_2 \left(\dot{I}_1 - \dot{I}_2\right), \tag{5.48}$$

$$V_s = R_1 \left(I_s + I_5\right), \qquad V_p = R_2 \left(I_p - I_1\right) \tag{5.49}$$

$$I_3 = C_0 \left(\dot{V}_p - \dot{V}_s\right) + 2C_N \left(V_p - V_s\right) \left(\dot{V}_p - \dot{V}_s\right) \tag{5.50}$$

These equations are solved in Appendix D. As usual, we use slowly-varying amplitudes in that derivation by setting $V_s(t) = A_s(t) \cos\left(\omega_0 t + 3\pi/4\right)$ and $V_p(t) = A_p(t) \cos\left(2\omega_0 t\right)$. Note that we are already dropping the sine component of the signal voltage $V_s(t)$ because it will decay anyway. The result is:

$$\dot{A}_s = \frac{I_s}{2\left(C_0 + C_1\right)} - \frac{A_s}{2R_1\left(C_0 + C_1\right)} + \frac{C_N \omega_0 A_s A_p}{2\left(C_0 + C_1\right)} \tag{5.51}$$

$$\dot{A}_p = \frac{I_p}{2\left(C_0 + C_2\right)} - \frac{A_p}{2R_2\left(C_0 + C_2\right)} - \frac{C_N \omega_0 A_s^2}{2\left(C_0 + C_2\right)} \tag{5.52}$$

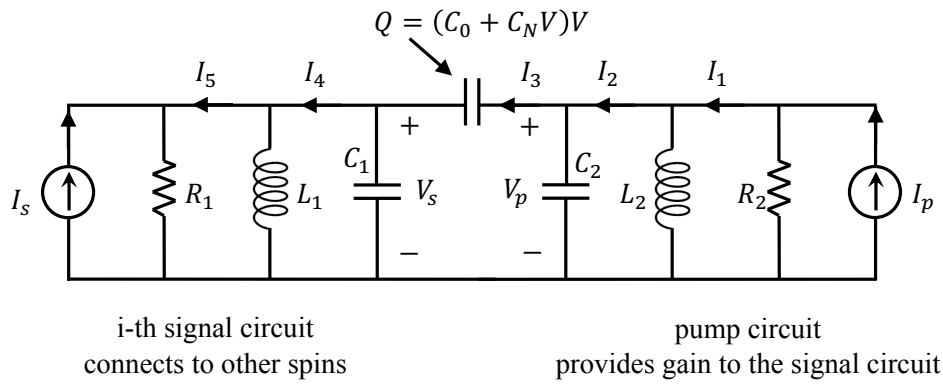

Figure 5.4: The i-th 'signal' oscillator, on the left, shown along with its pumping oscillator, on the right. The signal oscillator and the pump oscillator are coupled via a nonlinear capacitor, shown at the top. The terminals of the left signal capacitor C_1 are coupled to the other capacitors via the parallel and antiparallel coupling scheme.

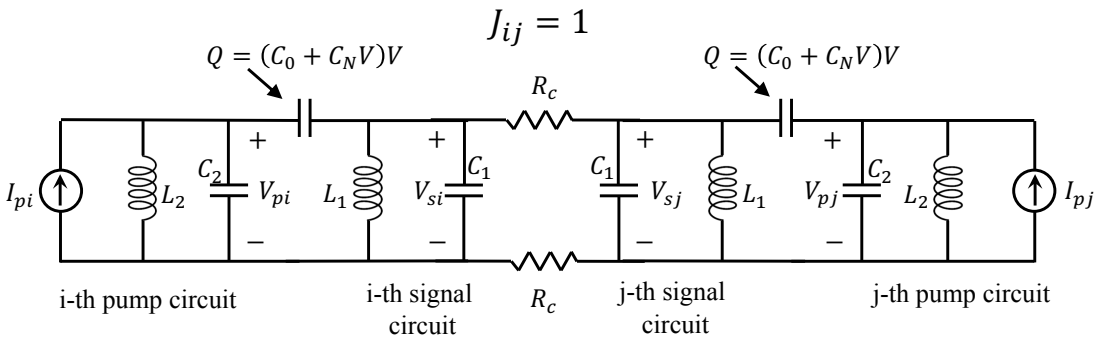

Figure 5.5: The complete circuit for two coupled magnets. The inner signal oscillators are exactly as shown in Fig. 5.2. The only difference is that the pumping circuits for the two oscillators are also depicted now.

where A_s and A_p are the slowly-varying amplitudes of voltage across the capacitors in the signal and pump circuits respectively. Both equations are very intuitive. The signal equation has an input term, I_s, the resistive decay term due to the internal resistance, and the final term is the gain which is directly proportional to V_p, the pump capacitor voltage. The pump equation has an input term, I_p, the resistive decay term due to the internal resistance, and the final term is the 'pump depletion' term. Pump depletion refers to the loss of the pump amplitude in response to an increase in the signal amplitude.

To make these equations look like the Lagrange multiplier equations, we set $I_s = 0$ and

$R_1 = R_2 = \infty$. The equations reduce to:

$$\dot{A}_s = \frac{C_N \omega_0 A_s A_p}{2\left(C_0 + C_1\right)} \tag{5.53}$$

$$\dot{A}_p = \frac{I_p}{2\left(C_0 + C_2\right)} - \frac{C_N \omega_0 A_s^2}{2\left(C_0 + C_2\right)} \tag{5.54}$$

These are the equations for a signal-pump pair that is not connected to anything else. In Fig. 5.5, we show that the signal part of the signal-pump pair has to be connected with other spins in the usual way using straight-linking and cross-linking resistors. The signal and pump voltage amplitudes in the i-th spin then satisfy:

$$\dot{A}_{si} = \left(\sum_{j=1}^{N} \frac{J_{ij} - (N-1)\,\delta_{ij}}{4R_c\left(C_0 + C_1\right)} A_{sj}\right) + \left(\frac{C_N \omega_0 A_{pi}}{2\left(C_0 + C_1\right)}\right) A_{si} \tag{5.55}$$

$$\dot{A}_{pi} = \frac{I_{pi}}{2\left(C_0 + C_2\right)} - \frac{C_N \omega_0 A_{si}^2}{2\left(C_0 + C_2\right)} \tag{5.56}$$

where δ_{ij} is the Kronecker delta. Equation (5.55) is only a rewriting of Eq. (5.19), but with a more accurate treatment of the final parametric gain term. There is nothing entirely new there. The pump equation is the new thing here, and we look at it a bit more closely. If I_{pi}, the pump driving current source, is chosen to be $I_{pi} = C_N \omega_0$ for all the spins i, then we can rewrite Eq. (5.56) as:

$$\dot{A}_{pi} = \frac{I_{pi}}{2\left(C_0 + C_2\right)}\left(1 - A_{si}^2\right) \tag{5.57}$$

This is exactly like the Lagrange multiplier gradient ascent equation. We summarize the exact equivalence below.

Exact equivalence between coupled LC oscillators and Lagrange multipliers

The Lagrange multiplier 'Method of Multipliers' is given by:

$$\dot{x}_i(t) = 2\kappa\left(\sum_{j=1}^{N} J_{ij} x_j(t) + \lambda_i(t) x_i(t)\right) \tag{5.58}$$

$$\dot{\lambda}_i(t) = \kappa'\left(1 - x_i^2(t)\right) \tag{5.59}$$

The coupled LC oscillator network is described by:

$$\dot{A}_{si}(t) = \left(\sum_{j=1}^{N} \frac{J_{ij} - (N-1)\,\delta_{ij}}{4R_c\left(C_0 + C_1\right)} A_{sj}(t)\right) + \left(\frac{C_N \omega_0 A_{pi}(t)}{2\left(C_0 + C_1\right)}\right) A_{si}(t) \tag{5.60}$$

$$\dot{A}_{pi}(t) = \frac{I_{pi}}{2\left(C_0 + C_2\right)}\left(1 - A_{si}^2(t)\right) \tag{5.61}$$

The equivalence is exact. The gains play the role of the Lagrange multipliers, the signal circuit performs the gradient descent in the original variable directions while the pump circuit performs gradient ascent in the Lagrange multiplier directions.

5.5 Augmented Lagrange function

Essence of this section: In this section, a drawback of the 'method of multipliers' is discussed, and a well-known way to fix it through the use of the 'Augmented Lagrange function' is presented. Finally, a circuit implementation of the 'Augmented Lagrange method' is provided and its equations of motion are derived.

The 'Method of Multipliers' finds constrained global minima in special optimization problems (those that satisfy strong duality; these problems tend to be convex). It doesn't work very well in finding constrained local minima of $f(\boldsymbol{x})$ in general optimization problems. The reason for this is explained in the next subsection.

Why the method of multipliers might fail in finding constrained local minima

Every constrained local minimum $\boldsymbol{x}^*$ satisfies the following KKT conditions for some $\boldsymbol{\lambda}^*$:

$$\boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = 0, \ \boldsymbol{\nabla}_{\boldsymbol{\lambda}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = 0. \tag{5.62}$$

However, these are only first-order conditions. If the objective function and the constraints are twice-differentiable (as is the case with the Ising objective function and its constraints), we can say more. Every constrained local minimum $\boldsymbol{x}^*$ satisfies the following *second-order* KKT conditions for the same $\boldsymbol{\lambda}^*$ as before:

$$\boldsymbol{\Delta x}^{\boldsymbol{T}} \boldsymbol{\nabla}^{\boldsymbol{2}}_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{\Delta x} \geq 0 \ \forall \text{ feasible displacements } \boldsymbol{\Delta x} \in \mathbb{R}^n \tag{5.63}$$

By 'feasible displacements', we mean displacements that are tangential to the constraint surfaces $g_1(\boldsymbol{x}) = 0, \ldots, g_p(\boldsymbol{x}) = 0$. Small tangential displacements do not move the point away from the constraint surfaces, which means $\boldsymbol{x}^* + \boldsymbol{\Delta x}$ still lies within the feasible region. The meaning of the second-order condition is that, for all feasible displacements away from $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ in the $\boldsymbol{x}$ directions, the Lagrange function $L(\boldsymbol{x}, \boldsymbol{\lambda})$ necessarily increases. In the 'infeasible' directions on the other hand, the Lagrange function could decrease. This decrease of the Lagrange function in the other directions around $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ poses problems for the method of multipliers when it is employed to find $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$.

The method of multipliers solves a nested max-min problem. For a fixed $\boldsymbol{\lambda}$, it minimizes $L(\boldsymbol{x}, \boldsymbol{\lambda})$ over $\boldsymbol{x}$ using gradient descent. Let us say we know the $\boldsymbol{\lambda}^*$ corresponding to a particular constrained local minimum $\boldsymbol{x}^*$. Let us also say that our initial point for the method of multipliers is $\left(\boldsymbol{x}^{(\boldsymbol{0})}, \boldsymbol{\lambda}^*\right)$ for some $\boldsymbol{x}^{(\boldsymbol{0})}$ that is already close to $\boldsymbol{x}^*$. The method

of multipliers then performs gradient descent in the $\boldsymbol{x}$ directions starting from $\left(\boldsymbol{x^{(0)}}, \boldsymbol{\lambda^*}\right)$. If $\boldsymbol{x^{(0)}}$ is separated from $\boldsymbol{x^*}$ by a feasible displacement (that is, if $\boldsymbol{\Delta x} = \boldsymbol{x^{(0)}} - \boldsymbol{x^*}$ is a feasible displacement), the gradient descent will work and we will reach $(\boldsymbol{x^*}, \boldsymbol{\lambda^*})$. For infeasible displacements, however, we know that Lagrange function can decrease as one moves *away* from $(\boldsymbol{x^*}, \boldsymbol{\lambda^*})$. If $\boldsymbol{x^{(0)}}$ is separated from $\boldsymbol{x^*}$ by an infeasible displacement, the gradient descent could take the algorithm in a direction that is opposite to where $(\boldsymbol{x^*}, \boldsymbol{\lambda^*})$ is situated. In summary, even if one is given the right $\boldsymbol{\lambda^*}$ *and* a starting $\boldsymbol{x^{(0)}}$ that is close to $\boldsymbol{x^*}$, the method of multipliers could fail spectacularly in finding $\boldsymbol{x^*}$.

The solution: Method of multipliers on the Augmented Lagrange function

The solution to this problem is to modify the Lagrange function $L(\boldsymbol{x}, \boldsymbol{\lambda})$ and get a new function, $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$, that has the property that *every displacement* $\boldsymbol{\Delta x}$ in the $\boldsymbol{x}$ directions away from $(\boldsymbol{x^*}, \boldsymbol{\lambda^*})$, feasible or otherwise, leads to an increase in $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$. The **Augmented Lagrange** function $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$ does the job [118]. The Augmented Lagrange function for a problem with objective function $f(\boldsymbol{x})$ and constraints $g_1(\boldsymbol{x}) = 0, \ldots, g_p(\boldsymbol{x}) = 0$ is given by:

$$L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}) + \frac{c}{2}\left(\sum_{i=1}^{p} \left(g_i(\boldsymbol{x})\right)^2\right) \tag{5.64}$$

$$= L(\boldsymbol{x}, \boldsymbol{\lambda}) + \frac{c}{2}\left(\sum_{i=1}^{p} \left(g_i(\boldsymbol{x})\right)^2\right) \tag{5.65}$$

It is simply the original Lagrange function but with the sum of squares of the constraints functions added on. It is shown in [118] that every constrained strict local minimum $\boldsymbol{x^*}$, along with its associated $\boldsymbol{\lambda^*}$, satisfies the following first-order 'Augmented Lagrange' KKT conditions:

$$\boldsymbol{\nabla_x} L_c(\boldsymbol{x^*}, \boldsymbol{\lambda^*}) = 0, \ \boldsymbol{\nabla_\lambda} L_c(\boldsymbol{x^*}, \boldsymbol{\lambda^*}) = 0. \tag{5.66}$$

$\boldsymbol{x^*}$ also satisfies the following second-order 'Augmented Lagrange' KKT conditions for the same $\boldsymbol{\lambda^*}$ as before, this time, for *all* displacements $\boldsymbol{\Delta x}$:

$$\boldsymbol{\Delta x^T \nabla^2_{xx}} L_c(\boldsymbol{x^*}, \boldsymbol{\lambda^*}) \boldsymbol{\Delta x} \geq 0 \ \forall \text{ displacements } \boldsymbol{\Delta x} \in \mathbb{R}^n \tag{5.67}$$

The method of multipliers will now succeed in finding constrained local minima if it is applied to $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$. Returning to the particular example we saw before, let us again say we know the $\boldsymbol{\lambda^*}$ corresponding to $\boldsymbol{x^*}$, and that our initial point for the method of multipliers is $\left(\boldsymbol{x^{(0)}}, \boldsymbol{\lambda^*}\right)$ for some $\boldsymbol{x^{(0)}}$ that is already close to $\boldsymbol{x^*}$. Since there are no directions $(\boldsymbol{\Delta x}, \boldsymbol{0})$ around $(\boldsymbol{x^*}, \boldsymbol{\lambda^*})$ in which the Augmented Lagrange function $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$ decreases, there is no danger of the algorithm slipping away from $(\boldsymbol{x^*}, \boldsymbol{\lambda^*})$ whatever the initial point $\left(\boldsymbol{x^{(0)}}, \boldsymbol{\lambda^*}\right)$ is.

The 'Augmented Lagrange' method of multipliers works as follows. Choose an initial point $\left(\boldsymbol{x^{(0)}}, \boldsymbol{\lambda^{(0)}}\right)$, run gradient descent on $L_c(\boldsymbol{x}, \boldsymbol{\lambda^{(0)}})$ in the $\boldsymbol{x}$ directions till we reach a local

optimum $\left(\boldsymbol{x}^{(\mathbf{1})}, \boldsymbol{\lambda}^{(\mathbf{0})}\right)$, then take one gradient ascent step on $L_c(\boldsymbol{x}^{(\mathbf{1})}, \boldsymbol{\lambda})$ in the $\boldsymbol{\lambda}$ directions to reach $L(\boldsymbol{x}^{(\mathbf{1})}, \boldsymbol{\lambda}^{(\mathbf{1})})$, and repeat. The $i+1$-th iteration looks like the following:

1. Starting from $(\boldsymbol{x}^{(\boldsymbol{i})}, \boldsymbol{\lambda}^{(\boldsymbol{i})})$, perform **gradient descent** on $L_c(\boldsymbol{x}, \boldsymbol{\lambda}^{(\boldsymbol{i})})$ in the $\boldsymbol{x}$ directions till a local minimum is reached. Label that local minimum $(\boldsymbol{x}^{(\boldsymbol{i+1})}, \boldsymbol{\lambda}^{(\boldsymbol{i})})$.
2. Perform **one step of gradient ascent** on $L_c(\boldsymbol{x}^{(\boldsymbol{i+1})}, \boldsymbol{\lambda})$ in the $\boldsymbol{\lambda}$ directions to reach $(\boldsymbol{x}^{(\boldsymbol{i+1})}, \boldsymbol{\lambda}^{(\boldsymbol{i+1})})$.

More details of this method, including the proof that it works, may be found in the textbook by Bertsekas [118].

Augmented Lagrange function for the Ising problem

The Augmented Lagrange function for the Ising problem is:

$$L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = -\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}x_i x_j + \sum_{i=1}^{N} \lambda_i \left(1 - x_i^2\right) + \frac{c}{2}\left(\sum_{i=1}^{N}\left(x_i^2 - 1\right)^2\right) \tag{5.68}$$

Then, the Augmented Lagrange method of multipliers can be expressed as:

$$\dot{x}_i(t) = 2\kappa\left(\sum_{j=1}^{N} J_{ij}x_j(t) + \lambda_i(t)x_i(t) + cx_i(t) - cx_i^3(t)\right) \tag{5.69}$$

$$\dot{\lambda}_i(t) = \kappa'\left(1 - x_i^2(t)\right) \tag{5.70}$$

In the next subsection, we show how the coupled LC oscillator network can be modified to implement the Augmented Lagrange method of multipliers.

Augmented Lagrange coupled LC oscillator Ising solver

Since the $\boldsymbol{\lambda}$ equation is the same in both the methods of multipliers, we need not make any changes to the pump circuit. The appearance of a cubic decay term in the $\boldsymbol{x}$ equation suggests that we should add a nonlinear resistor with a cubic nonlinearity to each oscillator. Let the cubic resistor be described by:

$$I = VG_0 + V^3G_N = VG_{\text{nlin}}(V) \tag{5.71}$$

The nonlinear resistor R_{nlin} is added in parallel to the parametric capacitor in the signal circuit as shown in Fig. 5.6. The equation of motion that was derived in Sec. 1 of this chapter for the original circuit was:

$$\left(\boldsymbol{J} - (N-1)\,\boldsymbol{I}\right)\dot{\boldsymbol{v}} = 2R_c\left(\dot{\boldsymbol{f}} + \frac{\boldsymbol{v}}{L}\right) \tag{5.72}$$

The current passing through the i-th capacitor was $i_{ci} = f_i(v_i, \dot{v}_i)$. We had then set f_i to:

$$i_{ci} = f_i(v_i, \dot{v}_i) = (C_0 + \Delta C_i \cos(2\omega_0 t))\, \dot{v}_i - 2\omega_0 \Delta C_i \sin(2\omega_0 t) v_i \tag{5.73}$$

To account for the nonlinear resistor in our new circuit, we simply add the current through the nonlinear resistor to f_i:

$$i_{ci} = f_i(v_i, \dot{v}_i) = (C_0 + \Delta C_i \cos(2\omega_0 t))\, \dot{v}_i - 2\omega_0 \Delta C_i \sin(2\omega_0 t) v_i + G_0 v_i + G_N v_i^3 \tag{5.74}$$

In Appendix D, we substitute this expression into the equation of motion and perform the usual slowly-varying amplitude approximation and retain only the cosine component of the capacitor voltage in each oscillator to obtain the following equation:

$$\dot{\boldsymbol{A}} = \left(\frac{\boldsymbol{J} - (N-1)\,\boldsymbol{I}}{4R_c C_0} - \frac{G_0}{2C_0} + \boldsymbol{g}(t) \right) \boldsymbol{A} - \frac{3G_N}{8C_0} \boldsymbol{A^3} \tag{5.75}$$

where $\boldsymbol{g}(t)$ is a diagonal matrix that contains the gain of the i-th oscillator, $g_i(t)$, as its i-th diagonal element for every i, and $\boldsymbol{A^3}$ is the vector whose i-th element is the cube of the i-th element of $\boldsymbol{A}$ for every i. It is clear that this equation of motion implements Eq. (5.69).

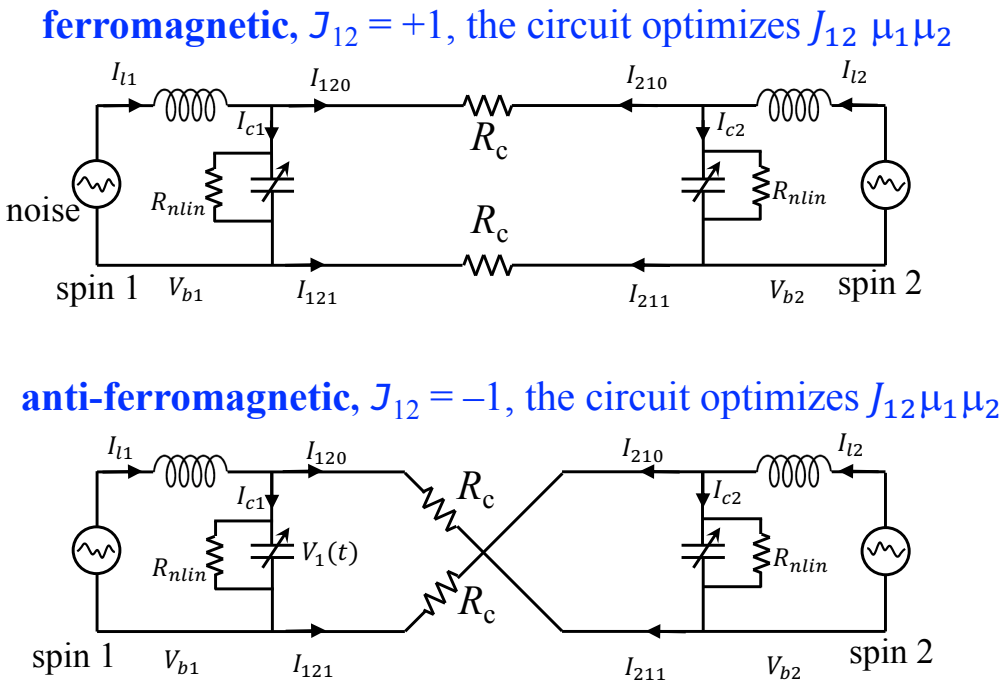

Figure 5.6: The coupled LC circuit that implements the Augmented Lagrange method for two spins. A nonlinear resistor R_{nlin} is placed in parallel with the parametric capacitor in each spin. The pumping circuit is only represented by an arrow over the capacitors for brevity.

It should be noted that we inserted cubic resistors in parallel with the parametric capacitors to implement our particular choice of the Augmented Lagrange function, taken from [118]. It is possible that there are other ways of constructing Augmented Lagrange functions in the optimization literature. One would simply use the appropriate nonlinear resistive element in parallel with the parametric capacitor to implement those functions. For instance, Xiao [96] used p-n diodes as the nonlinear resistive element in their circuit.

In the next chapter, we discuss a variety of coupled oscillator Ising solvers published in the literature in the past decade and show that they all work by the same mechanism as the coupled LC oscillator system.

5.6 Conclusion

The summary of this chapter is:

1. We presented a coupled nonlinear LC oscillator network circuit that functions as an Ising solver. The spins are mapped onto parametrically pumped oscillators and the Ising connections are implemented using parallel or antiparallel resistive connections between the oscillators.

2. The slowly-varying amplitude equations of motion of the capacitor voltages in the circuit perform gradient descent on the power dissipation function. This power dissipation is the same function discussed by Onsager in his work.

3. The network actually performs Lagrange multiplier optimization of the Ising problem. The net gain of each oscillator plays the role of the Lagrange multiplier of that spin.

4. The signal circuit performs gradient descent of the Lagrange function in the spin variables while the pump circuit performs gradient ascent of the Lagrange function in the gain/multiplier variables.

5. It is possible to implement more sophisticated procedures such as the 'Augmented Lagrange method of multipliers' by introducing additional nonlinear circuit elements.

