UNIVERSITY OF CALIFORNIA SAN DIEGO

Collective Phenomena in Memristive Networks: Engineering phase transitions into computation

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics

by

Forrest C. Sheldon

Committee in charge:

Professor Massimiliano Di Ventra, Chair
Professor Daniel Arovas
Professor Gert Cauwenberghs
Professor Clifford Surko
Professor Massimo Vergassola

2019

The dissertation of Forrest C. Sheldon is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

To Leo Szilard, the only physicist who seemed to like pastries as much as I do. And to my East and West coast families, especially Twig, Kendall, H and Savanna(h). Their confidence in me and aid has buoyed me through the most trying times and enabled me to pursue my interests fully. I would never have made it this far without them and cannot fully express by gratitude.

EPIGRAPH

Inspiration exists

but it must find you working

—Pablo Picasso

You know a lot of people go to college for 7 years.

I know, they're called doctors.

—Chis Farley and David Spade in Tommy Boy

There were mysterious questions,

but a mysterious answer was a contradiction in terms.

—Eliezer Yudkowsky via Harry James Potter Evans Veres

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I suspect all dissertations are a constant struggle between ballistic and diffusive progress. I would like to acknowledge those who pushed me towards the ballistic regime. First and foremost, my advisor and chair Max Di Ventra has been a patient advocate during my time at UCSD. He gave me an opportunity in the group, the most interesting project I could have hoped for and has supported my many enthusiastic diversions. For this he has my sincere thanks.

In the group, I would like to acknowledge the many helpful discussions and guidance from my senior members, Fabio Traversa, Sebastiano Peotta and Paul Boynton. In the same vein, the collaboration, enthusiasm and companionship of Haik Manukian, Sean Bearden and Rudy Pei have been invaluable. I enjoy pursuing science most in collaboration with others, and much of this dissertation would not have been possible without their help.

Chapter 2, in full, is a reprint of the material as it appears in, Physical Review E Volume 95. Sheldon, Forrest; Di Ventra, Massimiliano, APS, 2017. ©1997 American Physical Society. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is a reprint of the material which was submitted for publication as, Sheldon, Forrest; Cicotti, Pietro; Traversa, Fabio L.; Di Ventra, Massimiliano. Stress testing memcomputing on hard combinatorial optimization problems. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material which was submitted for publication as, Sheldon, Forrest; Traversa, Fabio L.; Di Ventra, Massimiliano. Taming a non-convex landscape with dynamical long-range order: Memcomputing the Ising spin-glass. The dissertation author was the primary investigator and author of this paper.

Appendix A, in full, is a reprint of the material as it appears in, European Physical Journal Applied Physics Volume 81. Sheldon, Forrest; Peotta, Sebastiano; Di Ventra, Massimiliano, and is included with the kind permission of The European Physical Journal (EPJ). The dissertation author was the primary investigator and author of this paper.

VITA

2009	B.S. in Physics *cum laude*, Duke University, Durham, NC
2009-20011	Research Assistant, Petrella Lab, Duke Hospital Radiology, Durham, NC
2010-2011	Research Assistant, Triangle Universities Nuclear Laboratory, Durham, NC
2011-2015	Graduate Teaching/Demonstration Assistant, University of California San Diego, La Jolla, CA
2014-2018	Di Ventra Lab, University of California San Diego, La Jolla, CA
2019	Ph.D. in Physics, University of California San Diego, La Jolla, CA

PUBLICATIONS

FC Sheldon, FL Traversa and M Di Ventra, "Taming a non-convex landscape with dynamical long-range order: memcomputing the ising spin glass", *In Submission*, 2018.

FC Sheldon, P Ciccoti, FL Traversa and M Di Ventra, "Stress-testing memcomputing on hard combinatorial optimization problems", *arXiv preprint arXiv:1807.00107*, 2018.

FL Traversa, P Ciccoti, FC Sheldon and M Di Ventra, "Evidence of exponential speed-up in the solution of hard optimization problems", *Complexity*, 2018.

FC Sheldon, S Peotta, and M Di Ventra, "Phase-dependent noise in Josephson junctions", *Eur. Phys. J. Appl. Phys.*, 81, 2018.

FC Sheldon and M Di Ventra, "Conducting-insulating transition in adiabatic memristive networks", *Phys. Rev. E*, 95, 2017.