Chapter 6

Other Physical Ising Solvers

In this chapter, we discuss some published coupled oscillator Ising solvers and show that they all perform Lagrange multiplier optimization and work by the same mechanism as the coupled LC oscillator system. We will also comment on a couple of published physics-based solvers which are less obviously connected to Lagrange multipliers. Finally, we will present a simple application of these solvers to perform linear regression in statistics.

6.1 Other Physical Ising Solvers

Essence of this section: In this section, we discuss a variety of physical Ising solvers in the literature and show that almost all of them perform Lagrange multiplier optimization. Moreover, there is an equivalence between the power dissipation and the Lagrange function in many of them.

We now discuss some physical methods proposed in the literature and show how each scheme implements the method of Lagrange multipliers. They all obtain good performance on the Gset benchmark problem set [120], and many of them demonstrate better performance than the heuristic algorithm, Breakout Local Search [121].

The available physical solvers in the literature, we entitle as follows: Optical Parametric Oscillators, Coupled laser cavities using multicore fibers, Coupled Radio Oscillators on the Unit Circle, Coupled polariton condensates.

Optical Parametric Oscillators

Overview

An early optical machine for solving the Ising problem was presented by Yamamoto et al. [92] and [122]. Their system consists of several pulses of light circulating in an optical fiber loop, with the a fixed amplitude in one phase representing an Ising spin and the same fixed amplitude in the opposite phase representing the other spin. The pulses are pumped

through parametric gain which serves a dual purpose. (1) When the gain overcomes the intrinsic losses of the fiber, the optical pulse builds up. (2) It restricts the oscillatory phase to the Real Axis of the complex plane, that is, only the cosine phases survive.

The Ising coupling J is achieved in this system by specified interactions between the optical pulses. In Yamamoto's approach, one pulse i is first plucked out by an optical gate, amplitude modulated by the proper connection weight J_{ij} specified in the Ising Hamiltonian, and then reinjected and superposed onto the other optical pulse j, producing constructive or destructive interference, representing ferromagnetic or anti-ferromagnetic coupling. This 'plucking out' and reinjection happens in every round-trip of the pulses around the fiber loop.

Our findings from the previous chapter hold here. The pulse amplitudes evolve according to the Principle of Minimum Power Dissipation. Each Lagrange Multiplier turns out to be equal to the gain or loss associated with the corresponding optical pulse.

Equations of motion

Yamamoto et al. [93] analyze their parametric oscillator system using slowly varying coupled wave equations for the circulating optical pulses. Owing to the nature of parametric amplification, the quadrature sine components s_i of the electric fields die out rapidly. The equation for the slowly varying amplitude c_i of the cosine component of the i-th optical pulse is as follows:

$$\frac{dc_i}{dt} = \frac{(-\alpha_i + \gamma_i)}{2} c_i + \sum_j J_{ij} c_j \tag{6.1}$$

where the weights, J_{ij}, are the magnetic cross-couplings, γ_i represents the parametric gain supplied to the i-th pulse, and α_i is the corresponding loss. For clarity of discussion, we dropped the cubic terms in (6.1) that Yamamoto et al. originally had. A discussion of these terms is given at the end of this subsection. A quick way to derive the above continuous-time equation from the iterative dynamics of this system is as follows. The amplitude of the i-th pulse at the beginning of the k-th iteration or round-trip is $c_i[k]$. In the course of the k-th iteration, this pulse loses amplitude to fiber losses α, gains amplitude due to the parametric gain γ, and is enhanced by the injection of modulated amplitude from the other pulses. If the length of the fiber loop is L, the amplitude of the i-th pulse at the beginning of the $k+1$-th iteration is:

$$c_i[k+1] = e^{-(\alpha-\gamma)L/2} c_i[k] + \beta \sum_j J_{ij} c_j[k] \tag{6.2}$$

where β is some small constant. If T is the round-trip time and $\alpha L \ll 1$ and $\gamma L \ll 1$, we have:

$$\frac{c_i[k+1] - c_i[k]}{T} = \frac{(-\alpha + \gamma)}{2} \frac{L}{T} c_i[k] + \frac{\beta}{T} \sum_j J_{ij} c_j[k] \tag{6.3}$$

which reproduces Eq. (6.1).

Power dissipation

The average power dissipation in the system is computed by dividing the average energy change of the system in one round-trip by the round-trip time T:

$$\langle P[k] \rangle = -\frac{\epsilon_0 V}{T} \sum_i \left(c_i^2[k+1] - c_i^2[k] \right) \tag{6.4}$$

$$= \epsilon_0 V \left(-2\frac{\beta}{T} \sum_{i,j} J_{ij} c_i[k] c_j[k] + \sum_i \alpha_i \frac{L}{T} c_i^2[k] - \sum_i \gamma_i \frac{L}{T} c_i^2[k] \right) \tag{6.5}$$

where V is the volume of each optical pulse and ϵ_0 is the permittivity of free space. Clearly, the partial derivative of the power dissipation with respect to the amplitude $c_i[k]$ is proportional to the right-hand side of Eq. (6.1). Therefore, the dynamics of the system perform gradient descent on the power dissipation function. Each mode, represented by the subscripts in c_i, must adjust to a particular gain γ_i and the other amplitudes such that the total power dissipation is minimized.

Lagrange function

The Lagrange function on which gradient descent is being performed is:

$$L(\boldsymbol{c}, \boldsymbol{\gamma}) = -\sum_{i,j} J_{ij} c_i c_j + \sum_i \left(\frac{\gamma_i - \alpha_i}{2} \right) \left(1 - c_i^2 \right) \tag{6.6}$$

The power dissipation function (6.5) and the Lagrange function (6.6) differ only in the (+1) in the final term. (6.6) is also identical in form to the Lagrange function for the LC oscillator solver. As before, the net gains play the role of Lagrange Multipliers. Minimization of the Lagrange function (6.6) provides the final steady state of the system dynamics. In fact, the right-hand side of (6.1) is the gradient of (6.6), demonstrating that the dynamical system performs gradient descent on the Lagrange function.

Cubic terms

The full equation derived in [93] is of the form:

$$\frac{dc_i}{dt} = \frac{(-\alpha_i + \gamma_i)}{2} c_i - \zeta c_i^3 + \sum_j J_{ij} c_j \tag{6.7}$$

for some nonlinear coefficient ζ. This equation looks like it was obtained through the use of a nonlinear resistor in order to implement the Augmented Lagrange method. However, [123] explains that it was obtained from the plain signal-pump system but with very fast pump

depletion. Recalling the signal-pump equations from the previous chapter:

$$\dot{A}_s = \frac{I_s}{2\left(C_0 + C_1\right)} - \frac{A_s}{2R_1\left(C_0 + C_1\right)} + \frac{C_N \omega_0 A_s A_p}{2\left(C_0 + C_1\right)} \tag{6.8}$$

$$\dot{A}_p = \frac{I_p}{2\left(C_0 + C_2\right)} - \frac{A_p}{2R_2\left(C_0 + C_2\right)} - \frac{C_N \omega_0 A_s^2}{2\left(C_0 + C_2\right)} \tag{6.9}$$

we see that if the pump dynamics is very fast, the latter rate will almost always be close to 0, giving $A_p \propto I_p - C_N \omega_0 A_s^2$. When this quadratic expression for A_p is substituted back into the signal equation, we get a A_s^3 term. Therefore, a cubic 'Augmented Lagrange'-like saturation is obtained at the price of losing the critical pump evolution equation which performed ascent on the Lagrange function in the λ directions. In our opinion, a more complete approach would be to include a nonlinear resistor to implement the Augmented Lagrange cubic term, and retain the important Lagrange function gradient ascent performed by the pump.

Coupled laser cavities using multicore fibers

Overview

The Ising solver designed by Babaeian et al., [97], makes use of coupled laser modes in a multicore optical fiber. Polarized light in each core of the optical fiber corresponds to a spin in the Ising problem. The number of cores is equal to the number of magnets in the given Ising instance. The right-hand and left-hand circular polarization of the laser light in each core represent the two polarities (up and down) of the corresponding magnet. The mutual coherence of the various cores is maintained by injecting seed light from a master laser.

The coupling between the fiber cores is achieved through amplitude mixing of the laser modes by Spatial Light Modulators at one end of the multicore fiber [97]. These Spatial Light Modulators couple light amplitude from the i-th core to the j-th core according to the prescribed connection weight J_{ij}. This coupling scheme is very similar to the one in the previous subsection.

Equations and comparison with Lagrange multipliers

As in prior physical examples, the dynamics can be expressed using slowly-varying equations for the polarization modes of the i-th core, E_{iL} and E_{iR}, where the two electric field amplitudes are in-phase temporally, are positive real, but have different polarization. They are,

$$\frac{d}{dt}E_{iL} = -\frac{\alpha_i}{2}E_{iL} + \frac{\gamma_i}{2}E_{iL} + \frac{1}{2}\sum_j J_{ij}\left(E_{jR} - E_{jL}\right)$$

$$\frac{d}{dt}E_{iR} = -\frac{\alpha_i}{2}E_{iR} + \frac{\gamma_i}{2}E_{iR} - \frac{1}{2}\sum_j J_{ij}\left(E_{jR} - E_{jL}\right)$$

where α_i is the decay rate in the i-th core, and γ_i is the gain in the i-th core. The third term on the right-hand side represents the coupling between the j-th and i-th cores that is provided by the Spatial Light Modulators. They next define the degree of polarization as $\mu_i \equiv E_{iL} - E_{iR}$. Subtracting the two equations above, we obtain the following evolution equation for μ_i:

$$\frac{d}{dt}\mu_i = -\frac{\alpha_i}{2}\mu_i + \frac{\gamma_i}{2}\mu_i + \sum_j J_{ij}\mu_j \tag{6.10}$$

The power dissipation is proportional to $|E_{iL}|^2 + |E_{iR}|^2$. But this can also be written $|E_{iL} - E_{iR}|^2 + |E_{iL} + E_{iR}|^2 = |\mu_i|^2 + |E_{iL} + E_{iR}|^2$. $|E_{iL} + E_{iR}|^2$ can be regarded as relatively constant as energy switches back and forth between right and left circular polarization. Then, power dissipation $h(\boldsymbol{\mu})$ would be most influenced by quadratic terms in $\boldsymbol{\mu}$:

$$h(\boldsymbol{\mu}, \boldsymbol{\gamma}) \propto -2\sum_{i,j} J_{ij}\mu_i\mu_j + \sum_i \alpha_i\mu_i^2 - \sum_i \gamma_i\mu_i^2$$

The right-hand side of the dynamics is again the partial derivative of the power dissipation function with respect to μ_i. Thus, the system performs a gradient descent minimization of the power dissipation, subject to the optical gain γ_i. Further, as before, we note that the dynamics equivalently perform gradient descent on the Lagrange function of the Ising problem defined by the J_{ij} couplings:

$$L(\boldsymbol{\mu}, \boldsymbol{\gamma}) = -\sum_{i,j} J_{ij}\mu_i\mu_j + \sum_i \left(\frac{\gamma_i - \alpha_i}{2}\right)\left(1 - \mu_i^2\right) \tag{6.11}$$

The net gains $\gamma_i - \alpha_i$ play the role of Lagrange multipliers.

Coupled Electrical Oscillators on the Unit Circle

We label this section 'Oscillators on the Unit Circle' to point out that the amplitudes of all the oscillators in this system are always nearly constant. The entire dynamics is in the phases. Hence, in quadrature space, the voltages move about on the unit circle. This is in contrast to Xiao's LC oscillator system from the previous chapter, where the phases of the oscillations collapsed early on to the cosine axis or the real axis. The dynamics was entirely in the amplitudes, and for that reason, Xiao's system may also be called 'Oscillators on the Real Axis'.

Overview

Networks of nonlinear, amplitude-stable electrical oscillators were designed by Roychowdhury et al. [98] to represent Ising systems. Their system consisted of dissipative coupled LC oscillators with clamped amplitude and the phase $\phi_i = 0$ or π revealing the preferred magnetic dipole orientation $\mu_{iz} = \pm 1$. It is noteworthy that Roychowdhury goes beyond Ising

machines and constructs general digital logic gates using these amplitude-stable oscillators in [124].

In their construction, Roychowdhury et al. [98] use nonlinear elements that behave like phase-independent negative resistors at low voltage amplitudes but as saturating resistance at high voltage amplitudes in order to keep the voltage amplitudes clamped at an intermediate value. This produces amplitude-stable oscillators. In addition, Roychowdhury et al. [98] provide a weak second harmonic pump and use a form of parametric amplification (referred to as sub-harmonic injection locking in [98]) to obtain bistability with respect to phase.

With the amplitudes being essentially clamped, it is the phases, 0 or π, that provide the magnetic dipole orientation $\mu_{iz} = \pm 1$. One key difference between this system and Yamamoto's system is that the latter had fast phase dynamics and slow amplitude dynamics, while Roychowdhury's system has the reverse. In the end, both methods map the Ising states to oscillations of a predetermined fixed amplitude but opposite phases.

Power dissipation

Roychowdhury et al. [98] derived the dynamics of their amplitude stable oscillator network using perturbation concepts developed in [125]. While a circuit diagram is not explicitly shown, [98] invokes the following dynamical equation for the phases of their electrical oscillators:

$$\frac{d\phi_i}{dt} = -\frac{1}{R_c}\sum_j J_{ij}\sin\left(\phi_i(t) - \phi_j(t)\right) - \lambda_i \sin\left(2\phi_i(t)\right) \tag{6.12}$$

where R_c is a coupling resistance in their system, ϕ_i is the phase of the i-th oscillator, and the λ_i are decay parameters that dictate how fast the phase angles settle towards their steady state values. The λ_i is related to the strength of the second-harmonic pump signal injected into the i-th oscillator.

(6.12) can be reproduced by iteratively minimizing the power dissipation in their system. Power dissipation across a resistor R_c is $(V_i - V_j)^2/R_c$ where $(V_i - V_j)$ is the voltage difference. Since V_i and V_j are sinusoidal, the power dissipation $P(\phi_i, \phi_j)$ consists of constant terms and a cross-term of the form:

$$P(\phi_i, \phi_j) = \frac{|V|^2 \cos\left(\phi_i - \phi_j\right)}{R_c}$$

Differentiating this with respect to ϕ_i leads to sine terms that resemble the first terms on the right-hand side of Eq. (6.12). We believe differentiating the power dissipation between the oscillators and their pumps will lead to the second term in Eq. (6.12).

Lagrange Multipliers

Magnetic dipole orientation parallel or anti-parallel is represented by whether $\phi_i - \phi_j = 0$ or π respectively. If the pump phase is chosen appropriately, we get $\phi_i = 0$ or π for all i. This can be implemented as:

$$g_i(\phi_i) = \left(\cos\left(2\phi_i\right) - 1\right) = 0$$

The Ising Lagrange function with constraints $g_i(\phi_i) = 0$ is proportional to:

$$L(\boldsymbol{\phi}, \boldsymbol{\lambda}) = \frac{1}{R_c} \sum_{i,j} J_{ij} \cos(\phi_i - \phi_j) + \sum_i \lambda_i (\cos(2\phi_i) - 1) \tag{6.13}$$

where λ_i is the Lagrange Multiplier corresponding to the phase angle constraint, and $J_{ij} = \pm 1$. The right-hand side of (6.12) is the gradient of (6.13), demonstrating that the dynamical system performs gradient descent on the Lagrange function.

The Lagrange function above implies that the system optimizes the following merit function f subject to constraints g_i:

$$f(\boldsymbol{\phi}) = \frac{1}{R_c} \sum_{i,j} J_{ij} \cos(\phi_i - \phi_j),$$
$$g_i(\phi_i) = (\cos(2\phi_i) - 1) = 0, \text{ for } i = 1, 2, \ldots, n.$$

Coupled polariton condensates

Overview

Kalinin and Berloff [99] proposed a system consisting of coupled polariton condensates to minimize the XY Hamiltonian. The XY Hamiltonian is a 2 dimensional version of the Ising Hamiltonian and is given by:

$$H(\boldsymbol{\mu}) = \sum_{ij} J_{ij} \boldsymbol{\mu_i} \cdot \boldsymbol{\mu_j}$$

where the $\boldsymbol{\mu_i}$ represents the magnetic moment vector of the i-th spin restricted to the spin-space XY plane. This formulation of the XY problem in the 2D plane is in fact easy to solve [126] because it can be recast as an semi-definite program (SDP), a type of very well-understood convex program.

Kalinin et al. pump a grid of coupled semiconductor microcavities with laser beams and observe the formation of strongly coupled exciton-photon states called polaritons. The laser pump, like those used in normal lasers, provides phase-independent gain. For our purposes, the polaritonic nomenclature is irrelevant. For us, these are simply coupled electromagnetic cavities that operate by the principle of minimum power dissipation similar to the previous cases. The complex electromagnetic amplitude in the i-th microcavity can be written $E_i = c_i + \mathrm{j}s_i$, where c_i and s_i represent the cosine and sine quadrature components of E_i, and j is the unit imaginary. c_i is mapped to the X-component of the magnetic dipole vector, and s_i to the Y-component. The electromagnetic microcavity system settles into a state of minimum power dissipation as the laser pump and optical gain are ramped up to compensate for the intrinsic cavity losses. The phase angles in the complex plane of the final electromagnetic modes are then reported as the corresponding $\boldsymbol{\mu}$-magnetic moment angles in the XY plane.

Since the electromagnetic cavities experience phase-independent gain, this system does not seek phase bistability. However, since we are actually searching for the magnetic dipole

vector angles on the *unit circle* in the XY plane, an amplitude constraint is required to remain on the unit circle, and it is indeed implemented.

Power dissipation and Lagrange function

Ref. [99] uses 'Ginzburg-Landau' equations to analyze their system resulting in equations for the complex amplitudes Ψ_i of the polariton wavefunctions. But the Ψ_i are actually the complex electric field amplitudes E_i of the i-th cavity. The electric field amplitudes satisfy the slowly-varying-amplitude equation:

$$\frac{dE_i}{dt} = \left(\frac{\gamma_i}{2} - \frac{\alpha_i}{2} - \beta|E_i|^2\right) E_i - iU|E_i|^2 E_i - \sum_j J_{ij} E_j \tag{6.14}$$

where γ_i is optical gain, α_i is linear loss, β is nonlinear attenuation, U is nonlinear phase shift, and J_{ij} is a dissipative cross-coupling-term representing linear loss. We note that both the amplitudes and phases of the electromagnetic modes are coupled to each other and evolve on comparable timescales. This is in contrast to ref. [98] where the main dynamics were embedded in phase—amplitude was fast and almost fixed—and to [96] where the dynamics were embedded in amplitude—phase was fast and almost fixed.

We show now that the method of ref. [99] is essentially the method of Lagrange multipliers with an added 'rotation'. The power dissipation rate is:

$$\begin{aligned} h(\boldsymbol{E}) =& -\frac{d}{dt}\sum_{i=1}^{N}\left(\frac{E_i^*}{2} + \frac{E_i}{2}\right)^2 \\ =& \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} J_{ij}\left(E_i^* E_j + E_i E_j^*\right) + \sum_{i=1}^{N}\beta|E_i|^4 + \sum_{i=1}^{N}\frac{\alpha_i}{2}|E_i|^2 - \sum_{i=1}^{N}\frac{\gamma_i}{2}|E_i|^2. \end{aligned}$$

The Lagrange function $L(\boldsymbol{E}, \boldsymbol{\gamma})$ for the saturation constraint, $g_i(E_i) = (1 - |E_i|^2) = 0$, is:

$$L(\boldsymbol{E}, \boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} J_{ij}\left(E_i^* E_j + E_i E_j^*\right) + \sum_{i=1}^{N}\beta|E_i|^4 + \sum_{i=1}^{N}\left(\frac{\gamma_i - \alpha_i}{2}\right)\left(1 - |E_i|^2\right) \tag{6.15}$$

Consistently with the previous subsections, the net linear gains are again reinterpreted as Lagrange Multipliers.

The steady state of the coupled polaritonic resonators scheme renders the power dissipation/Lagrange function stationary with respect to perturbations in the electric field. However, the dynamics does not go to that minimum via simple gradient descent as was the case in the previous systems. All the previous systems were performing gradient descent in time to reach the final steady-state, the steady-state which is characterized by the principle of minimum power dissipation. Dynamics (6.14), on the other hand, performs gradient descent on the power dissipation function (or, equivalently, the Lagrange function) **in conjunction with a rotation about the origin,** iU. This rotation term, iU, is not captured

by the power dissipation or Lagrange multiplier interpretations. It could, however, be useful in developing more sophisticated algorithms than the method of Lagrange multipliers and we comment on this prospect in Section 6.2 where a system with a more general 'rotation' term is discussed. The β term in the above equation is simply the nonlinear Augmented Lagrange term.

Evolution of Lagrange Multipliers

The papers of the previous subsections used simple heuristics like linear ramping to adjust their gains (which we have shown to be Lagrange multipliers). However, Kalinin et al. adjust their gains exactly according to the method of multipliers. We recall that the method of multipliers finds the optimal $\boldsymbol{x}^*$ and $\boldsymbol{\lambda}^*$ by performing gradient descent of L in $\boldsymbol{x}$ and gradient ascent of L in $\boldsymbol{\lambda}$.

Kalinin et al. employ the following dynamical equation for the gains γ_i in their coupled polariton system:

$$\frac{d\gamma_i}{dt} = \kappa' \left(1 - |E_i|^2\right) \tag{6.16}$$

where κ' is some constant. This exactly matches the gradient ascent equation in the method of multipliers for the λ variables.

We conclude this sub-section by splitting the Lagrange function into the effective merit function f, and the constraint function g_i. The extra 'phase rotation' U is not captured by this interpretation.

$$f(E_1, \ldots, E_n) = \frac{1}{2} \sum_{i,j} J_{ij} \left(E_i^* E_j + E_i E_j^*\right),$$
$$g_i(E_i) = \left(1 - |E_i|^2\right) = 0, \text{ for } i = 1, 2, \ldots, n$$

We do not show the β terms because they are simply the Augmented Lagrange penalty terms.

General conclusions from coupled multi-oscillator array Ising solvers

1. When subjected to input driving or gain, physical systems reach steady-states that minimize the net power dissipation rate.

2. All these systems actually perform Lagrange Multiplier optimization with the net gain $\gamma_i - \alpha_i$ (or some other appropriate physical variable in each oscillator i) playing the role of the Lagrange Multiplier for the i-th digital constraint.

3. Under the digital constraint, amplitudes $c_i = \pm 1$ or phases $\phi_i = 0$ or π, power dissipation minimization schemes are actually binary, similar to a flip-flop.

4. In many of the studied cases, the system dynamics approached a power dissipation minimum through gradient descent on the power dissipation function in the appropriate variables x_i. In one of the cases, the polariton condensate case, there was a rotation superimposed on this gradient descent.

6.2 Other methods in the literature

We now look at two other methods in the literature that do not explicitly implement the method of Lagrange Multipliers but nevertheless end up with dynamics that resemble it to varying extents. Both these methods offer operation regimes where the dynamics is not analogous to Lagrange multiplier optimization, and we believe it is an interesting avenue of future work to study the capabilities of these regimes.

Iterative Analog Matrix Multipliers

Soljacic et al. [100] developed an iterative procedure consisting of repeated matrix multiplication to solve the Ising problem. Their algorithm was implemented on a photonic circuit that utilized on-chip optical matrix multiplication units composed of Mach-Zehnder interferometers that were first introduced for matrix algebra by Zeilinger et al. in [127]. Soljacic et al. [100] showed that their algorithm optimized an effective merit function; we recast this function in the form of a Lagrange function in Appendix E. In this subsection, we use our insights from the previous sections to implement a simplified iterative Lagrange optimization using their platform of optical matrix multipliers. Details of the specific implementation of [100] can be found in Appendix E.

A block diagram of our iterative scheme is shown in Fig. 6.1. We found out after our work was published [128] that a similar scheme was proposed earlier in [129]. Let the multiple magnetic moment configuration of the Ising problem be represented as a vector of electric field amplitudes, E_i, of the spatially-separated optical modes. Each mode field amplitude represents the value of an Ising spin. In each iteration, the optical modes are fed into the optical circuit which performs matrix multiplication, and the resulting output optical modes are then fed back to the optical circuit input for the next iteration. Optical parametric gain sustains the successive iterations and enforces phase collapse on the optical pulses. Let c_i be the cosine component of the electric field of the i-th pulse E_i.