ABSTRACT OF THE DISSERTATION

Collective Phenomena in Memristive Networks: Engineering phase transitions into computation

by

Forrest C. Sheldon

Doctor of Philosophy in Physics

University of California San Diego, 2019

Professor Massimiliano Di Ventra, Chair

This dissertation aims to address the dual goals of (1) proposing practical computing devices that meet a growing need for alternatives to von Neumann architecture, and (2) leveraging these to build new connections between computation and physics. As an avenue towards the former, I have focused on dynamical systems inspired by the digital memcomputing machines (DMMs) proposed by M. Di Ventra and F. Traversa. These are continuous dynamical systems, embeddable directly in hardware, which can be utilized to solve discrete combinatorial problems. Through several benchmarking studies, we have established that this approach to combinatorial optimization can outperform standard algorithmic methods, both in time to solution, and in scaling.

In many cases we are able to show that the time to solution of a digital memcomputing machine scales polynomially in instance size when other approaches scale exponentially. This indicates that the unique niche occupied by DMMs as deterministic, continuous dynamical systems solvers, possesses features not present in traditional approaches. As an avenue towards uncovering these features, we have worked to construct simplified models of DMMs that would facilitate theoretical work to rigorously establish their capabilities. In navigating their configuration space, these devices manifest a form of dynamical long-range order in which widely separated variables transition between states collectively. The structure of the equations borrowed from a DMM combined with the 'heuristic' requirement of long-range order allows us to construct a set of equations that reproduce several features of the DMM dynamics and that are able to solve a set of problems derived from spin-glasses. The form of these equations clarifies the essential role of continuity in the dynamics of the solver and the role of memory in DMMs.

Chapter 1

The Menu: From Memristors to Optimization

The purpose of this chapter is to give a bird's eye view of the work I have undertaken under the supervision of Prof. Massimiliano Di Ventra at UCSD. The central player in this is the memristor, and much of this chapter will be devoted to developing the notion of a memristor from a circuit element based on resistive switching, to a controlled way of incorporating non-linearity and memory in a dynamical system. Subsequent chapters are composed of papers we have written and which are either published or in submission.

The nonlinearity introduced by memristors can lead to novel behaviors when these devices are incorporated into networks. In particular, current sharing amongst the memristors can lead to collective switching. As a first approach to this topic, we investigate the presence of collective behavior in networks of memristive elements [SDV17].

Our primary interest is in leveraging this behavior in circuits that perform a useful computation. Similar effects have been implicated in the dynamics of *digital memcomputing machines* (DMMs) and are thought to play a role in accelerating their convergence to the solution of a computational problem. Before examining the dynamics of these types of solvers, we first

examine their performance on a class of combinatorial optimization problems known as *maximum satisfiablility* (MAXSAT) which aims to find a configuration of boolean variables satisfying the maximum number of constraints. While this is somewhat distinct from their original, intended use on satisfiable problems, we show that they are very effective on a wide variety of benchmarks drawn from the yearly MAXSAT competitions and that they display a marked advantage in their scaling when producing approximate solutions [TCSD18, SCTD].

As a starting point for building a theoretical characterization of DMMs and the role of collective behavior in their dynamics, we attempt to generalize their construction to find the ground states of the Ising spin-glass. These systems have simple and homogeneous constraints and are typically embedded on a lattice, facilitating the analysis of collective behavior. They are also the native instance form for the quantum annealers manufactured by D-Wave [HJA$^+$15, DBI$^+$16] making the performance of classical solvers a topic of current interest for determining quantum speedup.

The DMM approach shows a substantial improvement in the scaling of solution time over local annealing approaches and one combinatorial solver, scaling as $N^{1.5}$ while the other solvers tested scaled exponentially [STD]. Borrowing the general structure of equations present in DMMs, we construct a simplified model capable of solving small instances and that proceeds through a transient phase of avalanches, similar to those shown in memristive networks. The avalanches derive from terms analogous to an elastic potential energy that enforce a form of rigidity on clusters of spins within the instance. This is closely connected to the continuity of the dynamics and gives further evidence that features of continuous dynamical systems solvers can provide unique benefits in solving natively discrete problems.

1.1 Resistive Switching

Examples of materials or systems that display resistive-switching (RS) stretch back to the 19th century with Sir Humphrey Davy's studies of carbon arc lamps and Branley's coherer [CC18]. In the first case, a potential difference is applied between two electrodes immersed in an inert gas [PTC12]. For a sufficiently large potential difference the gas is ionized leading to a sudden increase in the current and thus the conductivity. Once ignited, the potential difference may then be decreased while maintaining the ionized 'arc' of gas between the electrodes. In the coherer [FC05], fine metal shavings which are initially disconnected will assemble into a conductive pathway between electrodes when subjected to a sufficiently large electric field, greatly increasing the conductivity. Once assembled the shavings would remain connected until agitated to 'decohere' the device.