We wish to design the matrix multiplication unit such that it has a power dissipation function similar to the one found in Yamamoto et al.'s system:

$$h(\boldsymbol{c}) \propto -2\sum_{i,j} J_{ij}c_ic_j + \sum_i (\alpha_i - \gamma_i)\, c_i^2$$

The Lagrange function, including the binary constraint $c_i^2 = 1$, is given by:

$$L(\boldsymbol{c}, \boldsymbol{\gamma}) = -\sum_{i,j} J_{ij}c_ic_j + \sum_i \left(\frac{\gamma_i - \alpha_i}{2}\right)\left(1 - c_i^2\right), \tag{6.17}$$

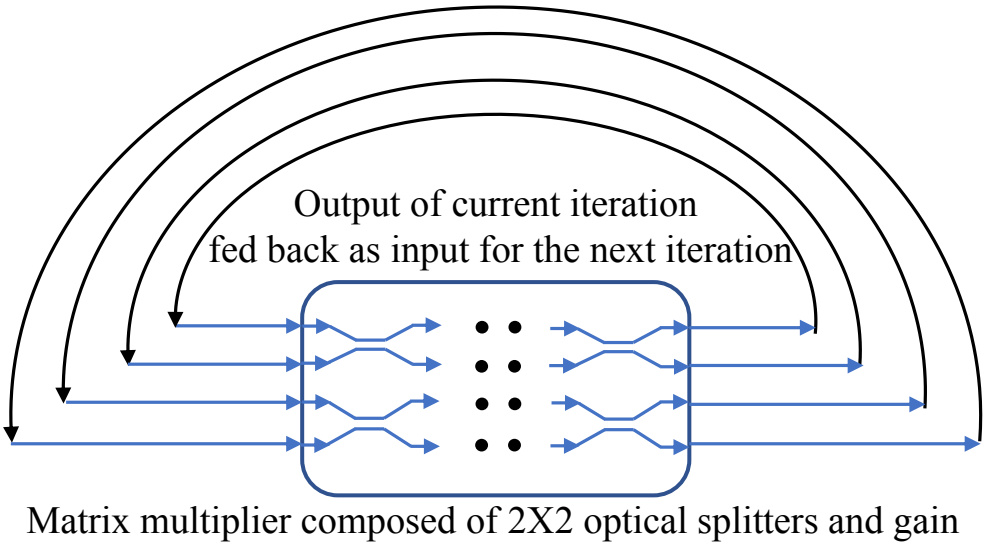

Figure 6.1: An optical circuit performing iterative multiplications converges on a solution of the Ising problem. Optical pulses are fed as input from the left hand side at the beginning of each iteration, pass through the matrix multiplication unit and are passed back from the outputs to the inputs for the next iteration. Distributed optical gain sustains the iterations.

where the J_{ij} is the dissipative loss associated with electric field interference between optical modes in the Mach-Zehnder interferometers, and γ_i is the optical gain.

The iterative multiplicative procedure that evolves the electric fields toward the minimum of the Lagrange function (6.17) is given by:

$$c_i(t+1) - c_i(t) = -\kappa \Delta t \frac{\partial}{\partial c_i} \left(-\sum_{i,j} J_{ij} c_i c_j + \sum_i \left(\frac{\gamma_i - \alpha_i}{2} \right) \left(1 - c_i^2 \right) \right),$$

where κ is a constant step size with the appropriate units and each iteration involves taking steps in c_i proportional to the partial derivatives $\partial / \partial c_i$ of the Lagrange function. Simplifying and sending all the terms involving time step t to one side, we get:

$$c_i(t+1) = \sum_j \left[\left(1 + \kappa \Delta t \left(\gamma_i - \alpha_i \right) \right) \delta_{ij} + 2 \kappa \Delta t J_{ij} \right] c_j(t), \tag{6.18}$$

where δ_{ij} is the Kronecker delta (1 only if $i = j$). The Mach-Zehnder interferometers should be tuned to the matrix $\left[\left(1 + \kappa \Delta t \left(\gamma_i - \alpha_i \right) \right) \delta_{ij} + 2 \kappa \Delta t J_{ij} \right]$. Thus, we have an iterative matrix multiplier scheme that minimizes the Lagrange function of the Ising problem. In effect, a lump of dissipative optical circuitry, compensated by optical gain, will, in a series of iterations, settle into a solution of the Ising problem.

The simple system above differs from that of Soljacic et al. [100] in that their method has added noise and nonlinear thresholding in each iteration. A description of their original approach is presented in Appendix E.

Leleu Mathematical Ising Solver

Leleu et al. [6] proposed a modified version of Yamamoto's Ising machine [93] that significantly resembles the Augmented Lagrange method while incorporating important new features. To understand the similarities and differences between Leleu's method and that of Lagrange multipliers, we recall the Augmented Lagrange function for the Ising problem from the previous chapter:

$$L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = -\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij}x_i x_j + \sum_{i=1}^{N}\left(\frac{\gamma_i - \alpha_i}{2}\right)\left(1 - x_i^2\right) + \frac{c}{2}\left(\sum_{i=1}^{N}\left(x_i^2 - 1\right)^2\right) \tag{6.19}$$

In the above, x_i are the optimization variables, J_{ij} is the interaction matrix, γ_i is the gain provided to the i-th variable, and α_i is the loss experienced by the i-th variable. To find a local optimum $(\boldsymbol{x}^*, \boldsymbol{\gamma}^*)$ that satisfies the constraints, one performs gradient descent on the Lagrange function in the $\boldsymbol{x}$ variables and gradient ascent in the $\boldsymbol{\gamma}$ variables. The equations are:

$$\dot{x}_i(t) = 2\kappa\left(\left(\frac{\gamma_i(t) - \alpha_i}{2}\right)x_i(t) + cx_i(t) - cx_i^3(t) + \sum_{j=1}^{N} J_{ij}x_j(t)\right) \tag{6.20}$$

$$\dot{\gamma}_i(t) = \frac{\kappa'}{2}\left(1 - x_i^2(t)\right) \tag{6.21}$$

On the other hand, Leleu et al. [6] propose the following system:

$$\frac{dx_i}{dt} = (\gamma - \alpha)x_i - \zeta x_i^3 + e_i\sum_j J_{ij}x_j, \tag{6.22}$$

$$\frac{de_i}{dt} = \beta(1 - x_i^2)e_i, \tag{6.23}$$

where the x_i are the optimization variables, α is the loss experienced by each variable, γ is a common gain supplied to each variable, ζ is a nonlinear loss coefficient, β is a positive parameter, and the e_i are error coefficients that capture how far away each x_i is from its saturation amplitude. Leleu's system achieves excellent performance on the Gset problem set as demonstrated in [6]. In their analysis, Leleu et al. show that all the fixed points of their system correspond to 1-flip local optima of the Ising problem, that is, flipping any one spin in a fixed point solution of the system cannot lead to a lower Ising energy.

It is clear that there are significant similarities between Leleu's system and the Lagrange multiplier system. The optimization variables in both systems experience linear losses and

gains and have interaction terms that capture the Ising interaction. Both systems have auxiliary variables that are varied according to how far away each degree of freedom is from its preferred saturation amplitude. However, the similarities end here. A major differentiation in Leleu's system is that e_i multiplies the Ising interaction felt by the i-th variable resulting in $e_i J_{ij}$. The complementary coefficient is $e_j J_{ij}$. Consequently, Leleu's equations implement asymmetric interactions $e_i J_{ij} \neq e_j J_{ij}$ between vector components x_i and x_j. We obtain some intuition about this system by splitting the asymmetric term $e_i J_{ij}$ into a symmetric and anti-symmetric part. This follows from the fact that any matrix A can be written as the sum of a symmetric matrix, $(A + A^T)/2$, and an anti-symmetric matrix, $(A - A^T)/2$. The symmetric part leads to gradient descent dynamics similar to all the other systems in this chapter. The anti-symmetric part causes a energy-conserving 'rotary' motion in the vector space of x_i.

It is possible that the secret of Leleu et al.'s improved performance lies in this anti-symmetric part. The dynamical freedom associated with asymmetry might provide a fruitful future research direction in optimization and deserves further study to ascertain its power.

6.3 Applications in Linear Algebra and Statistics

We have seen that minimum power dissipation machines serve as good Lagrange multiplier solvers for the Ising problem and other **NP**-hard problems that reduce to the Ising problem. In this section, we provide a straightforward application of minimum power dissipation solvers to the simple problem of linear regression in statistics. The problem of linear least squares regression, linear curve fitting with a quadratic error function, resembles the Ising problem. Since the error function to be minimized is quadratic anyway, the equivalence between linear least squares and Ising becomes exact when one demands that the regression weights be found in binary form to a certain number of bits. For this reason, the LC electrical circuit we presented in the previous chapter can be applied directly to linear regression. The circuit provides digital weights as the output but requires a series of binary resistance values, that is, $\ldots, 2R_0, R_0, 0.5R_0, \ldots$, to represent the input statistical observations/training data set.

The goal of linear least squares regression is to fit a linear function to a given set of data $\{(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), (\boldsymbol{x_3}, y_3), \ldots, (\boldsymbol{x_n}, y_n)\}$. The $\boldsymbol{x_i}$ s are input vectors of dimension d while the y_i are the observed outputs that we want our regression to capture. The linear function that is being fit to the data set is of the form $y(\boldsymbol{a}) = \sum_{i=1}^{d} w_i a_i$ where $\boldsymbol{a}$ is a feature vector of length d and $\boldsymbol{w}$ is a vector of unknown weights. The vector $\boldsymbol{w}$ is calculated by minimizing the sum of the squared errors it causes when used on the data set:

$$\boldsymbol{w^*} = \arg\min_{\boldsymbol{w}} \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{d} w_j x_{ij} \right) - y_i \right]^2,$$

where x_{ij} is the j-th component of the vector $\boldsymbol{x_i}$. This functional form is identical to the Ising Hamiltonian and we may construct an Ising circuit with $J_{ij} = \sum_{k=1}^{n} x_{ki} x_{kj}$ with the

weights $\boldsymbol{w}$ acting like the unknown magnetic moments. There is an effective magnetic field in the problem, $h_i = -2\sum_{j=1}^{n} x_{ji}y_j$. These local magnetic field terms h_i can be implemented in the LC Ising solver in a very straightforward fashion by introducing 'artificial spins with fixed phases and amplitudes', which is just a verbose definition of AC voltage sources. A simple circuit that solves this problem for $d = 2$ (each instance has two features) is provided in Fig. 6.2. Since the input data set decides the J matrix of the equivalent Ising problem, the input data are encoded in the values of the cross-coupling resistors. This circuit outputs each of the two weights to 2-'bit' precision. When we expand weights w_i in terms of 'bits' in our system, we use ± 1 'bits' and not the usual 0,1 bits.

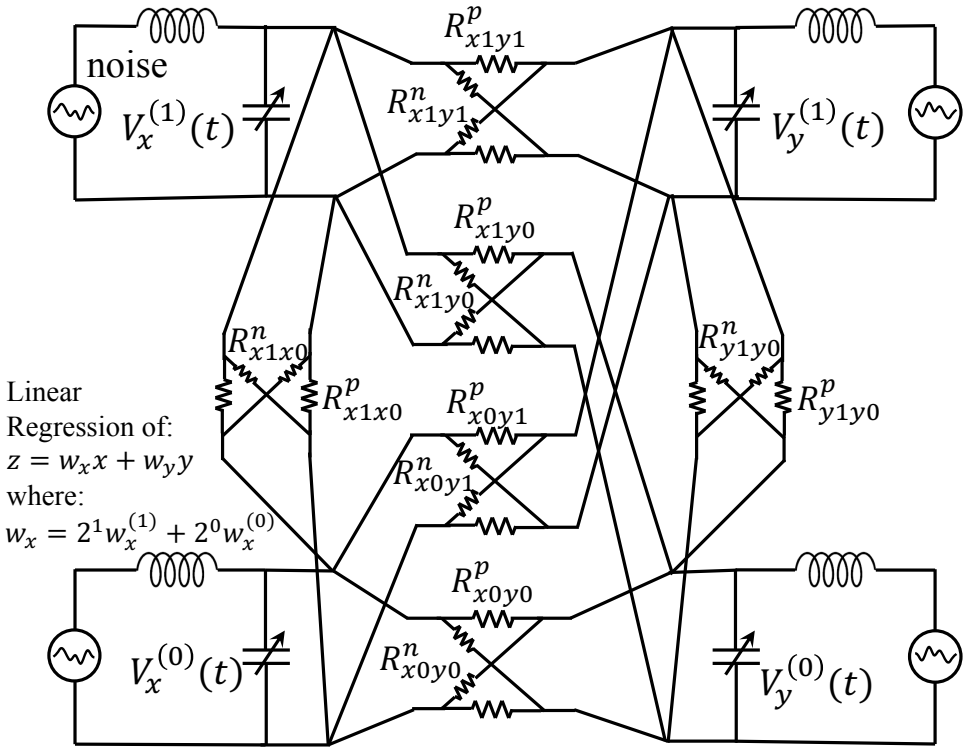

Figure 6.2: A two-bit, linear regression circuit, to find the best two curve-fitting weights w_d, using the Principle of Minimum Power Dissipation.

The oscillators on the left-hand side of the figure represent the 2^0 and 2^1 bits of the first weight while the oscillators on the other side represent the second weight.

The cross-resistance R that one would need to represent the J_{ij} that connects the units bit oscillator of the i-th weight with the units bit oscillator of the j-th weight is calculated as:

$$\frac{1}{R_{ij}} = \frac{b_1}{R_{-1}} + \frac{b_0}{R_0} + \frac{b_{-1}}{R_1} = \frac{1}{R_0}\left|J_{ij}\right|,$$

where $R_m = 2^m R_0$ is a binary hierarchy of resistances based on a reference resistor R_0, and b_m are the bits of $|J_{ij}|$: $|J_{ij}| = b_1 \times 2^1 + b_0 \times 2^0 + b_{-1} \times 2^{-1}$. This expansion for J_{ij} is the traditional binary expansion, with the b_i taking on values 0 or 1. The J_{ij} are encoded in the resistors. It is only the regression weights w_i, which are mapped onto the oscillator phase and amplitude, that are represented in a ± 1 'bit' representation. Representing J_{ij} to 3-bit precision requires resistors that span a dynamic range $2^2 = 4$. Further, the sign of the coupling is allotted according to whether the resistors R are parallel-connected or cross-connected. For connections between other bit places, say the 2^1 bit place oscillator of the i-th weight with the 2^{-2} bit place oscillator of the j-th weight, one needs to take those bit positions into account while computing the J, and hence the resistance value, that connects them. For this example, the effective J that connects the 2^1 bit of the i-th weight with the 2^{-2} bit of the i-th weight will be $J_{ij} \times 2^1 \times 2^{-2}$. This effective J is then expressed in binary form and the corresponding bits b_1, b_0, b_{-1} are used to construct the resistor between those two oscillators. In practice, the resistors R_{ij} would be externally programmed to the correct binary values using MOSFET switches that will determine which resistors take part in the coupling. If we have access to a resistor series that spans several orders of magnitude, we can implement J_{ij} to several bits of accuracy.

We have just solved the regression problem of the form $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$, where matrix $\boldsymbol{X}$ and vector $\boldsymbol{y}$ were known measurements and the corresponding best weight vector $\boldsymbol{w}$ for fitting was the unknown. We conclude by noting that this same procedure can be adopted to solve linear systems of equations of the form $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$.

6.4 Conclusion

The summary of this chapter is:

1. We discussed a variety of physical Ising solvers in the literature and showed that almost all of them performed Lagrange multiplier optimization. Moreover, there was an equivalence between the power dissipation and the Lagrange function in many of them.

2. Leleu's solver shows that it is possible to go beyond the basic Lagrange multipliers framework in intelligent ways to improve performance on benchmark problems.

3. We presented a straightforward application of the coupled LC oscillator Ising solver to perform least squares linear regression. Future work in this direction will tackle the challenge of extending these methods to the large-scale neural network optimizations prevalent today.

Chapter 7

Conclusion

7.1 Part 1: MOSFETs, subthreshold slope, tunnel transistors

To achieve a very steep slope in tunnel transistor response, many requirements must be fulfilled. Gate efficiency must be good. The concentration of defect energy levels, which are responsible for producing parasitic leakage, must be very low. In addition, the energy levels must themselves be very sharp. The usual assumption of Lorentzian spectral broadening would not provide sufficient ON/OFF ratio to satisfy system needs.

A correct use of Fermi's golden rule provides a mechanism for exponentially falling spectral tails, as required for a steep response. This demands that a narrowband, heavy effective mass wire should mediate between the tunneling energy levels and the external metallic contacts. There are many additional requirements, including the need for the inter-dot tunneling matrix element $|M|$ to be reduced such that the system does not enter strong coupling but large enough to provide the maximum conductance. In summary, we have shown that exponential spectral tails can arise naturally as part of lifetime broadening, from the initial parabolic decay of population as provided by the more complete form of Fermi's golden rule given here. In turn, exponential spectral tails justify the concept of exponential steepness or voltage swing, as had always been assumed in tFETs. We look forward to experimental verification of our predictions of ultra-sharp switching in the near future by groups working on high-purity, defect-free tunnel transistor systems.

We also presented a sequence of calculations of optical absorption in various model systems with the main aim of deriving the Urbach tail in the optical absorption of crystals. While we weren't able to achieve our main goal, we think it should be possible to decompose the spectrum of the crystal into a superposition of 2-level system spectra in a clever fashion. Then, the phonon-population-induced temperature dependent 2-level system spectrum that we derived can be plugged into that expression and an Urbach tail could possibly be obtained. We think this is a good direction for future work on this problem.

7.2 Part 2: Physics-based optimization—Ising solvers

In Part 2, we made the observation that physics has optimization principles at its heart, and that they can be exploited to design fast, low power, digital solvers that avoid the limits of standard computational paradigms. By appropriate choice of parameter values, the physical solvers we discussed can be used to perform Lagrange multiplier optimization orders-of-magnitude faster and with lower power than conventional digital computers. This performance advantage can be utilized for optimization in machine learning applications where energy and time considerations are critical. Large chip-scale implementations of minimum power dissipation Ising solvers have already been built by [103] and [98]. The important challenge ahead is to extend the scope of these machines to cover the large optimization problems involved in neural network training.

We add a note here about the power consumption of these machines. The act of computation can be regarded as a search among many possible answers. Finally, the circuit converges to a final correct configuration. Thus the initial conditions may include a huge phase space volume=2^n of possible solutions, ultimately transitioning into a final configuration representing a small or modest-sized binary number. This type of computing implies a substantial entropy reduction. This led to Landauer's admonition that computation costs $kn\log 2$ of entropy decrease, and $kTn\log 2$ of energy, for a final answer with n binary digits. By the 2$^{\text{nd}}$ Law of Thermodynamics, such an entropy reduction must be accompanied by an entropy increase elsewhere. In Landauer's viewpoint, the energy and entropy limit of computing was associated with the final acting of writing out the answer in n-bits, assuming the rest of the computer was reversible. In practice, technology consumes $\sim 10^4\times$ times more than the Landauer limit, owing to the insensitivity of the transistors operating at $\sim$1Volt, when they could be operating at $\sim$10mVolts.

In the continuously dissipative circuits we have described here, the total energy consumed would be very large if the machine is never turned off. If we terminate the powering of our optimizer systems after a desired final state answer is reached, the energy consumed becomes finite. By operating at voltage <1Volt and by powering off after the desired answer is achieved, our continuously dissipating Lagrange optimizers could actually be closer to the Landauer limit than a conventional digital computer.

It is important to keep in mind that the physical systems we studied evolve by steepest descent toward a local optimum, not a global optimum. Some of these systems make use of special adjustments to improve the quality of their solutions. Undoubtedly, more improvements might be possible, but none of the methods we discussed can always find the one global optimum which would be **NP**-hard [130].

In conclusion, the systems we studied here evolve toward an extremum through Lagrange function optimization where the Lagrange Multipliers are given by the gain or loss coefficients that keep the machine running. Nature provides us with a series of physical Optimization machines that are much faster and possibly more energy efficient than conventional computers. The next important challenge is to discover techniques to harness this power to solve power-hungry optimization problems in artificial intelligence.

Bibliography

[1] *https://irds.ieee.org/images/files/pdf/2020/2020IRDS_MM.pdf.*

[2] Sayeef Salahuddin and Supriyo Datta. "Use of negative capacitance to provide voltage amplification for low power nanoscale devices". In: *Nano letters* 8.2 (2008), pp. 405–410.

[3] Yichen Shen et al. "Deep learning with coherent nanophotonic circuits". en. In: *Nature Photonics* 11.7 (July 2017), pp. 441–446. ISSN: 1749-4885, 1749-4893. DOI: 10.1038/nphoton.2017.93. URL: http://www.nature.com/articles/nphoton.2017.93 (visited on 01/07/2020).

[4] Lars Onsager. "Reciprocal Relations in Irreversible Processes. II." en. In: *Physical Review* 38.12 (Dec. 1931), pp. 2265–2279. ISSN: 0031-899X. DOI: 10.1103/PhysRev.38.2265. URL: https://link.aps.org/doi/10.1103/PhysRev.38.2265 (visited on 07/10/2020).

[5] Andrew Lucas. "Ising formulations of many NP problems". In: *Frontiers in Physics* 2 (2014), p. 5.

[6] Timothée Leleu et al. "Destabilization of Local Minima in Analog Spin Systems by Correction of Amplitude Heterogeneity". en. In: *Physical Review Letters* 122.4 (Feb. 2019), p. 040607. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.122.040607. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.040607 (visited on 01/05/2020).

[7] *https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.*

[8] *https://www.mythic-ai.com/.*

[9] Ryan Hamerly et al. "Large-scale optical neural networks based on photoelectric multiplication". In: *Physical Review X* 9.2 (2019), p. 021032.

[10] Franz Urbach. "The Long-Wavelength Edge of Photographic Sensitivity and of the Electronic Absorption of Solids". en. In: *Physical Review* 92.5 (Dec. 1953), pp. 1324–1324. ISSN: 0031-899X. DOI: 10.1103/PhysRev.92.1324. URL: https://link.aps.org/doi/10.1103/PhysRev.92.1324 (visited on 07/31/2018).

[11] Adrian M. Ionescu and Heike Riel. "Tunnel field-effect transistors as energy-efficient electronic switches". en. In: *Nature* 479.7373 (Nov. 2011), pp. 329–337. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature10679. URL: http://www.nature.com/articles/nature10679 (visited on 01/10/2019).

[12] J.J. Quinn, Galen Kawamoto, and B.D. McCombe. "Subband spectroscopy by surface channel tunneling". en. In: *Surface Science* 73 (May 1978), pp. 190–196. ISSN: 00396028. DOI: 10.1016/0039-6028(78)90489-2. URL: http://linkinghub.elsevier.com/retrieve/pii/0039602878904892 (visited on 12/31/2018).

[13] S. Banerjee et al. "A new three-terminal tunnel device". en. In: *IEEE Electron Device Letters* 8.8 (Aug. 1987), pp. 347–349. ISSN: 0741-3106. DOI: 10.1109/EDL.1987.26655. URL: http://ieeexplore.ieee.org/document/1487205/ (visited on 12/31/2018).

[14] W Hansch et al. "A vertical MOS-gated Esaki tunneling transistor in silicon". en. In: *Thin Solid Films* 369.1-2 (July 2000), pp. 387–389. ISSN: 00406090. DOI: 10.1016/S0040-6090(00)00896-8. URL: http://linkinghub.elsevier.com/retrieve/pii/S0040609000008968 (visited on 12/31/2018).

[15] Junji Koga and Akira Toriumi. "Negative differential conductance in three-terminal silicon tunneling device". en. In: *Applied Physics Letters* 69.10 (Sept. 1996), pp. 1435–1437. ISSN: 0003-6951, 1077-3118. DOI: 10.1063/1.117606. URL: http://aip.scitation.org/doi/10.1063/1.117606 (visited on 12/31/2018).

[16] Toshio Baba. "Proposal for Surface Tunnel Transistors". en. In: *Japanese Journal of Applied Physics* 31.Part 2, No. 4B (Apr. 1992), pp. L455–L457. ISSN: 0021-4922. DOI: 10.1143/JJAP.31.L455. URL: http://stacks.iop.org/1347-4065/31/L455 (visited on 12/31/2018).

[17] William M. Reddick and Gehan A. J. Amaratunga. "Silicon surface tunnel transistor". en. In: *Applied Physics Letters* 67.4 (July 1995), pp. 494–496. ISSN: 0003-6951, 1077-3118. DOI: 10.1063/1.114547. URL: http://aip.scitation.org/doi/10.1063/1.114547 (visited on 12/31/2018).

[18] C. Aydin et al. "Lateral interband tunneling transistor in silicon-on-insulator". en. In: *Applied Physics Letters* 84.10 (Mar. 2004), pp. 1780–1782. ISSN: 0003-6951, 1077-3118. DOI: 10.1063/1.1668321. URL: http://aip.scitation.org/doi/10.1063/1.1668321 (visited on 12/31/2018).

[19] Saurabh Mookerjea et al. "Temperature-Dependent I– V Characteristics of a Vertical $\hbox{In}_{0.53}\hbox{Ga}_{0.47}\hbox{As}$ Tunnel FET". en. In: *IEEE Electron Device Letters* 31.6 (June 2010), pp. 564–566. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2010.2045631. URL: http://ieeexplore.ieee.org/document/5451179/ (visited on 07/30/2018).

[20] Mingda (Oscar) Li et al. "Single particle transport in two-dimensional heterojunction interlayer tunneling field effect transistor". en. In: *Journal of Applied Physics* 115.7 (Feb. 2014), p. 074508. ISSN: 0021-8979, 1089-7550. DOI: 10.1063/1.4866076. URL: http://aip.scitation.org/doi/10.1063/1.4866076 (visited on 07/30/2018).

[21] Anne S. Verhulst et al. "Tunnel field-effect transistor without gate-drain overlap". en. In: *Applied Physics Letters* 91.5 (July 2007), p. 053102. ISSN: 0003-6951, 1077-3118. DOI: 10.1063/1.2757593. URL: http://aip.scitation.org/doi/10.1063/1.2757593 (visited on 07/30/2018).

[22] K. E. Moselund et al. "InAs–Si Nanowire Heterojunction Tunnel FETs". en. In: *IEEE Electron Device Letters* 33.10 (Oct. 2012), pp. 1453–1455. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2012.2206789. URL: http://ieeexplore.ieee.org/document/6266688/ (visited on 07/27/2018).

[23] K. E. Moselund et al. "Comparison of VLS grown Si NW tunnel FETs with different gate stacks". en. In: *2009 Proceedings of the European Solid State Device Research Conference*. Athens, Greece: IEEE, Sept. 2009, pp. 448–451. ISBN: 978-1-4244-4351-2. DOI: 10.1109/ESSDERC.2009.5331446. URL: http://ieeexplore.ieee.org/document/5331446/ (visited on 12/31/2018).

[24] Osama M. Nayfeh et al. "Design of Tunneling Field-Effect Transistors Using Strained-Silicon/Strained-Germanium Type-II Staggered Heterojunctions". en. In: *IEEE Electron Device Letters* 29.9 (Sept. 2008), pp. 1074–1077. ISSN: 0741-3106. DOI: 10.1109/LED.2008.2000970. URL: http://ieeexplore.ieee.org/document/4604863/ (visited on 12/31/2018).

[25] F. Mayer et al. "Impact of SOI, Si1-xGexOI and GeOI substrates on CMOS compatible Tunnel FET performance". en. In: *2008 IEEE International Electron Devices Meeting*. San Francisco, CA, USA: IEEE, Dec. 2008, pp. 1–5. ISBN: 978-1-4244-2377-4. DOI: 10.1109/IEDM.2008.4796641. URL: http://ieeexplore.ieee.org/document/4796641/ (visited on 12/31/2018).

[26] Ram Krishna Ghosh and Santanu Mahapatra. "Monolayer Transition Metal Dichalcogenide Channel-Based Tunnel Transistor". en. In: *IEEE Journal of the Electron Devices Society* 1.10 (Oct. 2013), pp. 175–180. ISSN: 2168-6734. DOI: 10.1109/JEDS.2013.2292799. URL: http://ieeexplore.ieee.org/document/6675794/ (visited on 12/31/2018).

[27] Joachim Knoch. "Optimizing tunnel FET performance - Impact of device structure, transistor dimensions and choice of material". en. In: *2009 International Symposium on VLSI Technology, Systems, and Applications*. Hsinchu, Taiwan: IEEE, Apr. 2009, pp. 45–46. ISBN: 978-1-4244-2784-0. DOI: 10.1109/VTSA.2009.5159285. URL: http://ieeexplore.ieee.org/document/5159285/ (visited on 12/31/2018).

[28] Ian A Young, Uygar E. Avci, and Daniel H. Morris. "Tunneling field effect transistors: Device and circuit considerations for energy efficient logic opportunities". en. In: *2015 IEEE International Electron Devices Meeting (IEDM)*. Washington, DC, USA: IEEE, Dec. 2015, pp. 22.1.1–22.1.4. ISBN: 978-1-4673-9894-7. DOI: 10.1109/IEDM.2015.7409755. URL: http://ieeexplore.ieee.org/document/7409755/ (visited on 01/09/2019).

[29] Hamilton Carrillo-Nunez, Mathieu Luisier, and Andreas Schenk. "InAs-GaSb/Si heterojunction tunnel MOSFETs: An alternative to TFETs as energy-efficient switches?" en. In: *2015 IEEE International Electron Devices Meeting (IEDM)*. Washington, DC: IEEE, Dec. 2015, pp. 34.6.1–34.6.4. ISBN: 978-1-4673-9894-7. DOI: 10.1109/IEDM.2015.7409829. URL: http://ieeexplore.ieee.org/document/7409829/ (visited on 01/09/2019).

[30] Alan Seabaugh et al. "Steep subthreshold swing tunnel FETs: GaN/InN/GaN and transition metal dichalcogenide channels". en. In: *2015 IEEE International Electron Devices Meeting (IEDM)*. Washington, DC, USA: IEEE, Dec. 2015, pp. 35.6.1–35.6.4. ISBN: 978-1-4673-9894-7. DOI: 10.1109/IEDM.2015.7409835. URL: http://ieeexplore.ieee.org/document/7409835/ (visited on 01/09/2019).

[31] Chunlei Wu et al. "Design Guideline for Complementary Heterostructure Tunnel FETs With Steep Slope and Improved Output Behavior". en. In: *IEEE Electron Device Letters* 37.1 (Jan. 2016), pp. 20–23. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2015.2499183. URL: http://ieeexplore.ieee.org/document/7322213/ (visited on 01/10/2019).

[32] Jacopo Franco et al. "Intrinsic Robustness of TFET Subthreshold Swing to Interface and Oxide Traps: A Comparative PBTI Study of InGaAs TFETs and MOSFETs". en. In: *IEEE Electron Device Letters* 37.8 (Aug. 2016), pp. 1055–1058. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2016.2584983. URL: http://ieeexplore.ieee.org/document/7502178/ (visited on 01/10/2019).

[33] Zhaonian Yang. "Tunnel Field-Effect Transistor With an L-Shaped Gate". en. In: *IEEE Electron Device Letters* 37.7 (July 2016), pp. 839–842. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2016.2574821. URL: http://ieeexplore.ieee.org/document/7482645/ (visited on 01/10/2019).

[34] Chenming Hu et al. "Prospect of tunneling green transistor for 0.1V CMOS". en. In: *2010 International Electron Devices Meeting*. San Francisco, CA, USA: IEEE, Dec. 2010, pp. 16.1.1–16.1.4. ISBN: 978-1-4424-7418-5. DOI: 10.1109/IEDM.2010.5703372. URL: http://ieeexplore.ieee.org/document/5703372/ (visited on 12/31/2018).

[35] Alan C. Seabaugh and Qin Zhang. "Low-Voltage Tunnel Transistors for Beyond CMOS Logic". en. In: *Proceedings of the IEEE* 98.12 (Dec. 2010), pp. 2095–2110. ISSN: 0018-9219, 1558-2256. DOI: 10.1109/JPROC.2010.2070470. URL: http://ieeexplore.ieee.org/document/5608485/ (visited on 07/27/2018).

[36] Emanuele Baravelli et al. "Optimization of n- and p-type TFETs Integrated on the Same ${\rm InAs}/{\rm Al}_{x}{\rm Ga}_{1-x}{\rm Sb}$ Technology Platform". en. In: *IEEE Transactions on Electron Devices* 61.1 (Jan. 2014), pp. 178–185. ISSN: 0018-9383, 1557-9646. DOI: 10.1109/TED.2013.2289739. URL: http://ieeexplore.ieee.org/document/6670064/ (visited on 07/30/2018).

[37] Ram Asra et al. "A Tunnel FET for V_{DD} Scaling Below 0.6 V With a CMOS-Comparable Performance". en. In: *IEEE Transactions on Electron Devices* 58.7 (July 2011), pp. 1855–1863. ISSN: 0018-9383, 1557-9646. DOI: 10.1109/TED.2011.2140322. URL: http://ieeexplore.ieee.org/document/5765666/ (visited on 12/31/2018).

[38] Pengyu Long et al. "Design and Simulation of GaSb/InAs 2D Transmission-Enhanced Tunneling FETs". en. In: *IEEE Electron Device Letters* 37.1 (Jan. 2016), pp. 107–110. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2015.2497666. URL: http://ieeexplore.ieee.org/document/7317763/ (visited on 01/10/2019).

[39] Jiadi Zhu et al. "Design and Simulation of a Novel Graded-Channel Heterojunction Tunnel FET With High ${I}_{\scriptscriptstyle\text{ON}}/{I}_{\scriptscriptstyle\text{OFF}}$ Ratio and Steep Swing". en. In: *IEEE Electron Device Letters* 38.9 (Sept. 2017), pp. 1200–1203. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2017.2734679. URL: http://ieeexplore.ieee.org/document/7999254/ (visited on 01/10/2019).

[40] Joshua T. Smith, Saptarshi Das, and Joerg Appenzeller. "Broken-Gap Tunnel MOSFET: A Constant-Slope Sub-60-mV/decade Transistor". en. In: *IEEE Electron Device Letters* 32.10 (Oct. 2011), pp. 1367–1369. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2011.2162220. URL: http://ieeexplore.ieee.org/document/5977000/ (visited on 01/10/2019).

[41] Hyunjae Lee, Jung-Dong Park, and Changhwan Shin. "Performance Booster for Vertical Tunnel Field-Effect Transistor: Field-Enhanced High- $\kappa $ Layer". en. In: *IEEE Electron Device Letters* 37.11 (Nov. 2016), pp. 1383–1386. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2016.2606660. URL: http://ieeexplore.ieee.org/document/7562532/ (visited on 01/10/2019).

[42] Hesameddin Ilatikhameneh et al. "Dielectric Engineered Tunnel Field-Effect Transistor". en. In: *IEEE Electron Device Letters* 36.10 (Oct. 2015), pp. 1097–1100. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2015.2474147. URL: http://ieeexplore.ieee.org/document/7229273/ (visited on 01/10/2019).

[43] J. Appenzeller et al. "Band-to-Band Tunneling in Carbon Nanotube Field-Effect Transistors". en. In: *Physical Review Letters* 93.19 (Nov. 2004). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.93.196805. URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.196805 (visited on 12/31/2018).

[44] Lars Knoll et al. "Inverters With Strained Si Nanowire Complementary Tunnel Field-Effect Transistors". en. In: *IEEE Electron Device Letters* 34.6 (June 2013), pp. 813–815. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2013.2258652. URL: http://ieeexplore.ieee.org/document/6517573/ (visited on 07/30/2018).

[45] Qianqian Huang et al. "A novel Si tunnel FET with 36mV/dec subthreshold slope based on junction depleted-modulation through striped gate configuration". en. In: *2012 International Electron Devices Meeting*. San Francisco, CA, USA: IEEE, Dec. 2012, pp. 8.5.1–8.5.4. ISBN: 978-1-4673-4871-3 978-1-4673-4872-0 978-1-4673-4870-6. DOI: 10.1109/IEDM.2012.6479005. URL: http://ieeexplore.ieee.org/document/6479005/ (visited on 12/31/2018).

[46] Ramanathan Gandhi et al. "CMOS-Compatible Vertical-Silicon-Nanowire Gate-All-Around p-Type Tunneling FETs With ≤ 50-mV/decade Subthreshold Swing". en. In: *IEEE Electron Device Letters* 32.11 (Nov. 2011), pp. 1504–1506. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2011.2165331. URL: http://ieeexplore.ieee.org/document/6029281/ (visited on 12/31/2018).

[47] Tejas Krishnamohan et al. "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and ≪60mV/dec subthreshold slope". en. In: *2008 IEEE International Electron Devices Meeting*. San Francisco, CA, USA: IEEE, Dec. 2008, pp. 1–3. ISBN: 978-1-4244-2377-4. DOI: 10.1109/IEDM.2008.4796839. URL: http://ieeexplore.ieee.org/document/4796839/ (visited on 12/31/2018).

[48] Kanghoon Jeon et al. "Si tunnel transistors with a novel silicided source and 46mV/dec swing". en. In: *2010 Symposium on VLSI Technology*. Honolulu, HI, USA: IEEE, June 2010, pp. 121–122. ISBN: 978-1-4244-5451-8. DOI: 10.1109/VLSIT.2010.5556195. URL: http://ieeexplore.ieee.org/document/5556195/ (visited on 12/31/2018).

[49] Deblina Sarkar et al. "A subthermionic tunnel field-effect transistor with an atomically thin channel". en. In: *Nature* 526.7571 (Oct. 2015), pp. 91–95. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature15387. URL: http://www.nature.com/articles/nature15387 (visited on 01/02/2019).

[50] E.A. Casu et al. "Hybrid phase-change — Tunnel FET (PC-TFET) switch with sub-threshold swing ¡ 10mV/decade and sub-0.1 body factor: Digital and analog benchmarking". en. In: *2016 IEEE International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE, Dec. 2016, pp. 19.3.1–19.3.4. ISBN: 978-1-5090-3902-9. DOI: 10.1109/IEDM.2016.7838452. URL: http://ieeexplore.ieee.org/document/7838452/ (visited on 01/02/2019).

[51] E. Memisevic et al. "Vertical InAs/GaAsSb/GaSb tunneling field-effect transistor on Si with S = 48 mV/decade and Ion = 10 A/m for Ioff = 1 nA/m at Vds = 0.3 V". en. In: *2016 IEEE International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE, Dec. 2016, pp. 19.1.1–19.1.4. ISBN: 978-1-5090-3902-9. DOI: 10.1109/

IEDM.2016.7838450. URL: http://ieeexplore.ieee.org/document/7838450/ (visited on 01/02/2019).

[52] Peng Zhang et al. "CMOS-compatible FDSOI bipolar-enhanced tunneling FET". In: *2015 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*. Rohnert Park, CA, USA: IEEE, Oct. 2015, pp. 1–3. ISBN: 978-1-5090-0259-7. DOI: 10.1109/S3S.2015.7333513. URL: http://ieeexplore.ieee.org/document/7333513/ (visited on 01/09/2019).

[53] Y. Morita et al. "Experimental realization of complementary p- and n- tunnel FinFETs with subthreshold slopes of less than 60 mV/decade and very low (pA/μm) off-current on a Si CMOS platform". en. In: *2014 IEEE International Electron Devices Meeting*. San Francisco, CA, USA: IEEE, Dec. 2014, pp. 9.7.1–9.7.4. ISBN: 978-1-4799-8001-7. DOI: 10.1109/IEDM.2014.7047020. URL: http://ieeexplore.ieee.org/document/7047020/ (visited on 01/09/2019).

[54] Ali Saeidi et al. "Negative Capacitance as Performance Booster for Tunnel FETs and MOSFETs: An Experimental Study". en. In: *IEEE Electron Device Letters* 38.10 (Oct. 2017), pp. 1485–1488. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2017.2734943. URL: http://ieeexplore.ieee.org/document/8000343/ (visited on 01/10/2019).

[55] Xin Zhao, Alon Vardi, and Jesus A. del Alamo. "Sub-Thermal Subthreshold Characteristics in Top–Down InGaAs/InAs Heterojunction Vertical Nanowire Tunnel FETs". en. In: *IEEE Electron Device Letters* 38.7 (July 2017), pp. 855–858. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2017.2702612. URL: http://ieeexplore.ieee.org/document/7922554/ (visited on 01/10/2019).

[56] Elvedin Memisevic et al. "Vertical Nanowire TFETs With Channel Diameter Down to 10 nm and Point S $_{MIN}$ of 35 mV/Decade". en. In: *IEEE Electron Device Letters* 39.7 (July 2018), pp. 1089–1091. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2018.2836862. URL: https://ieeexplore.ieee.org/document/8359117/ (visited on 01/10/2019).

[57] Ali Saeidi et al. "Near Hysteresis-Free Negative Capacitance InGaAs Tunnel FETs with Enhanced Digital and Analog Figures of Merit below $V_{DD} = 400$mV". In: *2018 IEEE International Electron Devices Meeting (IEDM)*. IEEE. 2018, pp. 13–4.

[58] Bijesh Rajamohanan et al. "0.5 V Supply Voltage Operation of In $_{0.65}$ Ga $_{0.35}$ As/GaAs $_{0.4}$ Sb $_{0.6}$ Tunnel FET". en. In: *IEEE Electron Device Letters* 36.1 (Jan. 2015), pp. 20–22. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2014.2368147. URL: http://ieeexplore.ieee.org/document/6949120/ (visited on 01/10/2019).

[59] Sapan Agarwal et al. "Engineering the Electron–Hole Bilayer Tunneling Field-Effect Transistor". en. In: *IEEE Transactions on Electron Devices* 61.5 (May 2014), pp. 1599–1606. ISSN: 0018-9383, 1557-9646. DOI: 10.1109/TED.2014.2312939. URL: http://ieeexplore.ieee.org/document/6784495/ (visited on 07/30/2018).

[60] J. L. Padilla, F. Gamiz, and A. Godoy. "A Simple Approach to Quantum Confinement in Tunneling Field-Effect Transistors". en. In: *IEEE Electron Device Letters* 33.10 (Oct. 2012), pp. 1342–1344. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2012.2207876. URL: http://ieeexplore.ieee.org/document/6272315/ (visited on 07/30/2018).

[61] Kathy Boucart and Adrian Mihai Ionescu. "Double-Gate Tunnel FET With High-κ Gate Dielectric". en. In: *IEEE Transactions on Electron Devices* 54.7 (July 2007), pp. 1725–1733. ISSN: 0018-9383. DOI: 10.1109/TED.2007.899389. URL: http://ieeexplore.ieee.org/document/4252356/ (visited on 07/27/2018).

[62] Amey M. Walke et al. "Fabrication and Analysis of a ${\rm Si}/{\rm Si}_{0.55}{\rm Ge}_{0.45}$ Heterojunction Line Tunnel FET". en. In: *IEEE Transactions on Electron Devices* 61.3 (Mar. 2014), pp. 707–715. ISSN: 0018-9383, 1557-9646. DOI: 10.1109/TED.2014.2299337. URL: http://ieeexplore.ieee.org/document/6727530/ (visited on 07/27/2018).

[63] Sorin Cristoloveanu, Jing Wan, and Alexander Zaslavsky. "A Review of Sharp-Switching Devices for Ultra-Low Power Applications". en. In: *IEEE Journal of the Electron Devices Society* 4.5 (Sept. 2016), pp. 215–226. ISSN: 2168-6734. DOI: 10.1109/JEDS.2016.2545978. URL: http://ieeexplore.ieee.org/document/7465689/ (visited on 07/31/2018).

[64] Anil W. Dey et al. "High-Current GaSb/InAs(Sb) Nanowire Tunnel Field-Effect Transistors". en. In: *IEEE Electron Device Letters* 34.2 (Feb. 2013), pp. 211–213. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2012.2234078. URL: http://ieeexplore.ieee.org/document/6410332/ (visited on 07/27/2018).

[65] Joachim Knoch, Siegfried Mantl, and J Appenzeller. "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices". In: *Solid-State Electronics* 51.4 (2007), pp. 572–578.

[66] Sapan Agarwal and Eli Yablonovitch. "Pronounced Effect of pn-Junction Dimensionality on Tunnel Switch Threshold Shape". In: *arXiv:1109.0096 [cond-mat]* (Sept. 2011). arXiv: 1109.0096. URL: http://arxiv.org/abs/1109.0096 (visited on 07/27/2018).

[67] Haojun Zhang et al. "Effect of band-tails on the subthreshold performance of 2D tunnel-FETs". en. In: *2016 IEEE International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE, Dec. 2016, pp. 30.3.1–30.3.4. ISBN: 978-1-5090-3902-9. DOI: 10.1109/IEDM.2016.7838512. URL: http://ieeexplore.ieee.org/document/7838512/ (visited on 01/09/2019).

[68] E. Memisevic et al. "Impact of Band-Tails on the Subthreshold Swing of III-V Tunnel Field-Effect Transistor". en. In: *IEEE Electron Device Letters* 38.12 (Dec. 2017), pp. 1661–1664. ISSN: 0741-3106, 1558-0563. DOI: 10.1109/LED.2017.2764873. URL: http://ieeexplore.ieee.org/document/8076904/ (visited on 01/10/2019).

[69] Sri Krishna Vadlamani et al. "Tunnel-fet switching is governed by non-lorentzian spectral line shape". In: *Proceedings of the IEEE* 108.8 (2019), pp. 1235–1244.

[70] Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloë. *Quantum mechanics*. New York: Wiley, 1977. ISBN: 978-0-471-16435-7.

[71] Claude Cohen-Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg. *Processus d'interaction entre photons et atomes*. Les Ulis France Paris: EDP Sciences CNRS Éditions, 2001. ISBN: 0-471-62556-6.

[72] Alexander I. Saichev and Wojbor Woyczynski. "Asymptotics of Fourier Transforms". In: *Distributions in the Physical and Engineering Sciences, Volume 1: Distributional and Fractal Calculus, Integral Transforms and Wavelets*. Cham: Springer International Publishing, 2018, pp. 93–135. ISBN: 978-3-319-97958-8. DOI: 10.1007/978-3-319-97958-8_4. URL: https://doi.org/10.1007/978-3-319-97958-8_4.

[73] H Bateman. *Table of integral transforms, Vol 1(1954), Ed A. Erdélyi*.

[74] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Vol. 55. US Government printing office, 1964.

[75] Sapan Agarwal et al. *Designing a low voltage, high current tunneling transistor*. Cambridge, UK: Cambridge Univ. Press, 2015.

[76] Zafer Mutlu et al. "Transfer-Free Synthesis of Atomically Precise Graphene Nanoribbons on Insulating Substrates". In: *ACS nano* 15.2 (2021), pp. 2635–2642.

[77] Kyoko Nakada et al. "Edge state in graphene ribbons: Nanometer size effect and edge shape dependence". In: *Physical Review B* 54.24 (1996), p. 17954.

[78] Mitsutaka Fujita et al. "Peculiar localized state at zigzag graphite edge". In: *Journal of the Physical Society of Japan* 65.7 (1996), pp. 1920–1923.

[79] Y. Pan et al. "Atomistic Origin of Urbach Tails in Amorphous Silicon". en. In: *Physical Review Letters* 100.20 (May 2008). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.100.206403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.206403 (visited on 07/26/2018).

[80] J Liebler and H Haug. "Calculation of the Urbach Tail Absorption in a Second-Order Cumulant Expansion". en. In: *Europhysics Letters (EPL)* 14.1 (Jan. 1991), pp. 71–76. ISSN: 0295-5075, 1286-4854. DOI: 10.1209/0295-5075/14/1/013. URL: http://stacks.iop.org/0295-5075/14/i=1/a=013?key=crossref.963d731db79067aa6ed8808a7b8ffbb1 (visited on 07/26/2018).

[81] Sajeev John et al. "Density of states for an electron in a correlated Gaussian random potential: Theory of the Urbach tail". en. In: *Physical Review B* 37.12 (Apr. 1988), pp. 6963–6976. ISSN: 0163-1829. DOI: 10.1103/PhysRevB.37.6963. URL: https://link.aps.org/doi/10.1103/PhysRevB.37.6963 (visited on 07/26/2018).

[82] B. Sadigh et al. "First-Principles Calculations of the Urbach Tail in the Optical Absorption Spectra of Silica Glass". en. In: *Physical Review Letters* 106.2 (Jan. 2011). ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.106.027401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.027401 (visited on 07/26/2018).

[83] Sajeev John et al. "Theory of electron band tails and the Urbach optical-absorption edge". In: *Physical review letters* 57.14 (1986), p. 1777.