These systems exemplify several prominent features that carry over to the nanoscale [Di 08, PD11, HTTA12]: they involve the flow of ions rather than only electrons, an external field enacts a geometrical rearrangement of the system, changes in the conductivity are observed only for a sufficiently large applied field or current, and the effects are sufficiently non-volatile to display hysteresis when the field/current is cycled.

As a canonical example, we consider the $Ag|Ag_2S|Ag$ metal-insulator-metal (MIM) atomic switches [THNA05, HTTA12] examined in Chapter 2. These systems consist of a thin (50 nm) layer of insulating silver sulfide (Ag_2S) between two metallic layers (usually either $Ag-Ag$ or $Ag-Pt$). In the absence of an electric field, the Ag_2S layer exists in its semiconducting α acanthite phase with a 1.2 eV band-gap and low diffusivity of Ag cations [XBW$^+$10]. This layer undergoes an electric field induced transition to a β argentite phase in which both the conductivity, and diffusivity of Ag^+ ions is greatly enhanced. This forms a conducting filament along which Ag cations migrate, eventually forming a bridge of pure silver between the metal interfaces, resulting in an increase of several orders of magnitude in the conductivity between the

metal layers [KTHA06] (the conductivity of Ag and β-argentite Ag_2S are $6.3 \times 10^5\, 1/\Omega/\text{cm}$ and $1.6 \times 10^3\, 1/\Omega/\text{cm}$ respectively versus $2.5 \times 10^{-3}\, 1/\Omega/\text{cm}$ for the α-acanthite phase [XBW$^+$10]).

This corresponds to an initial formation step required in RS materials. After this step, the devices may display either unipolar or bipolar resistive switching, depending on the physics of the interfaces [HTTA12]. In unipolar switching, after the formation step the conductivity may be reset by increasing the voltage in the same direction used during forming, effectively 'burning out' the filament. In bipolar RS, a reversed bias is applied in order to dissolve part of the filament and break the conductive pathway. The $Ag|Ag_2S|Ag$ interfaces discussed above display bipolar switching. A reversed bias ionizes Ag at the thinnest part of the filament through Joule assisted heating. These ions flow away from the filament following their concentration gradients and cause the filament to dissolve back into the substrate.

This is meant to illustrate the range of physical processes at work within a material undergoing resistive switching. The notion that a medium could show memory effects in its conductivity should perhaps be regarded as the norm rather than an exception; in general, the instantaneous response of a material will depend on its current state [DVP13a]. If, however the timescale in which the material reaches steady state is short compared to the timescale of interest, the material will display an Ohmic, or perhaps nonlinear conductance. If instead, the system possesses degrees of freedom with slow timescales relative to the timescale of interest, we will observe a time dependence and thus memory effects in its response function. At the nanoscale, the presence of ionic conduction supplies timescales slow enough for us to observe memory effects leading to resistive switching.

1.2 The Memristor

Historically, the memristor is conceptually distinct from the phenomenon of resistive-switching, although modern usage often confuses the two. The evolution of the ideas surrounding

this is lively enough that I've chosen to supply a brief historical survey of the memristor and its various interpretations.

The notion of the memristor as an 'ideal' circuit element was first posed by Chua in 1971 [Chu71]. Reasoning by analogy with the fundamental circuit components, he supposed a device characterized by a flux-charge curve, $g(\phi, q) = 0$ which can be written as the relation,

$$d\phi = M(q)dq, \quad M(q) = \frac{d\phi(q)}{dq} \tag{1.1}$$

where $M(q)$ has units of ohms and called the memristance. As $dq = i\,dt$ and $d\phi = v\,dt$ we must have

$$v = M(q)i.$$

If $M(q)$ is constant this degenerates into a linear resistor. A dependence on the charge $q(t)$ which has passed through the element implies a memory of the past history in contrast to a nonlinear resistor $R(i)$ which depends only on the instantaneous current and is characterized by a relationship $f(v,i) = 0$. The memory suggested by this construction lead to the portmanteau 'memristor.' A construction was given to simulate such a device using active elements and the question was left open whether a passive device, satisfying $M(q) > 0$ could exist. An identical construction in term of a dependence on the flux, $M(\phi)$ leads to a variant known as a 'flux-controlled' rather than 'charge-controlled' memristor.