[84] Thomas H. Keil. "Theory of the Urbach Rule". en. In: *Physical Review* 144.2 (Apr. 1966), pp. 582–587. ISSN: 0031-899X. DOI: 10.1103/PhysRev.144.582. URL: https://link.aps.org/doi/10.1103/PhysRev.144.582 (visited on 07/26/2018).

[85] John D. Dow and David Redfield. "Toward a Unified Theory of Urbach's Rule and Exponential Absorption Edges". en. In: *Physical Review B* 5.2 (Jan. 1972), pp. 594–610. ISSN: 0556-2805. DOI: 10.1103/PhysRevB.5.594. URL: https://link.aps.org/doi/10.1103/PhysRevB.5.594 (visited on 07/26/2018).

[86] Herbert Mahr. "Absorption band shape and Urbach's rule of localized excitons". In: *Physical Review* 132.5 (1963), p. 1880.

[87] B. I. Halperin and Melvin Lax. "Impurity-Band Tails in the High-Density Limit. I. Minimum Counting Methods". en. In: *Physical Review* 148.2 (Aug. 1966), pp. 722–740. ISSN: 0031-899X. DOI: 10.1103/PhysRev.148.722. URL: https://link.aps.org/doi/10.1103/PhysRev.148.722 (visited on 07/26/2018).

[88] Hitoshi Sumi and Yutaka Toyozawa. "Urbach-Martienssen rule and exciton trapped momentarily by lattice vibrations". In: *Journal of the Physical Society of Japan* 31.2 (1971), pp. 342–358.

[89] Amnon Yariv. *Quantum electronics*. New York: Wiley, 1989. ISBN: 9780471609971.

[90] Ryogo Kubo. "A stochastic theory of line shape". In: *Stochastic processes in chemical physics* 15 (1969), pp. 101–127.

[91] Marvin L Cohen and Steven G Louie. *Fundamentals of condensed matter physics*. Cambridge University Press, 2016.

[92] Shoko Utsunomiya, Kenta Takata, and Yoshihisa Yamamoto. "Mapping of Ising models onto injection-locked laser systems". In: *Optics express* 19.19 (2011), pp. 18091–18108.

[93] Yoshitaka Haribara, Shoko Utsunomiya, and Yoshihisa Yamamoto. "Computational Principle and Performance Evaluation of Coherent Ising Machine Based on Degenerate Optical Parametric Oscillator Network". en. In: *Entropy* 18.4 (Apr. 2016), p. 151. ISSN: 1099-4300. DOI: 10.3390/e18040151. URL: http://www.mdpi.com/1099-4300/18/4/151 (visited on 01/05/2020).

[94] Takahiro Inagaki et al. "Large-scale Ising spin network based on degenerate optical parametric oscillators". In: *Nature Photonics* 10.6 (2016), p. 415.

[95] T. Inagaki et al. "A coherent Ising machine for 2000-node optimization problems". en. In: *Science* 354.6312 (Nov. 2016), pp. 603–606. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aah4243. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.aah4243 (visited on 01/07/2020).

[96] Tianyao Patrick Xiao. "Optoelectronics for refrigeration and analog circuits for combinatorial optimization". PhD Thesis. UC Berkeley, 2019.

[97] Masoud Babaeian et al. "A single shot coherent Ising machine based on a network of injection-locked multicore fiber lasers". en. In: *Nature Communications* 10.1 (Dec. 2019), p. 3516. ISSN: 2041-1723. DOI: 10.1038/s41467-019-11548-4. URL: http://www.nature.com/articles/s41467-019-11548-4 (visited on 01/27/2020).

[98] Tianshi Wang and Jaijeet Roychowdhury. "OIM: Oscillator-Based Ising Machines for Solving Combinatorial Optimisation Problems". en. In: *Unconventional Computation and Natural Computation*. Ed. by Ian McQuillan and Shinnosuke Seki. Vol. 11493. Cham: Springer International Publishing, 2019, pp. 232–256. ISBN: 978-3-030-19310-2 978-3-030-19311-9. DOI: 10.1007/978-3-030-19311-9_19. URL: http://link.springer.com/10.1007/978-3-030-19311-9_19 (visited on 01/05/2020).

[99] Kirill P. Kalinin and Natalia G. Berloff. "Global optimization of spin Hamiltonians with gain-dissipative systems". en. In: *Scientific Reports* 8.1 (Dec. 2018), p. 17791. ISSN: 2045-2322. DOI: 10.1038/s41598-018-35416-1. URL: http://www.nature.com/articles/s41598-018-35416-1 (visited on 07/11/2020).

[100] Charles Roques-Carmes et al. "Heuristic recurrent algorithms for photonic Ising machines". en. In: *Nature Communications* 11.1 (Dec. 2020), p. 249. ISSN: 2041-1723. DOI: 10.1038/s41467-019-14096-z. URL: http://www.nature.com/articles/s41467-019-14096-z (visited on 07/11/2020).

[101] Simon Mahler et al. "Improved Phase Locking of Laser Arrays with Nonlinear Coupling". en. In: *Physical Review Letters* 124.13 (Apr. 2020), p. 133901. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.124.133901. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.133901 (visited on 08/19/2020).

[102] D. Pierangeli, G. Marcucci, and C. Conti. "Large-Scale Photonic Ising Machine by Spatial Light Modulation". en. In: *Physical Review Letters* 122.21 (May 2019), p. 213902. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.122.213902. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.213902 (visited on 07/20/2020).

[103] Ibrahim Ahmed, Po-Wei Chiu, and Chris H Kim. "A Probabilistic Self-Annealing Compute Fabric Based on 560 Hexagonally Coupled Ring Oscillators for Solving Combinatorial Optimization Problems". In: *2020 IEEE Symposium on VLSI Circuits*. IEEE. 2020, pp. 1–2.

[104] Mária Ercsey-Ravasz and Zoltán Toroczkai. "Optimization hardness as transient chaos in an analog approach to constraint satisfaction". en. In: *Nature Physics* 7.12 (Dec. 2011). Number: 12 Publisher: Nature Publishing Group, pp. 966–970. ISSN: 1745-2481. DOI: 10.1038/nphys2105. URL: https://www.nature.com/articles/nphys2105/ (visited on 08/19/2020).

[105] Botond Molnár et al. "A continuous-time MaxSAT solver with high analog performance". en. In: *Nature Communications* 9.1 (Dec. 2018), p. 4864. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07327-2. URL: http://www.nature.com/articles/s41467-018-07327-2 (visited on 07/11/2020).

[106] Fabio L. Traversa and Massimiliano Di Ventra. "Polynomial-time solution of prime factorization and NP-complete problems with digital memcomputing machines". en. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.2 (Feb. 2017), p. 023107. ISSN: 1054-1500, 1089-7682. DOI: 10.1063/1.4975761. URL: http://aip.scitation.org/doi/10.1063/1.4975761 (visited on 08/19/2020).

[107] Wolfgang Maass, Thomas Natschläger, and Henry Markram. "Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations". en. In: *Neural Computation* 14.11 (Nov. 2002), pp. 2531–2560. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/089976602760407955. URL: https://www.mitpressjournals.org/doi/abs/10.1162/089976602760407955 (visited on 08/19/2020).

[108] Gouhei Tanaka et al. "Recent advances in physical reservoir computing: A review". en. In: *Neural Networks* 115 (July 2019), pp. 100–123. ISSN: 08936080. DOI: 10.1016/j.neunet.2019.03.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0893608019300784 (visited on 08/19/2020).

[109] Hayato Goto, Kosuke Tatsumura, and Alexander R. Dixon. "Combinatorial optimization by simulating adiabatic bifurcations in nonlinear Hamiltonian systems". en. In: *Science Advances* 5.4 (Apr. 2019), eaav2372. ISSN: 2375-2548. DOI: 10.1126/sciadv.aav2372. URL: http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.aav2372 (visited on 01/05/2020).

[110] William A. Borders et al. "Integer factorization using stochastic magnetic tunnel junctions". en. In: *Nature* 573.7774 (Sept. 2019), pp. 390–393. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-019-1557-9. URL: http://www.nature.com/articles/s41586-019-1557-9 (visited on 08/19/2020).

[111] Rolf Landauer. "Irreversibility and heat generation in the computing process". In: *IBM journal of research and development* 5.3 (1961), pp. 183–191.

[112] Ilya Prigogine. *Etude Thermodynamique des Phénomènes irréversibles*. Chap V. Editions Desoer, Liège, 1947.

[113] S.R. de Groot. *Thermodynamics of Irreversible Processes*. Chap X. Interscience Publishers, Inc., New York, 1951.

[114] Peter JM Van Laarhoven and Emile HL Aarts. "Simulated annealing". In: *Simulated annealing: Theory and applications*. Springer, 1987, pp. 7–15.

[115] Neil G. Dickson et al. "Thermally assisted quantum annealing of a 16-qubit problem". In: *Nature communications* 4 (2013), p. 1903.

[116] Neumann John Von. *Non-linear capacitance or inductance switching, amplifying, and memory organs*. US Patent 2,815,488. Dec. 1957.

[117] Eiichi Goto. "The parametron, a digital computing element which utilizes parametric oscillation". In: *Proceedings of the IRE* 47.8 (1959), pp. 1304–1316.

[118] Dimitri Bertsekas. *Nonlinear programming*. Athena Scientific, 1999.

[119] *https://www.nii.ac.jp/qis/first-quantum/forStudents/lecture/pdf/noise/chapter11.pdf*.

[120] *https://web.stanford.edu/ yyye/yyye/Gset/*.

[121] Una Benlic and Jin-Kao Hao. "Breakout Local Search for the Max-Cutproblem". en. In: *Engineering Applications of Artificial Intelligence* 26.3 (Mar. 2013), pp. 1162–1173. ISSN: 09521976. DOI: 10.1016/j.engappai.2012.09.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0952197612002175 (visited on 01/05/2020).

[122] Peter L. McMahon et al. "A fully programmable 100-spin coherent Ising machine with all-to-all connections". In: *Science* 354.6312 (2016), pp. 614–617.

[123] Zhe Wang et al. "Coherent Ising machine based on degenerate optical parametric oscillators". In: *Physical Review A* 88.6 (2013), p. 063853.

[124] Jaijeet Roychowdhury. "Boolean Computation Using Self-Sustaining Nonlinear Oscillators". en. In: *Proceedings of the IEEE* 103.11 (Nov. 2015), pp. 1958–1969. ISSN: 0018-9219, 1558-2256. DOI: 10.1109/JPROC.2015.2483061. URL: http://ieeexplore.ieee.org/document/7299247/ (visited on 07/10/2020).

[125] Alper Demir, Amit Mehrotra, and Jaijeet Roychowdhury. "Phase noise in oscillators: A unifying theory and numerical methods for characterization". In: *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 47.5 (2000), pp. 655–674.

[126] Michel X Goemans and David P Williamson. "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming". In: *Journal of the ACM (JACM)* 42.6 (1995), pp. 1115–1145.

[127] Michael Reck et al. "Experimental realization of any discrete unitary operator". en. In: *Physical Review Letters* 73.1 (July 1994), pp. 58–61. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.73.58. URL: https://link.aps.org/doi/10.1103/PhysRevLett.73.58 (visited on 07/11/2020).

[128] Sri Krishna Vadlamani, Tianyao Patrick Xiao, and Eli Yablonovitch. "Physics successfully implements Lagrange multiplier optimization". In: *Proceedings of the National Academy of Sciences* 117.43 (2020), pp. 26639–26650.

[129] Nasim Mohammadi Estakhri, Brian Edwards, and Nader Engheta. "Inverse-designed metastructures that solve equations". In: *Science* 363.6433 (2019), pp. 1333–1338.

[130] Richard M. Karp. "Reducibility among combinatorial problems". In: *Complexity of computer computations*. Springer, 1972, pp. 85–103.

[131] Andrei Tokmakoff. *5.74 Introductory Quantum Mechanics II. Spring 2009.* MIT OpenCourseWare. Cambridge MA: Massachusetts Institute of Technology. URL: `https://ocw.mit.edu/courses/chemistry/5-74-introductory-quantum-mechanics-ii-spring-2009/`.

[132] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Appendix A

Optical absorption in a damped harmonic potential well—No rotating wave approximation

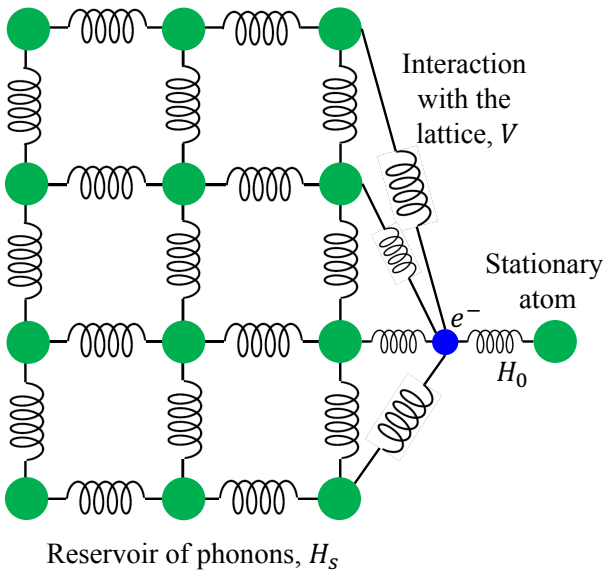

Figure A.1: The electron is in a harmonic potential well and is also coupled to a large crystal which is at thermal equilibrium. The phonons in the crystal interact with the electron and broaden its absorption spectrum. The phonons are described by the Hamiltonian H_s, the harmonic well by H_0, and their interaction by V.

In this appendix, we rederive the optical absorption of an electron in a phonon-damped harmonic potential, but for a more complete Hamiltonian that takes into account processes that were ignored in the main text.

Hamiltonian

The first two terms, the harmonic oscillator's own Hamiltonian, H_0, and the reservoir term, H_s, will be the same:

$$H_o = \hbar\omega_0 \left(b^\dagger b + \frac{1}{2} \right) \tag{A.1}$$

$$H_s = \sum_i \hbar\omega_i \left(a_i^\dagger a_i + \frac{1}{2} \right) \tag{A.2}$$

where $b/b^\dagger$, $a_i/a_i^\dagger$, ω_0, and ω_i have the same meanings as before. We are going to add additional terms to the coupling term, V:

$$V = \sum_i \left(f_i b a_i + f_i^* b^\dagger a_i^\dagger + g_i b^\dagger a_i + g_i^* b a_i^\dagger \right) \tag{A.3}$$

where the f_i and g_i are the coupling strengths between the electron and the i-th phonon. We originally had only the g_i terms, which physically correspond to the creation or destruction of an electron excitation in response to a phonon destruction or creation respectively. To make the analysis more complete, we have now added the f_i terms which physically correspond to the creation of an electron excitation in response to a phonon creation and the destruction of an electron excitation in response to a phonon destruction. These terms are ignored in the 'Rotating Wave Approximation', which we performed implicitly in the main text. However, this approximation is accurate only if the phonon frequencies are close to the electron frequency. If there is a big offset between the phonon and electron frequencies, as is often the case, all these terms have to be taken into account.

Heisenberg equations

The Heisenberg equations for the operators are:

$$i\hbar \frac{d}{dt} b(t) = \hbar\omega_0 b(t) + \sum_i g_i a_i(t) + \sum_i f_i^* a_i^\dagger(t) \tag{A.4}$$

$$i\hbar \frac{d}{dt} a_i(t) = \hbar\omega_i a_i(t) + g_i^* b(t) + f_i^* b^\dagger(t) \tag{A.5}$$

Introducing 'slowly-varying' operators:

$$b(t) = \hat{b}(t) e^{-i\omega_0 t} \tag{A.6}$$

$$a_i(t) = \hat{a}_i(t) e^{-i\omega_i t} \tag{A.7}$$

The Heisenberg equations become:

$$i\hbar \frac{d}{dt} \hat{b}(t) = \sum_i g_i \hat{a}_i(t) e^{i(\omega_0 - \omega_i)t} + \sum_i f_i^* \hat{a}_i{}^\dagger(t) e^{i(\omega_0 + \omega_i)t} \tag{A.8}$$

$$i\hbar \frac{d}{dt} \hat{a}_i(t) = g_i^* \hat{b}(t) e^{i(\omega_i - \omega_0)t} + f_i^* \hat{b}^\dagger(t) e^{i(\omega_i + \omega_0)t} \tag{A.9}$$

We integrate the $\hat{a}_i(t)$ differential equations (and the $\hat{a}_i^\dagger(t)$ equations too) from 0 to t and substitute the result into the $\hat{b}(t)$ differential equation to get:

$$\begin{aligned}\frac{d}{dt}\hat{b}(t) = &-\frac{i}{\hbar}\sum_i \left[g_i a_i e^{i(\omega_0-\omega_i)t} + f_i^* a_i^\dagger e^{i(\omega_0+\omega_i)t}\right] \\ &-\frac{1}{\hbar^2}\int_0^t d\tau \ \left[\sum_i |g_i|^2 e^{i(\omega_0-\omega_i)\tau} - \sum_i |f_i|^2 e^{i(\omega_0+\omega_i)\tau}\right]\hat{b}(t-\tau) \\ &-e^{2i\omega_0 t}\frac{1}{\hbar^2}\int_0^t d\tau \ \left[\sum_i g_i f_i^* e^{-i(\omega_0+\omega_i)\tau} - \sum_i g_i f_i^* e^{-i(\omega_0-\omega_i)\tau}\right]\hat{b}^\dagger(t-\tau)\end{aligned} \tag{A.10}$$

The equation for $\hat{b}^\dagger(t)$ is obtained simply by conjugating the above. We condense the above equation by introducing new functions:

$$\hat{F}(t) = -\frac{i}{\hbar}\sum_i \left[g_i a_i e^{i(\omega_0-\omega_i)t} + f_i^* a_i^\dagger e^{i(\omega_0+\omega_i)t}\right] \tag{A.11}$$

$$\kappa(\tau) = \frac{1}{\hbar^2}\left[\sum_i |g_i|^2 e^{i(\omega_0-\omega_i)\tau} - \sum_i |f_i|^2 e^{i(\omega_0+\omega_i)\tau}\right] \tag{A.12}$$

$$\lambda(\tau) = \frac{1}{\hbar^2}\left[\sum_i g_i f_i^* e^{-i(\omega_0+\omega_i)\tau} - \sum_i g_i f_i^* e^{-i(\omega_0-\omega_i)\tau}\right] \tag{A.13}$$

Then, Eq. (A.10) reduces to:

$$\frac{d}{dt}\hat{b}(t) = -\int_0^t d\tau \ \kappa(\tau)\hat{b}(t-\tau) - e^{2i\omega_0 t}\int_0^t d\tau \ \lambda(\tau)\hat{b}^\dagger(t-\tau) + \hat{F}(t) \tag{A.14}$$

The equivalent equation for $\hat{b}^\dagger(t)$ is:

$$\frac{d}{dt}\hat{b}^\dagger(t) = -\int_0^t d\tau \ \kappa^*(\tau)\hat{b}^\dagger(t-\tau) - e^{-2i\omega_0 t}\int_0^t d\tau \ \lambda^*(\tau)\hat{b}(t-\tau) + \hat{F}^\dagger(t) \tag{A.15}$$

We have coupled equations with memory kernels but they are still solvable through the use of Laplace transforms. Recall that we are denoting the Laplace transform of a function $x(t)$ by $x^L(s)$. Then, the solution to Eq. (A.14) is:

$$\hat{b}(t) = \left(\int_0^t d\tau \ h(t-\tau)\hat{F}(\tau)\right) - \left(\int_0^t d\tau \ h(t-\tau)e^{2i\omega_0\tau}\int_0^\tau d\tau' \ \lambda(\tau')\hat{b}^\dagger(\tau-\tau')\right) + h(t)\hat{b} \tag{A.16}$$

while $\hat{b}^\dagger(t)$ is given by:

$$\hat{b}^\dagger(t) = \left(\int_0^t d\tau \ h^*(t-\tau)\hat{F}^\dagger(\tau)\right) - \left(\int_0^t d\tau \ h^*(t-\tau)e^{-2i\omega_0\tau}\int_0^\tau d\tau' \ \lambda^*(\tau')\hat{b}(\tau-\tau')\right) + h^*(t)\hat{b}^\dagger \tag{A.17}$$

where $h(t)$, like in the main chapter, is the function whose Laplace transform is given by:

$$h^L(s) = \frac{1}{s + \kappa^L(s)} \tag{A.18}$$

$\kappa^L(s)$ is the Laplace transform of $\kappa(\tau)$, and is given by:

$$\kappa^L(s) = \frac{1}{\hbar^2} \sum_i \frac{|g_i|^2}{s - i\left(\omega_0 - \omega_i\right)} - \frac{1}{\hbar^2} \sum_i \frac{|f_i|^2}{s - i\left(\omega_0 + \omega_i\right)} \tag{A.19}$$

$\lambda^L(s)$ is the Laplace transform of $\lambda(\tau)$, and is given by:

$$\lambda^L(s) = \frac{1}{\hbar^2} \sum_i \frac{g_i f_i^*}{s + i\left(\omega_0 + \omega_i\right)} - \frac{1}{\hbar^2} \sum_i \frac{g_i f_i^*}{s + i\left(\omega_0 - \omega_i\right)} \tag{A.20}$$

The Laplace transform of $h^*(t)$ is denoted by $\left(h^*\right)^L(s)$ and given by:

$$\left(h^*\right)^L(s) = \left(h^L(s^*)\right)^* = \frac{1}{s + \left(\kappa^L(s^*)\right)^*} \tag{A.21}$$

Dipole commutator autocorrelation

The expression for the dipole commutator autocorrelation is the same as before:

$$\langle [p(t), p(0)] \rangle = \frac{e^2\hbar}{2m_e\omega_0} \left(\langle [b(t), b] \rangle + \left\langle \left[b(t), b^\dagger\right] \right\rangle + \left\langle \left[b^\dagger(t), b\right] \right\rangle + \left\langle \left[b^\dagger(t), b^\dagger\right] \right\rangle \right) \tag{A.22}$$

The first commutator is:

$$\left\langle \left[\hat{b}(t), b\right] \right\rangle = -\left(\int_0^t d\tau\ h(t-\tau) e^{2i\omega_0\tau} \int_0^\tau d\tau'\ \lambda(\tau') \left\langle \left[\hat{b}^\dagger(\tau - \tau'), b\right] \right\rangle \right) \tag{A.23}$$

The second commutator is:

$$\left\langle \left[\hat{b}(t), b^\dagger\right] \right\rangle = h(t) - \left(\int_0^t d\tau\ h(t-\tau) e^{2i\omega_0\tau} \int_0^\tau d\tau'\ \lambda(\tau') \left\langle \left[\hat{b}^\dagger(\tau - \tau'), b^\dagger\right] \right\rangle \right) \tag{A.24}$$

The third commutator is:

$$\left\langle \left[\hat{b}^\dagger(t), b\right] \right\rangle = -h^*(t) - \left(\int_0^t d\tau\ h^*(t-\tau) e^{-2i\omega_0\tau} \int_0^\tau d\tau'\ \lambda^*(\tau') \left\langle \left[\hat{b}(\tau - \tau'), b\right] \right\rangle \right) \tag{A.25}$$

The fourth commutator is:

$$\left\langle \left[\hat{b}^\dagger(t), b^\dagger\right] \right\rangle = -\left(\int_0^t d\tau\ h^*(t-\tau) e^{-2i\omega_0\tau} \int_0^\tau d\tau'\ \lambda^*(\tau') \left\langle \left[\hat{b}(\tau - \tau'), b^\dagger\right] \right\rangle \right) \tag{A.26}$$

Susceptibility $\chi(\omega)$

We take the Laplace transform on both sides of each of the four commutator equations above. Denoting the Laplace transform of $\left\langle \left[\hat{b}(t), b\right] \right\rangle$ by $\hat{B}B(s)$, that of $\left\langle \left[\hat{b}(t), b^\dagger\right] \right\rangle$ by $\hat{B}B^\dagger(s)$, and so on, we have:

$$\hat{B}B(s) = -h^L(s)\lambda^L(s-2i\omega_0)\hat{B}^\dagger B(s-2i\omega_0) \tag{A.27}$$

$$\hat{B}B^\dagger(s) = h^L(s) - h^L(s)\lambda^L(s-2i\omega_0)\hat{B}^\dagger B^\dagger(s-2i\omega_0) \tag{A.28}$$

$$\hat{B}^\dagger B(s) = -\left(h^*\right)^L(s) - \left(h^*\right)^L(s)\left(\lambda^*\right)^L(s+2i\omega_0)\hat{B}B(s+2i\omega_0) \tag{A.29}$$

$$\hat{B}^\dagger B^\dagger(s) = -\left(h^*\right)^L(s)\left(\lambda^*\right)^L(s+2i\omega_0)\hat{B}B^\dagger(s+2i\omega_0) \tag{A.30}$$

Solving these equations and denoting the Laplace transform of $\langle [b(t), b] \rangle$ by $BB(s)$, that of $\left\langle \left[b(t), b^\dagger\right] \right\rangle$ by $BB^\dagger(s)$, and so on, we have:

$$BB(s) = \frac{h^L(s+i\omega_0)\lambda^L(s-i\omega_0)\left(h^L(s^*+i\omega_0)\right)^*}{1-\left(h^L(s^*+i\omega_0)\right)^*\left(\lambda^L(s^*-i\omega_0)\right)^* h^L(s+i\omega_0)\lambda^L(s-i\omega_0)} \tag{A.31}$$

$$B^\dagger B(s) = \frac{-\left(h^L(s^*+i\omega_0)\right)^*}{1-\left(h^L(s^*+i\omega_0)\right)^*\left(\lambda^L(s^*-i\omega_0)\right)^* h^L(s+i\omega_0)\lambda^L(s-i\omega_0)} \tag{A.32}$$

$$BB^\dagger(s) = \frac{h^L(s+i\omega_0)}{1-h^L(s+i\omega_0)\lambda^L(s-i\omega_0)\left(h^L(s^*+i\omega_0)\right)^*\left(\lambda^L(s^*-i\omega_0)\right)^*} \tag{A.33}$$

$$B^\dagger B^\dagger(s) = \frac{-\left(h^*\right)^L(s-i\omega_0)\left(\lambda^L(s^*-i\omega_0)\right)^* h^L(s+i\omega_0)}{1-h^L(s+i\omega_0)\lambda^L(s-i\omega_0)\left(h^L(s^*+i\omega_0)\right)^*\left(\lambda^L(s^*-i\omega_0)\right)^*} \tag{A.34}$$

The susceptibility is then:

$$\chi(\omega) = \frac{i}{\hbar}\frac{e^2\hbar}{2m_e\omega_0}\int_0^\infty \langle [p(t), p(0)] \rangle\, e^{-i\omega\tau} d\tau \tag{A.35}$$

$$= \frac{i}{\hbar}\frac{e^2\hbar}{2m_e\omega_0}\lim_{\eta\to 0}\left(BB(i\omega+\eta) + B^\dagger B(i\omega+\eta) + BB^\dagger(i\omega+\eta) + B^\dagger B^\dagger(i\omega+\eta)\right) \tag{A.36}$$

$$= \frac{i}{\hbar}\frac{e^2\hbar}{2m_e\omega_0}\left(\frac{2\left(\Gamma_\lambda(\omega)+i\Delta_\lambda(\omega)\right) - i\left(2\omega_0+\Delta(\omega)+\Delta(-\omega)\right) - \Gamma(\omega)+\Gamma(-\omega)}{\left[i\left(\omega-\omega_0-\Delta(-\omega)\right)+\Gamma(-\omega)\right]\left[i\left(\omega+\omega_0+\Delta(\omega)\right)+\Gamma(\omega)\right] + \left(\Gamma_\lambda(\omega)+i\Delta_\lambda(\omega)\right)^2}\right) \tag{A.37}$$

where we have defined:

$$\Gamma(\omega) = \frac{\pi}{\hbar^2}\left(\sum_i |g_i|^2 \delta\left(\omega + \omega_i\right) - \sum_i |f_i|^2 \delta\left(\omega - \omega_i\right)\right) \tag{A.38}$$

$$\Delta(\omega) = \frac{1}{\hbar^2}\left(\sum_i \frac{|f_i|^2}{\omega - \omega_i} - \sum_i \frac{|g_i|^2}{\omega + \omega_i}\right) \tag{A.39}$$

$$\Gamma_\lambda(\omega) = \frac{1}{\hbar^2}\sum_i g_i f_i^* \pi\left[\delta\left(\omega + \omega_i\right) - \delta\left(\omega - \omega_i\right)\right] \tag{A.40}$$

$$\Delta_\lambda(\omega) = \frac{i}{\hbar^2}\sum_i g_i f_i^* \frac{2\omega_i}{\omega^2 - \omega_i^2} \tag{A.41}$$

We do not explicitly extract the imaginary part, $\chi''(\omega)$, from the above formula due to the tedium involved, and the fact that it doesn't add physically significant information to our understanding.

Discussion

Unfortunately, Eq. (A.37) is still completely independent of temperature. It seems like simple linear dynamics of the physical operators is insufficient to introduce temperature dependence in the absorption lineshape.

Appendix B

Motional Narrowing

Let us consider the setting of Nuclear Magnetic Resonance (NMR). A nuclear spin is placed in a certain medium that is entirely immersed in an external magnetic field. The magnetic field causes Zeeman splitting of the nucleus's energy levels and one can observe transitions between the two different spin states of the nucleus via absorption of light. If the nucleus and its environment are stationary and the magnetic field is time-invariant, the absorption takes place at a fixed frequency. However, if the nucleus randomly jumps around spatially, it could experience different values of local magnetic field through the course of its journey due to inhomogeneities in the applied magnetic field or inhomogeneities in the nucleus's environment. That is, the energy level splitting could vary with time, leading to a 'jumping' of the absorption frequency of the nucleus.

In this appendix, we consider a system whose absorption (or dipole oscillation) frequency jumps randomly, due to external disturbances, within a Gaussian distribution $G(\omega)$ of frequencies generated by its inhomogeneous surroundings. If the frequency jumps around slowly compared to the spread of frequencies in $G(\omega)$, the absorption spectrum of the system essentially matches the Gaussian distribution $G(\omega)$. If, on the other hand, the frequency jumps are very fast compared to the width of $G(\omega)$, one obtains an absorption spectrum that is surprisingly much narrower than the external distribution of frequencies $G(\omega)$. This phenomenon, called 'Motional Narrowing', was discovered by Bloembergen and was later extensively studied by researchers. Kubo [90] studied a particular 2-parameter statistical model of this problem and derived absorption lineshapes. We look at this model and point out that it does not yield exponential tails in frequency for any setting of parameter values. We present here the calculation in Andrei Tokmakoff's MIT OpenCourseWare notes [131].

B.1 Jumping within a Gaussian distribution of frequencies

The dipole moment at time t is given by $p(t) = p(0)e^{-i\int_0^t dt_1\ \omega(t_1)}$. The frequency $\omega(t_1)$ is jumping around randomly in time within a background Gaussian frequency distribution that

has variance Δ^2. Let the average time interval at which jumps occur be τ_c. We will assume for simplicity that the Gaussian is centered about $\omega = 0$.

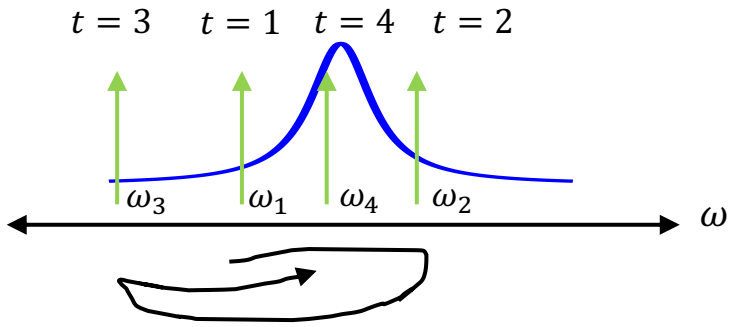

Figure B.1: The frequency of the absorption is jumping around randomly as time passes within a Gaussian distribution of frequencies. At t=1, it is at ω_1, at $t = 2$ at ω_2, and so on.

Now, it turns out that the above 'jump' dynamics of the frequency can be described using a stochastic Langevin equation:

$$\frac{d}{dt}\omega(t) = -\frac{1}{\tau_c}\omega(t) + n(t) \tag{B.1}$$

where τ_c is the jumping time constant and $n(t)$ is a Gaussian white-noise process with strength σ^2, that is, $\langle n(t_1)n(t_2)\rangle = \sigma^2\delta(t_1 - t_2)$. We will now relate σ^2 to Δ and τ_c. From the above, we have:

$$\omega(t) = e^{-t/\tau_c}\int_{-\infty}^{t} d\tau_1 \; e^{\tau_1/\tau_c} n(\tau_1) \tag{B.2}$$

which further yields, for $t_2 > t_1$:

$$\langle \omega(t_1)\omega(t_2)\rangle = e^{-(t_2-t_1)/\tau_c}\frac{\sigma^2}{2/\tau_c} \tag{B.3}$$

Since we know that the variance of ω's distribution is Δ^2, we have:

$$\Delta^2 = \frac{\sigma^2}{2/\tau_c} \tag{B.4}$$

Now that we have exactly captured our jumping process using the Langevin equation with the appropriate parameters, we compute the dipole autocorrelation:

$$\langle p(0)p(t)\rangle = p(0)^2 \left\langle e^{-i\int_0^t dt_1 \; \omega(t_1)} \right\rangle \tag{B.5}$$

The average on the right is the characteristic function of the Gaussian random variable $\int_0^t dt_1\ \omega(t_1)$. For Gaussian random variables, one can rewrite the characteristic function using the two-term cumulant expansion [131]:

$$\left\langle e^{-i\int_0^t dt_1\ \omega(t_1)} \right\rangle = e^{-i\int_0^t dt_1\ \langle\omega(t_1)\rangle} e^{-\frac{1}{2}\int_0^t d\tau' \int_0^t d\tau'' \langle\omega(\tau')\omega(\tau'')\rangle} \tag{B.6}$$

Since we have assumed that the distribution of ω is centered around $\omega = 0$, the first term in the product on the right is just 1. The right-side simplifies to:

$$\left\langle e^{-i\int_0^t dt_1\ \omega(t_1)} \right\rangle = e^{-\frac{1}{2}\int_0^t d\tau' \int_0^t d\tau'' \langle\omega(\tau')\omega(\tau'')\rangle} = e^{-\int_0^t d\tau\ (t-\tau)\langle\omega(\tau)\omega(0)\rangle} \tag{B.7}$$

Plugging in our expression for the autocorrelation of $\omega(t)$, we have [131]:

$$\langle p(0)p(t)\rangle = p(0)^2 \exp\left(-\Delta^2\tau_c^2\left(e^{-t/\tau_c} + \frac{t}{\tau_c} - 1\right)\right) \tag{B.8}$$

The absorption spectrum is the Fourier transform of the dipole autocorrelation. The autocorrelation is simplified in the two limits $\Delta \gg 1/\tau_c$ and $\Delta \ll 1/\tau_c$ to obtain the standard background Gaussian lineshape and the motionally narrowed Lorentzian lineshape respectively.

B.2 Impossibility of getting asymptotically exponential lineshapes

If we Taylor expand the exponent of e on the right-hand side, we see something interesting. The constant and linear terms of e^{-t/τ_c} get canceled by the other two terms, but the other Taylor terms remain. In particular, the odd powers of t, 3 and above, all survive.

We recall now the discussion of Chapter 2, Section 6. The absorption spectrum is the Fourier transform of the dipole autocorrelation. Since the dipole autocorrelation has odd powers of t in it, it has discontinuities in its odd derivatives. Its Fourier transform will therefore have a heavy, reciprocal polynomial tail in ω for any choices of the Gaussian width Δ and the jump rate $1/\tau_c$.

Appendix C

Minimum Power Dissipation

In this appendix, we will show the connection between minimum power dissipation and the Kirchoff circuit laws through the example of a simple linear circuit. The connections we discuss should be true for arbitrary linear circuits in steady-state though we will not prove them at that level of generality. In the final section of this appendix, we will also comment on the validity of minimum power dissipation for nonlinear systems.

C.1 Minimum power dissipation implies circuit laws

It makes sense to talk about the principle of minimum power dissipation in a circuit only if it is being driven. In an undriven circuit, the configuration of currents that minimizes the power dissipation is the all-zero current configuration. Therefore, we get meaningful results from the application of minimum power dissipation only if there are *constraints* associated with the minimization. We show next that if one minimizes the power dissipation subject to the constraint that the Kirchoff current law is satisfied, the resulting circuit steady-state configuration satisfies the Kirchoff voltage law. Conversely, one can also recover the Kirchoff current law by minimizing the power dissipation subject to the constraint that the Kirchoff voltage law is satisfied.

Minimization of power dissipation subject to the current law constraint yields the voltage law

Consider the circuit shown in Fig. C.1. There is a current source I_0 that is driving the two parallel RL branches, with the current in the left branch being I_1 and the current in the other branch being I_2. The power dissipation in the circuit is:

$$P(I_1, I_2) = I_1^2 R_1 + I_2^2 R_2 \tag{C.1}$$

The steady state of the circuit may be found by minimizing the power dissipation function with respect to I_1 and I_2 subject to the constraint that I_1 and I_2 respect the Kirchoff current

law in steady state. Therefore, the constrained optimization problem we have is:

$$\begin{aligned} &\text{minimize} && I_1^2 R_1 + I_2^2 R_2 \\ &\text{subject to} && (I_1, I_2) \in \{(I_1, I_2) \,|\, I_1 + I_2 = I_0\} \end{aligned}$$

We can solve this problem using Lagrange multipliers. The Lagrange function is:

$$L(I_1, I_2, \lambda) = I_1^2 R_1 + I_2^2 R_2 + \lambda\,(I_1 + I_2 - I_0) \tag{C.2}$$

Setting $\frac{\partial}{\partial I_1} L(I_1, I_2, \lambda) = 0$, $\frac{\partial}{\partial I_2} L(I_1, I_2, \lambda) = 0$, and $\frac{\partial}{\partial \lambda} L(I_1, I_2, \lambda) = 0$, we get:

$$\frac{\partial}{\partial I_1} L(I_1, I_2, \lambda) = 2 I_1 R_1 + \lambda = 0 \tag{C.3}$$

$$\frac{\partial}{\partial I_2} L(I_1, I_2, \lambda) = 2 I_2 R_2 + \lambda = 0 \tag{C.4}$$

$$\frac{\partial}{\partial \lambda} L(I_1, I_2, \lambda) = I_1 + I_2 - I_0 = 0 \tag{C.5}$$

The first two equations give us $I_1 R_1 = I_2 R_2$ which is the correct Kirchoff voltage law for the two parallel branches in steady state.

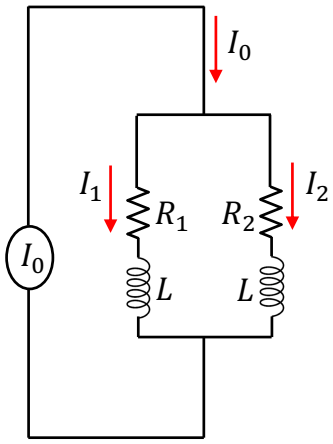

Figure C.1: Current source driving a parallel RL circuit. The Kirchoff current law and the principle of minimum power dissipation together imply the Kirchoff voltage law in steady state.

Minimization of power dissipation subject to the voltage law constraint yields the current law

Now, we consider the converse problem of deriving the current law by assuming the voltage law and minimum power dissipation. The circuit for this case is shown in Fig. C.2. There is

a voltage source V_0 that is driving the two parallel RL branches, with the current in the left branch being I_1 and the current in the other branch being I_2. R_s is the source resistance. The power dissipation in the circuit is:

$$P(I_0, I_1, I_2) = I_1^2 R_1 + I_2^2 R_2 + I_0^2 R_s \tag{C.6}$$

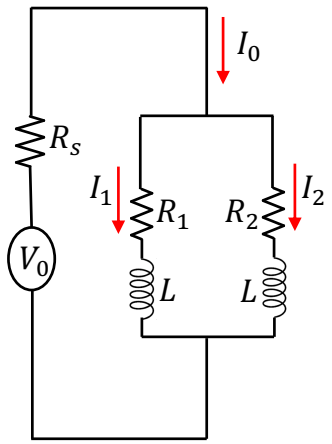

Figure C.2: Voltage source driving a parallel RL circuit. The Kirchoff voltage law and the principle of minimum power dissipation together imply the Kirchoff current law in steady state.

The steady state of the circuit may be found by minimizing the power dissipation function with respect to I_1 and I_2 subject to the constraint that I_0, I_1, and I_2 respect the Kirchoff voltage law in steady state. Therefore, the constrained optimization problem we have is:

$$\begin{aligned} &\text{minimize} && I_1^2 R_1 + I_2^2 R_2 + I_0^2 R_s \\ &\text{subject to} && (I_0, I_1, I_2) \in \{(I_0, I_1, I_2) \mid V_0 - I_0 R_s = I_1 R_1,\ V_0 - I_0 R_s = I_2 R_2\} \end{aligned}$$

We again solve this problem using Lagrange multipliers. The Lagrange function is:

$$L(I_0, I_1, I_2, \lambda_1, \lambda_2) = I_1^2 R_1 + I_2^2 R_2 + I_0^2 R_s + \lambda_1 (V_0 - I_0 R_s - I_1 R_1) + \lambda_2 (V_0 - I_0 R_s - I_2 R_2) \tag{C.7}$$

Setting all the partial derivatives to zero, we get:

$$\frac{\partial}{\partial I_0} L(I_0, I_1, I_2, \lambda_1, \lambda_2) = 2 I_0 R_s - \lambda_1 R_s - \lambda_2 R_s = 0 \tag{C.8}$$

$$\frac{\partial}{\partial I_1} L(I_0, I_1, I_2, \lambda_1, \lambda_2) = 2 I_1 R_1 - \lambda_1 R_1 = 0 \tag{C.9}$$

$$\frac{\partial}{\partial I_2} L(I_0, I_1, I_2, \lambda_1, \lambda_2) = 2 I_2 R_2 - \lambda_2 R_2 = 0 \tag{C.10}$$

$$\frac{\partial}{\partial \lambda_1} L(I_0, I_1, I_2, \lambda_1, \lambda_2) = V_0 - I_0 R_s - I_1 R_1 = 0 \tag{C.11}$$

$$\frac{\partial}{\partial \lambda_2} L(I_0, I_1, I_2, \lambda_1, \lambda_2) = V_0 - I_0 R_s - I_2 R_2 = 0 \tag{C.12}$$

The first three equations give us $I_0 = I_1 + I_2$ which is the correct Kirchoff current law for the two parallel branches.

C.2 Onsager's dissipation and the standard power dissipation

In Section 5.2 in the main text, we saw that Onsager recast the equations of linear systems in steady state as a minimization principle:

$$\boldsymbol{V} = \boldsymbol{R}\boldsymbol{J} \iff \boldsymbol{\nabla}_{\boldsymbol{J}} \left(\boldsymbol{J}^{\boldsymbol{T}} \boldsymbol{R} \boldsymbol{J} - 2 \boldsymbol{V}^{\boldsymbol{T}} \boldsymbol{J} \right) = \boldsymbol{\nabla}_{\boldsymbol{J}} H(\boldsymbol{J}) = 0 \tag{C.13}$$

The dissipation function $H(\boldsymbol{J})$, which we will call Onsager's dissipation in this section, is a bit strange because it is the sum of the power dissipated in the resistors and the negative of twice the power injected by the voltage sources. We will now discuss how Onsager's dissipation arises from the standard dissipation function $P(\boldsymbol{J})$.

We again consider the circuit in Fig. C.2. This time, let the inductors in the two branches have different values, L_1 in the left branch and L_2 in the right. The instantaneous Kirchoff current and voltage laws are:

$$\frac{d}{dt} I_1 = \frac{1}{L_1} \left(V_0 - I_1 R_s - I_2 R_s - I_1 R_1 \right) \tag{C.14}$$

$$\frac{d}{dt} I_2 = \frac{1}{L_2} \left(V_0 - I_1 R_s - I_2 R_s - I_2 R_2 \right) \tag{C.15}$$

It is clear that the expressions in the parentheses on the right-hand sides of the two equations above are proportional to the partial derivatives with respect to I_1 and I_2 of the following Onsager dissipation function:

$$H(I_1, I_2) = \left(I_1 + I_2 \right)^2 R_s + I_1^2 R_1 + I_2^2 R_2 - 2 V_0 \left(I_1 + I_2 \right) \tag{C.16}$$

At steady state, the equations of motion settle down at a minimum of $H(I_1, I_2)$. The true power dissipation function, on the other hand, is:

$$P(I_1, I_2) = \left(I_1 + I_2 \right)^2 R_s + I_1^2 R_1 + I_2^2 R_2 \tag{C.17}$$

How are the two related? To understand that, we will have to recall the concept of strong duality from the main text.

Lagrange duality

Let us say we have an optimization problem given by:

$$\begin{aligned} &\text{minimize} && f(\boldsymbol{x}) \\ &\text{subject to} && g_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, p. \end{aligned}$$

Let the optimal value of this problem be Z. The Lagrange function is:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}), \tag{C.18}$$

The optimization problem is said to be strongly dual if:

$$Z = \min_{\boldsymbol{x}} \left(\max_{\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right) = \max_{\boldsymbol{\lambda}} \left(\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right) \tag{C.19}$$

We will define the *dual function* $D(\boldsymbol{\lambda})$ as the function in the parenthesis in the right-most expression of the above equation:

$$D(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{C.20}$$

Onsager's dissipation is the negative of the dual function of the standard power dissipation

Let us consider the following minimization problem that we encountered in the previous section:

$$\begin{aligned} &\text{minimize} && I_1^2 R_1 + I_2^2 R_2 + I_0^2 R_s \\ &\text{subject to} && (I_0, I_1, I_2) \in \{(I_0, I_1, I_2) \mid V_0 - I_0 R_s = I_1 R_1,\ V_0 - I_0 R_s = I_2 R_2\} \end{aligned}$$

Next, we set $I_0 = I_1 + I_2$ in it:

$$\text{minimize} \quad I_1^2 R_1 + I_2^2 R_2 + (I_1 + I_2)^2 R_s \tag{C.21}$$

$$\text{subject to} \quad (I_1, I_2) \in \{(I_1, I_2) \mid V_0 - (I_1 + I_2) R_s = I_1 R_1,\ V_0 - (I_1 + I_2) R_s = I_2 R_2\} \tag{C.22}$$

The constraint set trivially contains only one point (I_1, I_2) because we incorporated both the current law and the voltage law into it. Nevertheless, it is a strongly dual problem so we will go ahead and compute its dual function $D(\lambda_1, \lambda_2)$. The Lagrange function is:

$$\begin{aligned} L(I_1, I_2, \lambda_1, \lambda_2) = & I_1^2 R_1 + I_2^2 R_2 + (I_1 + I_2)^2 R_s + \lambda_1 (V_0 - (I_1 + I_2) R_s - I_1 R_1) \\ & + \lambda_2 (V_0 - (I_1 + I_2) R_s - I_2 R_2) \end{aligned} \tag{C.23}$$

The Lagrange dual function $D(\lambda_1, \lambda_2)$ is defined as:

$$D(\lambda_1, \lambda_2) = \min_{I_1, I_2} L(I_1, I_2, \lambda_1, \lambda_2) \tag{C.24}$$

To find the I_1, I_2 that minimize L, we set the partial derivatives to 0:

$$\frac{\partial}{\partial I_1} L(I_1, I_2, \lambda_1, \lambda_2) = 2I_1R_1 - \lambda_1 R_1 + 2\left(I_1 + I_2\right) R_s - \left(\lambda_1 + \lambda_2\right) R_s = 0 \tag{C.25}$$

$$\frac{\partial}{\partial I_2} L(I_1, I_2, \lambda_1, \lambda_2) = 2I_2R_2 - \lambda_2 R_2 + 2\left(I_1 + I_2\right) R_s - \left(\lambda_1 + \lambda_2\right) R_s = 0 \tag{C.26}$$

From the above, we get $I_1 = \frac{1}{2}\lambda_1$ and $I_2 = \frac{1}{2}\lambda_2$, and the dual function works out to:

$$D(\lambda_1, \lambda_2) = \left(\lambda_1 + \lambda_2\right) V_0 - \frac{\left(\lambda_1 + \lambda_2\right)^2}{4} R_s - \frac{\lambda_1^2}{4} R_1 - \frac{\lambda_2^2}{4} R_2 \tag{C.27}$$

By duality, the power dissipation at the steady state, which is obtained by solving the original minimization problem Eqs. C.21 and C.22, can be obtained instead by maximizing Eq. C.27 over all λ_1, λ_2. The current configuration I_1, I_2 at steady state is obtained by taking the optimal λ_1^*, λ_2^* and multiplying them by 2. Instead of performing the 2-step procedure of maximizing $D(\lambda_1, \lambda_2)$ and multiplying the optimal λ_1^*, λ_2^* by 2, one could directly maximize the following function of I_1, I_2:

$$D'(I_1, I_2) = 2\left(I_1 + I_2\right) V_0 - \left(I_1 + I_2\right)^2 R_s - I_1^2 R_1 - I_2^2 R_2 \tag{C.28}$$

Maximizing the above function is equivalent to minimizing its negative:

$$-D'(I_1, I_2) = \left(I_1 + I_2\right)^2 R_s + I_1^2 R_1 + I_2^2 R_2 - 2\left(I_1 + I_2\right) V_0 = H(I_1, I_2) \tag{C.29}$$

This function is exactly the same as Onsager's dissipation function!