The requirement that M depend only on the charge/flux is quite stringent. In 1976, Chua and Kang published a generalization of the memristor which they called a memristive system [CK76]. This is a dynamical element satisfying

$$y = g(x,u,t)u \tag{1.2}$$

$$\dot{x} = f(x,u,t) \tag{1.3}$$

where u and y are the input and output of the system and $x \in \mathbb{R}^n$ denotes its state. This definition relaxes the requirement that the device be derived from a $\phi - q$ curve and as a result is capable of modeling a wide range of physical systems forming a useful abstraction of the original concept. In this work they also placed a heavy emphasis on pinched hysteresis in $i - v$ curves of the device as a marker of memristive behavior.

Publications surrounding memristors were mostly quiescent until the 2008 paper by Strukov et al. proclaimed "The missing memristor found" [SSSW08], precipitating an explosion in interest and publications. The proposed devices were titantium-oxide films which exhibit resistive switching due to the migration of oxygen vacancies. As a model, they suggested

$$\begin{aligned} v(t) &= \left(R_{ON} \frac{w(t)}{D} + R_{OFF} \left(1 - \frac{w(t)}{D} \right) \right) i(t) \\ \frac{dw}{dt} &= \mu_V \frac{R_{ON}}{D} i(t) \end{aligned}$$

for a semiconductor of thickness D with resistivity $\frac{R_{OFF}}{D}$ containing a dopant layer of thickness $0 \leq w(t) \leq D$ with resistivity $\frac{R_{ON}}{D}$ and where the dopants have mobility μ_V. This admits the simplified form, for $R_{ON} \ll R_{OFF}$

$$M(q) = R_{OFF} \left(1 - \frac{\mu_V R_{ON}}{D^2} q(t) \right) \tag{1.4}$$

which is a marker of an ideal memristor.

In addition to interest, this generated substantial controversy [Abr18, VM15, DVP13a]. The device was not new as RS in semiconducting films was well known even before Chua's 1971 article [Arg68]. It was also clear that the model ignored several important effects in the physical devices (such as volatility) which would require classifying them as memristive systems. Additionally, scrutiny over the original definition called into question its consistency with Landauer's principle and the laws of electricity and magnetism [DVP13a], and whether its

characterization as 'fundamental' is deserved or an artifact of Chua's particular organization of the circuit elements [Abr18, VM15].

In light of the fact that the other circuit elements R, L, C are not 'ideally' realized in a physical device, aspects of this debate are difficult to understand. The notion of an ideal element is not intended to be exactly realized in nature but that we use it as a productive abstraction in modeling aspects of physical devices. In the course of this debate it was also pointed out that memory in a conductive system (or capacitive/inductive system), leading to the response function relationships (1.2) is a direct consequence of linear response theory [DVP13a] and should perhaps be considered the norm rather than an exception. While some of the success of the 2008 publication was thus undoubtably due to hype, its consequence has been the realization that memristive behavior is quite ubiquitous and the popularization of the memristor as useful abstraction apart from its physical realization.

We will make progressively heavier use of this abstraction, posing memristive systems with desirable equations and assuming that we can realize them with a suitably designed physical system. In this interest, we will review a few basic models for memristive systems that see regular usage. Having cleared my conscience with this historical aside, I will now follow the modern convention of bungling the names, using memristor and memristive system interchangably, but reserve the term 'ideal memristor' for one in which the resistance is purely a function of the charge/flux.

A unique memristor arises from the imaginary component of the superconducting response function of a Josephson junction. This can be isolated with a properly chosen SQUID junction [PDV14a]. While the dynamics of single physical memristors is somewhat tangential to the primary thread of this thesis, investigations of the noise properties of this system occupied a substantial portion of my time at UCSD and we have included the result of this work in Appendix A. In this case, the dissipative component of the response function leads to a noise correlation function that depends on the instantaneous conductance, and thus the phase of the

junction.