C.3 Minimum power dissipation in nonlinear circuits

The principle of minimum power dissipation, in its native form, seems to invalid in general nonlinear circuits. We study a simple nonlinear circuit in this section to understand this point. It is possible that a more generalized version of the principle of minimum power dissipation holds in nonlinear systems but we are not currently aware of such an extension.

Consider the circuit in Fig. C.3. The resistor R_2 in the right branch is nonlinear and is given by $R_2 = R_l + I_2^2 R_{nl}$. Let us assume the current law and try to obtain the steady state voltage law by minimizing the power dissipation. The constrained optimization problem we have is:

$$\begin{aligned} &\text{minimize} && I_1^2 R_1 + I_2^2 R_l + I_2^4 R_{nl} \\ &\text{subject to} && (I_1, I_2) \in \{(I_1, I_2) \mid I_1 + I_2 = I_0\} \end{aligned}$$

The Lagrange function is:

$$L(I_1, I_2, \lambda) = I_1^2 R_1 + I_2^2 R_l + I_2^4 R_{nl} + \lambda (I_1 + I_2 - I_0) \tag{C.30}$$

Setting the partial derivatives to 0, we get:

$$\frac{\partial}{\partial I_1} L(I_1, I_2, \lambda) = 2 I_1 R_1 + \lambda = 0 \tag{C.31}$$

$$\frac{\partial}{\partial I_2} L(I_1, I_2, \lambda) = 2 I_2 R_l + 4 I_2^3 R_{nl} + \lambda = 0 \tag{C.32}$$

$$\frac{\partial}{\partial \lambda} L(I_1, I_2, \lambda) = I_1 + I_2 - I_0 = 0 \tag{C.33}$$

The first two equations give us $I_1 R_1 = I_2 (R_l + 2 I_2^2 R_{nl})$ which is **not** the correct steady state Kirchoff voltage law for the two parallel branches. The correct voltage law is $I_1 R_1 = I_2 (R_l + I_2^2 R_{nl})$.

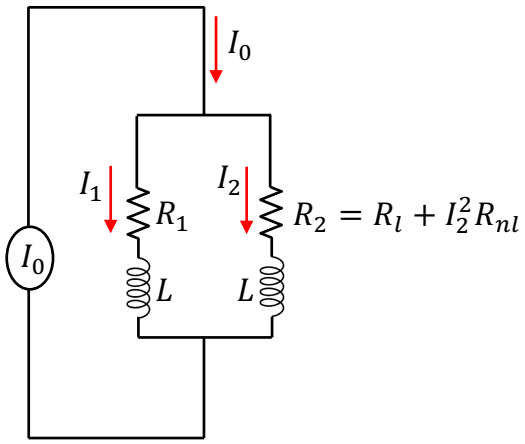

Figure C.3: Current source driving a parallel RL circuit with a nonlinear resistance in one branch. The Kirchoff current law and the principle of minimum power dissipation together do not yield the steady state Kirchoff voltage law.

Appendix D

Equations of motion derivation

In this appendix, we fill in the missing details of the derivations of the equations of motions for the coupled LC oscillator system. The first section deals with the equations for the main coupled oscillator system ('signal circuit') while the second section derives the interaction between each LC oscillator and its associated pump circuit.

D.1 Signal circuit

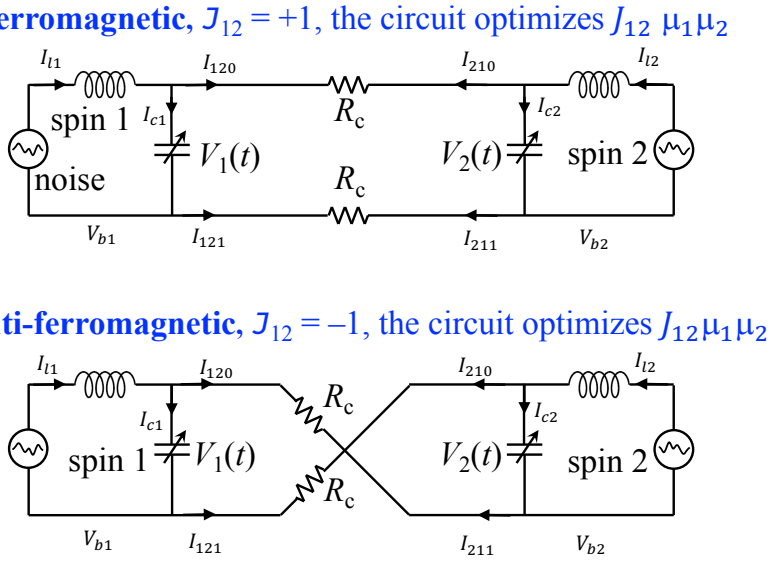

Figure D.1: Coupled LC oscillator circuit for two coupled magnets. The oscillation of the LC oscillators represents the magnetic moments, while the parallel or antiparallel cross-connections represent ferromagnetic $J_{ij} = 1$ or antiferromagnetic $J_{ij} = -1$ coupling, respectively. The nonlinear capacitors are pumped by $V(2\omega_0)$ at frequency $2\omega_0$, providing parametric gain at ω_0.

The circuit equations we had compiled in the main text are reproduced here:

$$i_{ci} = f(v_i, \dot{v_i}), \text{ for all } i \in \{1, 2, \ldots, N\} \tag{D.1}$$

$$v_i = -L\dot{i_{li}}, \text{ for all } i \in \{1, 2, \ldots, N\} \tag{D.2}$$

$$\sum_j i_{ij0} = i_{li} - i_{ci}, \text{ for all } i \in \{1, 2, \ldots, N\} \tag{D.3}$$

$$\sum_j i_{ij0} + \sum_j i_{ij1} = 0, \text{ for all } i \in \{1, 2, \ldots, N\} \tag{D.4}$$

$$v_{b1} = 0 \tag{D.5}$$

$$\begin{aligned} i_{ijk} = {} & k\left[\left(\frac{v_{bi} - v_{bj}}{R}\right)\left(\frac{1 + J_{ij}}{2}\right) + \left(\frac{v_{bi} - v_{bj} - v_j}{R}\right)\left(\frac{1 - J_{ij}}{2}\right)\right] \\ & + (1 - k)\left[\left(\frac{v_{bi} + v_i - v_{bj} - v_j}{R}\right)\left(\frac{1 + J_{ij}}{2}\right) + \left(\frac{v_{bi} + v_i - v_{bj}}{R}\right)\left(\frac{1 - J_{ij}}{2}\right)\right], \\ & \text{for all } i \in \{1, 2, \ldots, N\}, \text{ for all } j \in \{1, 2, \ldots, N\},\ j \neq i, \text{ for all } k \in \{0, 1\} \end{aligned} \tag{D.6}$$

Plugging Eq. (D.6) into Eq. (D.3), we get:

$$i_{li} - i_{ci} = \sum_{j:j\neq i}\left[\frac{v_{bi} - v_{bj} + v_i}{R} - \frac{v_j}{R}\left(\frac{1 + J_{ij}}{2}\right)\right] \tag{D.7}$$

Next, plugging Eq. (D.6) into Eq. (D.4), we get:

$$\sum_{j:j\neq i}\left[\frac{v_{bi} - v_{bj} + v_i}{R} - \frac{v_j}{R}\left(\frac{1 + J_{ij}}{2}\right)\right] = -\sum_{j:j\neq i}\left[\frac{v_{bi} - v_{bj}}{R} - \frac{v_j}{R}\left(\frac{1 - J_{ij}}{2}\right)\right] \tag{D.8}$$

which leads to:

$$\sum_{j:j\neq i} 2\left(\frac{v_{bi} - v_{bj}}{R}\right) = -\frac{(N-1)\, v_i}{R} + \sum_{j:j\neq i}\frac{v_j}{R} \tag{D.9}$$

Finally, substituting Eq. (D.9) into Eq. (D.7), we get:

$$2R\,(i_{li} - i_{ci}) = (N-1)\, v_i - \sum_{j:j\neq i} J_{ij} v_j \tag{D.10}$$

which upon differentiation leads to the following final equation in vector form:

$$(J - (N-1)\, I)\,\dot{v} = 2R\left(\dot{f} + \frac{v}{L}\right) \tag{D.11}$$

Slowly-varying amplitude approximation

We start with:

$$\begin{aligned}
&\left[\frac{J-(N-1)\,I}{2R_cC_0}+4\omega_0\frac{\Delta C}{C_0}\sin\left(2\omega_0 t\right)\right]\\
&\left(\left(\dot{A}+B\omega_0\right)\cos(\omega_0 t+\phi)+\left(\dot{B}-A\omega_0\right)\sin(\omega_0 t+\phi)\right)\\
=&\left(1+\frac{\Delta C}{C_0}\cos\left(2\omega_0 t\right)\right)\left(\left(-2\dot{A}\omega_0-B\omega_0^2\right)\sin(\omega_0 t+\phi)+\left(2\dot{B}\omega_0-A\omega_0^2\right)\cos(\omega_0 t+\phi)\right)\\
&+\left(\frac{1}{LC_0}-4\omega_0^2\frac{\Delta C}{C_0}\cos\left(2\omega_0 t\right)\right)\left(A\cos(\omega_0 t+\phi)+B\sin(\omega_0 t+\phi)\right)
\end{aligned} \tag{D.12}$$

For notational convenience, we set $J'=\frac{J-(N-1)I}{2R_cC_0}$ and $g'=\omega_0\frac{\Delta C}{C_0}$. Comparing the coefficient of $\cos\left(\omega_0 t\right)$ on both sides, we have:

$$\begin{aligned}
J'\left[\left(\dot{A}+B\omega_0\right)\cos\phi+\left(\dot{B}-A\omega_0\right)\sin\phi\right]=&\left(2\omega_0-g'\right)\left(\dot{B}\cos\phi-\dot{A}\sin\phi\right)\\
&-\frac{g'}{2}\omega_0\left(A\cos\phi+B\sin\phi\right)
\end{aligned} \tag{D.13}$$

Similarly, comparing the coefficient of $\sin\left(\omega_0 t\right)$ on both sides, we have:

$$\begin{aligned}
J'\left[-\left(\dot{A}+B\omega_0\right)\sin\phi+\left(\dot{B}-A\omega_0\right)\cos\phi\right]=&-\left(2\omega_0+g'\right)\left(\dot{B}\sin\phi+\dot{A}\cos\phi\right)\\
&-\frac{g'}{2}\omega_0\left(A\sin\phi-B\cos\phi\right)
\end{aligned} \tag{D.14}$$

Taking linear combinations of the above, we get:

$$J'\left(\dot{A}+B\omega_0\right)=2\omega_0\dot{B}-g'\dot{B}\cos 2\phi+g'\dot{A}\sin 2\phi-\frac{g'}{2}\omega_0\left(A\cos 2\phi+B\sin 2\phi\right) \tag{D.15}$$

$$J'\left(\dot{B}-A\omega_0\right)=-2\omega_0\dot{A}-g'\dot{B}\sin 2\phi-g'\dot{A}\cos 2\phi-\frac{g'}{2}\omega_0\left(A\sin 2\phi-B\cos 2\phi\right) \tag{D.16}$$

Writing it in matrix form, we get:

$$\begin{pmatrix} J'-g'\sin 2\phi & g'\cos 2\phi-2\omega_0\\ g'\cos 2\phi+2\omega_0 & J'+g'\sin 2\phi\end{pmatrix}\begin{pmatrix}\dot{A}\\ \dot{B}\end{pmatrix}=\begin{pmatrix}-\frac{g'}{2}\omega_0\cos 2\phi & -\frac{g'}{2}\omega_0\sin 2\phi-J'\omega_0\\ -\frac{g'}{2}\omega_0\sin 2\phi+J'\omega_0 & \frac{g'}{2}\omega_0\cos 2\phi\end{pmatrix}\begin{pmatrix}A\\ B\end{pmatrix} \tag{D.17}$$

J' contains the resistance value and is essentially the loss experienced by the system whereas g' is the parametric gain supplied to the system. They together decide the rate of growth of the amplitudes $A(t)$ and $B(t)$ as Eq. (D.17) indicates. We divide by ω_0 on both sides and drop terms of the order of J'/ω_0 or g'/ω_0. J'/ω_0 and g'/ω_0 are intentionally chosen

to be much smaller than 1 during the design as they directly control the evolution of the amplitudes $A(t)$ and $B(t)$. For the slowly-varying amplitude approximation to hold, the envelope, which evolves according to J' and g', has to change at a slower rate than the 'carrier' frequency of ω_0. Doing this (and the switching the order of the equations), we get:

$$\dot{A} = \left(\frac{J'}{2} - \frac{g'}{4}\sin 2\phi\right) A + \frac{g'}{4}\cos 2\phi B \tag{D.18}$$

$$\dot{B} = \left(\frac{J'}{2} + \frac{g'}{4}\sin 2\phi\right) B + \frac{g'}{4}\cos 2\phi A \tag{D.19}$$

Since our quadrature basis, $\cos(\omega_0 t + \phi)$ and $\sin(\omega_0 t + \phi)$, was very general, we can choose any particular basis we are interested in by picking ϕ. Setting $\phi = 3\pi/4$, and resubstituting J' and g', we have:

$$\dot{A} = \left(\frac{J - (N-1)I}{4R_c C_0} + \frac{\omega_0 \Delta C}{4C_0}\right) A \tag{D.20}$$

$$\dot{B} = \left(\frac{J - (N-1)I}{4R_c C_0} - \frac{\omega_0 \Delta C}{4C_0}\right) B \tag{D.21}$$

Both the cosine and the sine components evolve exponentially from their starting conditions. Since all the eigenvalues of the loss matrix $(J - (N-1)I)/(4R_c C_0)$ are negative, the cosine component A grows if there is enough gain $g = \omega_0 \Delta C/(4C_0)$ while the sine component decays always, even at zero gain.

D.2 Power dissipation

To obtain the power dissipation in the coupling resistors, we need to find the potentials at the top and bottom terminals of all the oscillators. This can be done by using Eq. D.9. We have:

$$\sum_{j:j\neq i} 2\left(\frac{v_{bi} - v_{bj}}{R}\right) + \frac{(N-1)v_i}{R} - \sum_{j:j\neq i}\frac{v_j}{R} = 0 \tag{D.22}$$

$$\implies (N-1)(2v_{bi} + v_i) - \sum_{j:j\neq i}(2v_{bj} + v_j) = 0 \tag{D.23}$$

In matrix form, this equation is:

$$\begin{pmatrix} N-1 & -1 & \dots & -1 \\ -1 & N-1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & N-1 \end{pmatrix} \begin{pmatrix} 2v_{b1} + v_1 \\ 2v_{b2} + v_2 \\ \vdots \\ 2v_{bN} + v_N \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{D.24}$$

The matrix on the left can be written as $NI - \mathbf{1}$ where $\mathbf{1}$ is the matrix with all ones. Since $\mathbf{1}$ is a rank one matrix with N as its only non-zero eigenvalue, $NI - \mathbf{1}$ has rank $N - 1$. The only vector in the nullspace of $NI - \mathbf{1}$ is the all ones vector. Therefore, the vector $2v_b + v$ has to be a multiple of the all ones vector which means that all its components are equal. Using $v_{b1} = 0$, we have for $i \neq 1$:

$$v_{bi} = \frac{v_1 - v_i}{2} \tag{D.25}$$

Then, the power dissipation in the coupling resistors that implements J_{ij} between the i-th and j-th oscillators is:

$$P = \begin{cases} 2\left(\frac{v_i - v_j}{2}\right)^2 \frac{1}{R_c} = \frac{v_i^2 + v_j^2 - 2v_i v_j}{2R_c}, & J_{ij} = 1 \\ 2\left(\frac{v_i + v_j}{2}\right)^2 \frac{1}{R_c} = \frac{v_i^2 + v_j^2 + 2v_i v_j}{2R_c}, & J_{ij} = -1 \end{cases} \tag{D.26}$$

The total power dissipation is then:

$$P(t) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\left(\frac{v_i^2(t) + v_j^2(t) - 2J_{ij}v_i(t)v_j(t)}{2R_c}\right) = v^T(t)\left(\frac{(N-1)\,I - J}{2R_c}\right)v(t) \tag{D.27}$$

Plugging in $v(t) = A(t)\cos\left(\omega_0 t + \frac{3\pi}{4}\right)$ and averaging over the cycle time period of $2\pi/\omega_0$ on both sides, we see that the time-averaged power can be expressed in terms of the slowly-varying amplitude $A(t)$ as follows:

$$\langle P \rangle = A^T(t)\left(\frac{(N-1)\,I - J}{4R_c}\right)A(t) \tag{D.28}$$

D.3 Pump circuit

The pump circuit equations are:

$$C_1\dot{V}_s = I_3 - I_4, \qquad C_2\dot{V}_p = I_2 - I_3, \tag{D.29}$$

$$V_s = L_1\left(\dot{I}_4 - \dot{I}_5\right), \qquad V_p = L_2\left(\dot{I}_1 - \dot{I}_2\right), \tag{D.30}$$

$$V_s = R_1\left(I_s + I_5\right), \qquad V_p = R_2\left(I_p - I_1\right) \tag{D.31}$$

$$I_3 = C_0\left(\dot{V}_p - \dot{V}_s\right) + 2C_N\left(V_p - V_s\right)\left(\dot{V}_p - \dot{V}_s\right) \tag{D.32}$$

I_3 can be eliminated by substituting the expression in the last line into the equations on the top line. Using the equations (D.29) and (D.31), we can express I_1, I_2, I_4, and I_5 in terms of voltages and the current sources. Finally, we plug all those expressions into

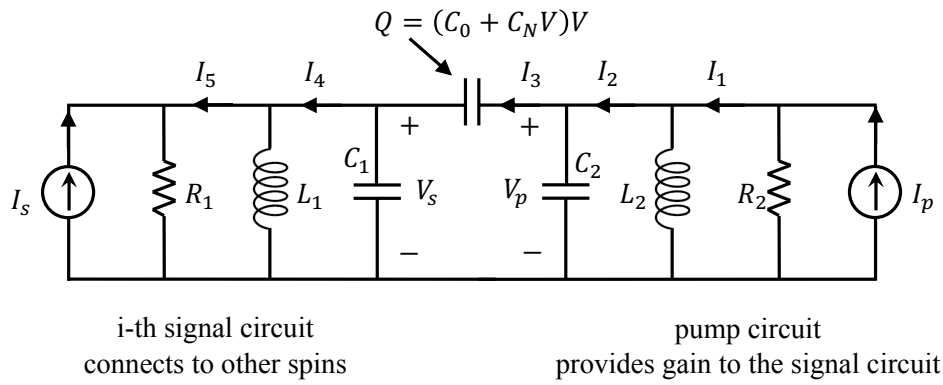

Figure D.2: The i-th 'signal' oscillator, on the left, shown along with its pumping oscillator, on the right. The signal oscillator and the pump oscillator are coupled via a nonlinear capacitor, shown at the top. The terminals of the left signal capacitor C_1 are coupled to the other capacitors via the parallel and antiparallel coupling scheme.

equations (D.30) to get:

$$V_s = L_1\left(C_0\left(\ddot{V}_p - \ddot{V}_s\right) + 2C_N\left(\dot{V}_p - \dot{V}_s\right)^2 + 2C_N\left(V_p - V_s\right)\left(\ddot{V}_p - \ddot{V}_s\right) - C_1\ddot{V}_s - \frac{\dot{V}_s}{R_1} + \dot{I}_s\right) \tag{D.33}$$

$$V_p = L_2\left(-C_0\left(\ddot{V}_p - \ddot{V}_s\right) - 2C_N\left(\dot{V}_p - \dot{V}_s\right)^2 - 2C_N\left(V_p - V_s\right)\left(\ddot{V}_p - \ddot{V}_s\right) - C_2\ddot{V}_p - \frac{\dot{V}_p}{R_2} + \dot{I}_p\right) \tag{D.34}$$

Next, we perform the slowly-varying amplitude approximation by using the following expressions for all the currents and voltages involved:

$$I_p = \frac{I_p}{2}e^{2i\omega_0 t} + \text{c.c} \qquad \dot{I}_p = 2i\omega_0\frac{I_p}{2}e^{2i\omega_0 t} + \text{c.c} \tag{D.35}$$

$$I_s = \frac{I_s}{2}e^{i(\omega_0 t+\phi_s)} + \text{c.c} \qquad \dot{I}_s = i\omega_0\frac{I_s}{2}e^{i(\omega_0 t+\phi_s)} + \text{c.c} \tag{D.36}$$

$$V_s = \frac{A_s}{2}e^{i(\omega_0 t+\phi_s)} + \text{c.c} \qquad \dot{V}_s = i\omega_0\frac{A_s}{2}e^{i(\omega_0 t+\phi_s)} + \frac{\dot{A}_s}{2}e^{i(\omega_0 t+\phi_s)} + \text{c.c} \tag{D.37}$$

$$\ddot{V}_s = 2i\omega_0\frac{\dot{A}_s}{2}e^{i(\omega_0 t+\phi_s)} - \omega_0^2\frac{A_s}{2}e^{i(\omega_0 t+\phi_s)} + \text{c.c} \tag{D.38}$$

$$V_p = \frac{A_p}{2}e^{i(2\omega_0 t)} + \text{c.c} \qquad \dot{V}_p = 2i\omega_0\frac{A_p}{2}e^{i(2\omega_0 t)} + \frac{\dot{A}_p}{2}e^{i(2\omega_0 t)} + \text{c.c} \tag{D.39}$$

$$\ddot{V}_p = 4i\omega_0\frac{\dot{A}_p}{2}e^{i(2\omega_0 t)} - 4\omega_0^2\frac{A_p}{2}e^{i(2\omega_0 t)} + \text{c.c} \tag{D.40}$$

All the slowly-varying amplitudes are assumed to be real, that is, we are already ignoring the 'sine' components and only considering the 'cosine' components of oscillation because they are the only ones that will eventually survive. In Eq. (D.33), we retain only terms that oscillate at ω_0 or contribute to oscillations at ω_0. Similarly, in Eq. (D.34) we retain only terms that oscillate at $2\omega_0$ or contribute to oscillations at $2\omega_0$. These equations simplify to:

$$\dot{I}_s = \frac{V_s}{L_1} + (C_0 + C_1)\,\ddot{V}_s + \frac{\dot{V}_s}{R_1} + 4C_N\dot{V}_p\dot{V}_s + 2C_N V_p\ddot{V}_s + 2C_N V_s\ddot{V}_p \tag{D.41}$$

$$\dot{I}_p = \frac{V_p}{L_2} + (C_0 + C_2)\,\ddot{V}_p + \frac{\dot{V}_p}{R_2} + 2C_N\dot{V}_s^2 + 2C_N V_s\ddot{V}_s \tag{D.42}$$

Plugging the slowly-varying expressions into the first of these equations, we get:

$$\begin{aligned} i\omega_0 I_s = {} & \frac{A_s}{L_1} + (C_0 + C_1)\left(2i\omega_0\dot{A}_s - \omega_0^2 A_s\right) + \frac{i\omega_0 A_s + \dot{A}_s}{R_1} \\ & + C_N\Big[2\left(2i\omega_0 A_p + \dot{A}_p\right)\left(-i\omega_0 A_s + \dot{A}_s\right) \\ & + A_p\left(-2i\omega_0\dot{A}_s - \omega_0^2 A_s\right) + A_s\left(4i\omega_0\dot{A}_p - 4\omega_0^2 A_p\right)\Big]\, e^{i(-2\phi_s)} \end{aligned} \tag{D.43}$$

Equating the imaginary parts on both sides and setting $\phi_s = 3\pi/4$, we have:

$$\omega_0 I_s = (C_0 + C_1)\,2\omega_0\dot{A}_s + \frac{\omega_0 A_s}{R_1} + C_N\left(2\dot{A}_s\dot{A}_p - \omega_0^2 A_s A_p\right) \tag{D.44}$$

In the last parenthesis, we ignore the first term because of the slowly-varying amplitude approximation leading to the following final equation:

$$\dot{A}_s = \frac{I_s}{2\,(C_0 + C_1)} - \frac{A_s}{2R_1\,(C_0 + C_1)} + \frac{C_N\omega_0 A_s A_p}{2\,(C_0 + C_1)} \tag{D.45}$$

This is how the 'cosine' component of the capacitor voltage in the signal circuit evolves in time. It has a noise term, I_s, the resistive decay term due to the internal resistance, and the final term is the gain which is directly proportional to A_p, the pump capacitor voltage. Performing an identical sequence of operations on the second equation (D.42), we get:

$$\dot{A}_p = \frac{I_p}{2\,(C_0 + C_2)} - \frac{A_p}{2R_2\,(C_0 + C_2)} - \frac{C_N\omega_0 A_s^2}{2\,(C_0 + C_2)} \tag{D.46}$$

This equation describes the evolution of the pump voltage A_p. The first two terms are again intuitive, but the last term is more important. It illustrates 'pump depletion', the fact that the pump loses its amplitude if the signal grows too much. When the pump falls, the gain it supplies to the signal gets reduced and the signal's amplitude gets saturated. It is not surprising that the equations for the signal and pump voltage amplitudes are exactly

isomorphic to the standard laser rate equations:

$$\frac{dS}{dt} = \Gamma\beta R_{sp} - \frac{S}{\tau_p} + \Gamma v_g g(N) S \tag{D.47}$$

$$\frac{dN}{dt} = \frac{\eta_i I}{qV} - \frac{N}{\tau} - v_g g(N) S \tag{D.48}$$

where N is the electron density, S is the photon density in the lasing mode, I is the injected current into the laser, η_i is the injection efficiency, βR_{sp} is the spontaneous emission rate into the lasing mode, $1/\tau$ is the electron population decay rate, $1/\tau_p$ is the cavity photon lifetime, v_g is the group velocity of the light, and $g(N)$ is the electron-density-dependent gain.

D.4 Augmented Lagrange equations of motion

For notational convenience, we set $J' = \frac{J-(N-1)I}{2R_c C_0}$ and $g' = \omega_0 \frac{\Delta C}{C_0}$. Comparing the coefficient of $\cos(\omega_0 t)$ on both sides, we have:

$$\begin{aligned}
&J'\left[\left(\dot{A} + B\omega_0\right)\cos\phi + \left(\dot{B} - A\omega_0\right)\sin\phi\right] \\
&- \frac{3G_N}{C_0}\cos\phi\left(\frac{\left(\dot{A} + B\omega_0\right)(3A^2 + B^2)}{4} + \frac{AB\dot{B} - A^2 B\omega_0}{2}\right) \\
&- \frac{3G_N}{C_0}\sin\phi\left(\frac{\left(\dot{B} - A\omega_0\right)(A^2 + 3B^2)}{4} + \frac{AB\dot{A} + AB^2\omega_0}{2}\right) = (2\omega_0 - g')\left(\dot{B}\cos\phi - \dot{A}\sin\phi\right) \\
&- \frac{g'}{2}\omega_0\left(A\cos\phi + B\sin\phi\right)
\end{aligned} \tag{D.49}$$

Similarly, comparing the coefficient of $\sin(\omega_0 t)$ on both sides, we have:

$$\begin{aligned}
&J'\left[-\left(\dot{A} + B\omega_0\right)\sin\phi + \left(\dot{B} - A\omega_0\right)\cos\phi\right] \\
&+ \frac{3G_N}{C_0}\sin\phi\left(\frac{\left(\dot{A} + B\omega_0\right)(3A^2 + B^2)}{4} + \frac{AB\dot{B} - A^2 B\omega_0}{2}\right) \\
&- \frac{3G_N}{C_0}\cos\phi\left(\frac{\left(\dot{B} - A\omega_0\right)(A^2 + 3B^2)}{4} + \frac{AB\dot{A} + AB^2\omega_0}{2}\right) \\
&= -(2\omega_0 + g')\left(\dot{B}\sin\phi + \dot{A}\cos\phi\right) - \frac{g'}{2}\omega_0\left(A\sin\phi - B\cos\phi\right)
\end{aligned} \tag{D.50}$$

Taking linear combinations of the above, we get:

$$J'\left(\dot{A}+B\omega_0\right)-\frac{3G_N}{C_0}\left(\frac{\left(\dot{A}+B\omega_0\right)\left(3A^2+B^2\right)}{4}+\frac{AB\dot{B}-A^2B\omega_0}{2}\right) \tag{D.51}$$
$$=2\omega_0\dot{B}-g'\dot{B}\cos 2\phi+g'\dot{A}\sin 2\phi-\frac{g'}{2}\omega_0\left(A\cos 2\phi+B\sin 2\phi\right)$$

$$J'\left(\dot{B}-A\omega_0\right)-\frac{3G_N}{C_0}\left(\frac{\left(\dot{B}-A\omega_0\right)\left(A^2+3B^2\right)}{4}+\frac{AB\dot{A}+AB^2\omega_0}{2}\right) \tag{D.52}$$
$$=-2\omega_0\dot{A}-g'\dot{B}\sin 2\phi-g'\dot{A}\cos 2\phi-\frac{g'}{2}\omega_0\left(A\sin 2\phi-B\cos 2\phi\right)$$

Simplifying the equations as before, we get:

$$\begin{pmatrix}\dot{B}\\ \dot{A}\end{pmatrix}-\frac{3G_N}{8C_0\omega_0}\begin{pmatrix}-\left(\dot{A}+B\omega_0\right)\left(3A^2+B^2\right)-2\left(AB\dot{B}-A^2B\omega_0\right)\\ \left(\dot{B}-A\omega_0\right)\left(A^2+3B^2\right)+2\left(AB\dot{A}+AB^2\omega_0\right)\end{pmatrix}=\begin{pmatrix}\left(\frac{J'}{2}-\frac{g'}{4}\right)B\\ \left(\frac{J'}{2}+\frac{g'}{4}\right)A\end{pmatrix} \tag{D.53}$$

We set $B=0$ and look only at the bottom equation:

$$\dot{A}=\left(\frac{J'}{2}+\frac{g'}{4}\right)A-\frac{3G_N}{8C_0}A^3 \tag{D.54}$$

D.5 Circuit for arbitrary real values in the Ising J matrix

So far, we have only considered J matrices in which all the entries were chosen from $\{-1,1\}$. In this section, we describe the modifications required to generalize the coupled LC oscillator circuit when the J values take on arbitrary real values expressed in binary form $\ldots b_2b_1b_0.b_{-1}b_{-2}\ldots$. We will let the sign of J_{ij}, positive or negative, be represented by s_{ij}. That is, we are going to set $J_{ij}=s_{ij}\left|J_{ij}\right|$. The circuit equations from before are reproduced here:

$$i_{ci}=f(v_i,\dot{v_i}),\ \text{for all } i\in\{1,2,\ldots,N\} \tag{D.55}$$
$$v_i=-L\dot{i_{li}},\ \text{for all } i\in\{1,2,\ldots,N\} \tag{D.56}$$
$$\sum_j i_{ij0}=i_{li}-i_{ci},\ \text{for all } i\in\{1,2,\ldots,N\} \tag{D.57}$$
$$\sum_j i_{ij0}+\sum_j i_{ij1}=0,\ \text{for all } i\in\{1,2,\ldots,N\} \tag{D.58}$$
$$v_{b1}=0 \tag{D.59}$$

$$\begin{aligned} i_{ijk} = k &\left[\left(\frac{v_{bi}-v_{bj}}{R_{ij}}\right)\left(\frac{1+s_{ij}}{2}\right)+\left(\frac{v_{bi}-v_{bj}-v_j}{R_{ij}}\right)\left(\frac{1-s_{ij}}{2}\right)\right] \\ &+(1-k)\left[\left(\frac{v_{bi}+v_i-v_{bj}-v_j}{R_{ij}}\right)\left(\frac{1+s_{ij}}{2}\right)+\left(\frac{v_{bi}+v_i-v_{bj}}{R_{ij}}\right)\left(\frac{1-s_{ij}}{2}\right)\right], \\ &\text{for all } i \in \{1,2,\ldots,N\},\ \text{for all } j \in \{1,2,\ldots,N\},\ j \neq i,\ \text{for all } k \in \{0,1\} \end{aligned} \tag{D.60}$$

Plugging Eq. (D.60) into Eq. (D.57), we get:

$$i_{li} - i_{ci} = \sum_{j:j\neq i}\left[\frac{v_{bi}-v_{bj}+v_i}{R_{ij}} - \frac{v_j}{R_{ij}}\left(\frac{1+s_{ij}}{2}\right)\right] \tag{D.61}$$

Next, plugging Eq. (D.60) into Eq. (D.58), we get:

$$\sum_{j:j\neq i}\left[\frac{v_{bi}-v_{bj}+v_i}{R_{ij}} - \frac{v_j}{R_{ij}}\left(\frac{1+s_{ij}}{2}\right)\right] = -\sum_{j:j\neq i}\left[\frac{v_{bi}-v_{bj}}{R_{ij}} - \frac{v_j}{R_{ij}}\left(\frac{1-s_{ij}}{2}\right)\right] \tag{D.62}$$

which leads to:

$$\sum_{j:j\neq i} 2\left(\frac{v_{bi}-v_{bj}}{R_{ij}}\right) = -\sum_{j:j\neq i}\frac{v_i}{R_{ij}} + \sum_{j:j\neq i}\frac{v_j}{R_{ij}} \tag{D.63}$$

In our circuit, we are going to use a reference resistor R_0, and binary multiples of it, $R_m = 2^m R_0$. That is, R_{-1} will be $R_0/2$ and R_2 will be $4R_0$. If $|J_{ij}|$ is written in binary form upto 3-bit precision as

$$|J_{ij}| = b_1 2^1 + b_0 2^0 + b_{-1}2^{-1}, \tag{D.64}$$

we can implement J_{ij} in our circuit by setting:

$$\frac{1}{R_{ij}} = \frac{b_1}{R_{-1}} + \frac{b_0}{R_0} + \frac{b_{-1}}{R_1} = \frac{1}{R_0}|J_{ij}| \tag{D.65}$$

To see that this setting indeed does the job, we plug this expression for R_{ij} into the preceding equations and substitute Eq. (D.63) into Eq. (D.61) to get:

$$2R_0\left(i_{li} - i_{ci}\right) = v_i \sum_{j:j\neq i}|J_{ij}| - \sum_{j:j\neq i} J_{ij} v_j \tag{D.66}$$

In the above equation, we used $J_{ij} = s_{ij}|J_{ij}|$. Differentiating the above leads to the following final equation in vector form:

$$(J - D)\,\dot{v} = 2R_0\left(\dot{f} + \frac{v}{L}\right) \tag{D.67}$$

where D is a diagonal matrix with the diagonal values given by $D_{ii} = \sum_{j:j\neq i}|J_{ij}|$.

Appendix E

Lagrange Multipliers

E.1 Lagrange multipliers theory

In this appendix, we shall study in greater detail the method of Lagrange multipliers, one of the most well-known techniques for the solution of constrained optimization problems. At its core, it simply involves the modification of the merit function by adding terms that penalize constraint violations. Much of the material here is drawn from [132] and [118].

In the setting of constrained optimization, we are required to minimize an objective function $f(\boldsymbol{x})$ with respect to all points that satisfy certain given inequality ($h_j(\boldsymbol{x}) \leq 0$) and equality ($g_i(\boldsymbol{x}) = 0$) constraints. Let us assume that the domain set D, that is, the intersection of the domains of $f(\boldsymbol{x}), g_i(\boldsymbol{x}), h_j(\boldsymbol{x})$, is a subset of $\mathbb{R}^n$. Further, let the set of points in D that satisfy all the constraints g_i, h_j be called F, the feasible set. Then, the optimization problem can be written as:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & h_j(\boldsymbol{x}) \leq 0, \ j = 1, \ldots, m, \\ & g_i(\boldsymbol{x}) = 0, \ i = 1, \ldots, p, \\ & \boldsymbol{x} \in D. \end{aligned}$$

Lagrange function

We now define a new function in $n+m+p$ variables, called the Lagrange function, as follows:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j h_j(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}).$$

The summations that were added to the plain objective function $f(\boldsymbol{x})$ serve as constraint violation penalty terms. The coefficients multiplying the penalty terms, λ_i and μ_j, are known as Lagrange multipliers. The inequality Lagrange multipliers μ_j are constrained to be non-negative in order that the penalty that arises when the inequality constraints are not satisfied

(i.e. when $h_j(\boldsymbol{x}) > 0$) is non-negative. The equality Lagrange multipliers, λ_i, have no such restrictions.

The Lagrange function has the advantage that it helps us express the constrained optimization of f as an unconstrained optimization of L. That is, it can be shown that:

$$\min_{\boldsymbol{x}\in F} f(\boldsymbol{x}) = \min_{\boldsymbol{x}\in D} \max_{(\boldsymbol{\mu}\geq 0),\boldsymbol{\lambda}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

The minimization over $\boldsymbol{x}$ in the feasible set, F, on the left-hand side has turned into a minimization over $\boldsymbol{x}$ in the entire domain, D, on the right-hand side.

Karush-Kuhn-Tucker (KKT) sufficient conditions

The Lagrange function also appears in an important, related role. While the conditions for a point $\boldsymbol{x}^*$ to be an unconstrained optimum of a differentiable function f are expressed in terms of the gradient of $f(\boldsymbol{x})$, the optimality conditions for constrained optimization problems are naturally expressed in terms of the Lagrange function. These conditions are presented next.

A point $\boldsymbol{x}^*$ is a local optimum of the function $f(\boldsymbol{x})$ subject to the constraints g_i, h_j if it satisfies the Karush-Kuhn-Tucker (KKT) sufficient conditions:

1. Primal feasibility: The point $\boldsymbol{x}^*$ is feasible, that is, it satisfies all the constraints.

$$h_j(\boldsymbol{x}^*) \leq 0,\ j = 1, \ldots, m,$$
$$g_i(\boldsymbol{x}^*) = 0,\ i = 1, \ldots, p.$$

2. First-order condition: There exist Lagrange multipliers $\mu_1^*, \ldots, \mu_m^*, \lambda_1^*, \ldots, \lambda_p^*$ such that the following equation is satisfied:

$$\boldsymbol{\nabla}_{\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j^* \boldsymbol{\nabla}_{\boldsymbol{x}} h_j(\boldsymbol{x}^*) + \sum_{i=1}^{p} \lambda_i^* \boldsymbol{\nabla}_{\boldsymbol{x}} g_i(\boldsymbol{x}^*) = 0.$$

3. Second-order condition: In addition to the first-order condition, if all the concerned functions are twice differentiable, we require that the Hessian of L with respect to $\boldsymbol{x}$ be positive definite along all directions that respect the active constraints. That is, the following equation has to be satisfied:

$$\boldsymbol{v}^T \boldsymbol{\nabla}^2_{\boldsymbol{xx}} L(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \boldsymbol{v} > 0,\ \forall \boldsymbol{v} \in T,$$

where T is defined as $T = \{\boldsymbol{v} :\ (\boldsymbol{\nabla}_{\boldsymbol{x}} h_j(\boldsymbol{x}^*))^T \boldsymbol{v} = 0,\ j = \text{active constraints at } \boldsymbol{x}^*,\ (\boldsymbol{\nabla}_{\boldsymbol{x}} g_i(\boldsymbol{x}^*))^T \boldsymbol{v} = 0,\ i = 1, \ldots, p\}$.

4. Complementary slackness: $\mu_j^* h_j(\boldsymbol{x}^*) = 0$ is satisfied for all the inequality constraints, $j = 1, \ldots, m$.

5. Dual feasibility: The Lagrange multipliers of all the inequality constraints satisfy $\mu_j^* \geq 0,\ j = 1, \ldots, m$, with the inequality being strict for active constraints.

E.2 Lagrange multipliers algorithms: Augmented Lagrangian

In this section, we discuss the so-called 'Augmented Lagrangian method of multipliers', a popular algorithm used to obtain locally optimal solutions $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ that satisfy the KKT conditions. We shall be only considering the case where there are no inequality constraints h_j (and, consequently, no $\boldsymbol{\mu}$ multipliers). This algorithm is discussed in detail in [118].

To motivate the 'Augmented Lagrangian' approach, let us observe the KKT conditions a bit closely. We conclude from the first-order condition that a locally optimum $\boldsymbol{x}^*$ renders the function $L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$ stationary. However, this doesn't guarantee that $\boldsymbol{x}^*$ is a minimum of $L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$. Indeed, observation of the second-order condition tells us that $\boldsymbol{x}^*$ could be a saddle point of $L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$. This means that gradient descent-based algorithms will not converge to $\boldsymbol{x}^*$ if the starting point is not in the correct region. It was to solve this problem that the 'Augmented Lagrangian' method was invented.

The Augmented Lagrange function, $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$, is given by:

$$L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i g_i(\boldsymbol{x}) + \frac{c}{2} \left(\sum_{i=1}^{p} \left(g_i(\boldsymbol{x}) \right)^2 \right).$$

This function $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$, for a suitable choice of c, has the property that $\boldsymbol{x}^*$ forms a minimum of $L_c(\boldsymbol{x}, \boldsymbol{\lambda}^*)$ and not just a saddle point as was the case with $L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$.

In other words, local optima of the original constrained optimization problem can be obtained if we perform an unconstrained optimization of $L_c(\boldsymbol{x}, \boldsymbol{\lambda}^*)$. However, for this procedure to be a feasible solution approach, we would have to know the right $\boldsymbol{\lambda}^*$. It has been shown in [118] that the way to do this is to perform gradient ascent of L_c in the $\boldsymbol{\lambda}$ variables.

The 'Augmented Lagrangian method of multipliers' involves the repeated minimization of $L_c(\boldsymbol{x}, \boldsymbol{\lambda}^{(k)})$ using progressively better estimates, $\boldsymbol{\lambda}^{(k)}$, of $\boldsymbol{\lambda}^*$. The algorithm starts off with an arbitrary starting point $(x^{(0)}, \lambda^{(0)})$. It then performs the following steps repeatedly:

1. Locally minimize $L_c(\boldsymbol{x}, \boldsymbol{\lambda}^{(k)})$ and call the minimum point $\boldsymbol{x}^{(k)}$.
2. $\lambda_i^{(k+1)} = \lambda_i^{(k)} + c g_i(\boldsymbol{x}^{(k)})$.

The second step above corresponds to gradient ascent of $L_c(\boldsymbol{x}, \boldsymbol{\lambda})$ in the $\boldsymbol{\lambda}$ variables. Basically, this method performs a *fast gradient descent* of L_c in the $\boldsymbol{x}$ directions in conjunction with a *slow gradient ascent* of L_c in the $\boldsymbol{\lambda}$ directions. A dynamical system that performs this process in continuous time is given below:

$$\begin{aligned} \frac{dx_i}{dt} &= -\kappa \frac{\partial}{\partial x_i} L_c(\boldsymbol{x}, \boldsymbol{\lambda}), \\ \frac{d\lambda_i}{dt} &= \kappa' \frac{\partial}{\partial \lambda_i} L_c(\boldsymbol{x}, \boldsymbol{\lambda}), \end{aligned}$$

where κ and κ' are suitable step sizes.

E.3 Application of Lagrange multipliers to the Ising problem; Cubic terms

In this section, we shall provide an explanation of the role that is played by the cubic terms—which were neglected in the main text—in the methods of Yamamoto et al. [93], Kalinin et al. [99], and others. It will turn out that the inclusion of cubic terms helps us implement the 'Augmented Lagrangian method of multipliers' from Supplementary Section 2.

The statement of the Ising problem is as follows:

$$\begin{aligned} &\text{minimize} && -\boldsymbol{x}^T \boldsymbol{J} \boldsymbol{x} \\ &\text{subject to} && x_i^2 - 1 = 0, \quad i = 1, \ldots, n. \end{aligned}$$

The corresponding Lagrange function is given by:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = -\boldsymbol{x}^T \boldsymbol{J} \boldsymbol{x} - \sum_{i=1}^{n} \lambda_i (x_i^2 - 1).$$

Next, we write down the Augmented Lagrange function:

$$L_c(\boldsymbol{x}, \boldsymbol{\lambda}) = -\boldsymbol{x}^T \boldsymbol{J} \boldsymbol{x} - \sum_{i=1}^{n} \lambda_i (x_i^2 - 1) + \frac{c}{2} \left(\sum_{i=1}^{n} \left(x_i^2 - 1 \right)^2 \right).$$

Substituting the above Augmented Lagrange function into the dynamical system provided at the end of Supplementary Section 2, we get:

$$\begin{aligned} \frac{dx_i}{dt} &= 2\kappa \left(\lambda_i x_i + 2c x_i - 2c x_i^3 + \sum_{j=1}^{n} J_{ij} x_j \right), \\ \frac{d\lambda_i}{dt} &= \kappa' (1 - x_i^2), \end{aligned}$$

where κ and κ' are appropriately chosen step sizes. We notice the equivalence in form between this dynamical system and those in the papers discussed in the main text and conclude that the cubic terms that appear in most of those systems are in fact helping to implement the Augmented Lagrangian method of multipliers.

E.4 Iterative Analog Matrix Multipliers

In this section, we present the system designed by Soljacic et al. [100]. We shall see that the simplified system we presented in Section 5.A of the main text differs from that of Soljacic et al. significantly in that their method has added noise and nonlinear thresholding after each iteration. It is possible that their modifications lead to performance improvements.

Formally, their iteration is given by:

$$\boldsymbol{E}(t+1) = u\left(2\boldsymbol{K}\boldsymbol{E}(t) + \boldsymbol{N}(t)\right),$$

where $\boldsymbol{E}(t)$ is the vector of electric field amplitude values at the beginning of the t-th iteration, $u(x)$ is the Heaviside step function that is 1 for positive x and 0 for negative x, $\boldsymbol{N}(t)$ is a zero-mean Gaussian random noise vector, and K is a matrix given by $K = \sqrt{\boldsymbol{J} + \alpha\boldsymbol{M}}$, $\boldsymbol{J}$ is the Ising connectivity matrix, α, a real number, and $\boldsymbol{M}$, some suitably chosen matrix. More specifically, $\boldsymbol{M}$ is chosen to have the same eigenvectors as $\boldsymbol{J}$. It will turn out that the eigenvalues of $\boldsymbol{M}$ play the role of Lagrange multipliers.

The authors showed that under the condition of high noise $\boldsymbol{N}(t)$, their system performs minimization of the following effective merit function:

$$H = -\frac{\beta}{2}\sum_{ij}\left(J_{ij} + \alpha M_{ij}\right)E_i E_j,$$

where β is some parameter dependent on the noise.

Using the fact that the matrix $\boldsymbol{M}$ is chosen to have the same eigenvectors as $\boldsymbol{J}$, we rewrite the above merit function, modulo additive constants, as the following Lagrange function:

$$H = L(\boldsymbol{E}, \boldsymbol{\gamma}) = -\frac{\beta}{2}\left(\sum_{ij} J_{ij}E_i E_j + \alpha\sum_i \gamma_i\left(z_i^2 - 1\right)\right),$$

where the γ_i are the eigenvalues of the matrix $\boldsymbol{M}$, and the vector $\boldsymbol{z}$ is the vector of electric field amplitudes $\boldsymbol{E}$ expressed in the basis of eigenvectors of $\boldsymbol{M}$ (that is, the eigenvectors of $\boldsymbol{J}$). We see that the eigenvalues of $\boldsymbol{M}$ play the role of Lagrange multipliers, albeit for different constraints than those required by the Ising problem. This difference is caused by $\boldsymbol{M}$ not being a diagonal matrix.

In conclusion, we interpret their algorithm as optimizing a Lagrange function with the merit function being the Ising Hamiltonian itself, and the constraints being that the components of the spin vector when expressed in the eigenvector basis of $\boldsymbol{J}$ be restricted to 1 and -1.