1.2.1 Memristor Models

For our purposes it is more convenient to formulate things in terms of conductance. Rather than use the unfortunate 'memductance', we will refer to these devices as memristors trusting that the context will make our meaning obvious. It is sufficient to consider a time-invariant memristive system with a scalar state variable $x \in \mathbb{R}$,

$$I = G(x)V$$
$$\dot{x} = f(x,V).$$

where we may alternatively take $\dot{x} = f(x,I)$ to be a function of the current rather than voltage. There is substantial freedom in choosing the state variable, and we could in principle eliminate x entirely by solving for $x(G)$ and using

$$I = G(t)V \tag{1.5}$$
$$\dot{G} = G'(x(G))f(x(G),V). \tag{1.6}$$

The choice of conductance function is thus mostly a matter of clarity and convenience. We will most often assume that the conductance varies between two limiting states, $g_{on} > g_{off}$, interpolated by a state variable $x \in [0,1]$ as,

$$G(x) = g_{off} + x(g_{on} - g_{off}). \tag{1.7}$$

If the evolution of the state variable x can be kept away from its boundary values and $g_{on} \gg g_{off}$, then by rescaling the interval $[0,1]$ we have the approximate form,

$$G(x) = \Delta g\, x \tag{1.8}$$

with $x \in \mathbb{R}^+$ and Δg a constant. This same form, with $x \in [0,1]$ is sometimes useful if the off conductance can be cancelled with another element in the circuit.

The dynamics of the state or memory variable x are such that when current or voltage is applied in one direction, the memristor increases its conductance and when applied in the other it decreases its conductance. Two natural choices are

$$\dot{x} = \beta V w(x) \quad \text{or} \quad \dot{x} = \beta I w(x) \tag{1.9}$$

where β is a timescale and $w(x)$ is a windowing function such that $w(x) > 0$ on the interval in which x is defined and $w(x) \leq 0$ outside of this interval. These correspond to voltage controlled and current controlled memristors. The two windowing functions we will make use of are, $w(x) = \text{rect}(x - \frac{1}{2})$ and $w(x) = x(1-x)$ which respectively give linear and exponential convergence to the endpoints.

Memristors may also exhibit a threshold in their dependence on the applied voltage/current such that they respond slowly for $V < V_T$ and rapidly past this threshold. In physical devices the difference in timescales between these two regimes can be marked, driven by transitions in the substrate as discussed in 1.1. This can be modelled by a piecewise linear function such as,

$$\dot{x} = w(x) \cdot \begin{cases} \beta_2(V - V_T) + \beta_1 V_T & V > V_T \\ \beta_1 V & V_T \geq V \geq -V_T \\ \beta_2(V + V_T) - \beta_1 V_T & -V_t > V \end{cases} \tag{1.10}$$

with $\beta_2 > \beta 1$. This may be similarly formulated as a current controlled device and/or with differing forward and backward thresholds.

Physical memristors also display volatility, causing them to relax to their g_{off} state when not subjected to a sufficient voltage/current. This may be modeled as,

$$\dot{x} = \beta w(x)(V - \gamma) \tag{1.11}$$

which will cause the memory variables to decay when subjected to a voltage $V < \gamma$, and also introduces a threshold in the forward switching.

1.3 Collective Switching in Memristive Networks

Our eventual goal will be to utilize memristive dynamics to perform useful computations in electric circuits. A primary feature we would like to leverage in this is their potential for collective behavior when arranged in networks. As a first pass at understanding this phenomena, we investigate the presence of collective switching in the simplest possible setting: a lattice of memristors subject to a slowly ramped voltage at the boundaries, discussed in detail in Chapter 2 [SDV17].

The memristor model we consider is the limit of (1.10) when we take $\beta_1 \to 0$ and $\beta_2 \to \infty$. These limits are a consequence of the slow driving limit imposed on the voltages which renders the individual switching time of a memristor negligible $(1/\beta_2 \ll \tau_{VOLTAGE})$ and the separation in timescales surrounding the threshold which allows for $\tau_{VOLTAGE} \gg 1/\beta_1$. In this limit the memristors switch discretely between g_{off} and g_{on} when their threshold is crossed.

Disorder is incorporated through a stochastic threshold among the memristors, drawn from $p(T)$ which may be due both to some degree of randomness in the memristors themselves, and a coarse-graining of structural disorder in the network. Within the network, the transition of a memristor from $g_{off} \to g_{on}$ diverts current from its neighbors in parallel, increasing the

current through its neighbors in series. The increase in current can cause neighboring memristors to transition, resulting in a series of avalanches analogous to a sandpile model. The size of these avalanches increases as more elements have switched, leading to to a transition in which a conducting backbone is formed, bridging the boundaries of the network. When biased in the opposite direction, memristors switching from $g_{on} \rightarrow g_{off}$ divert current away from neighbors in series, toward their neighbors in parallel. This can lead to a similar series of avalanches culminating in the formation of a crack separating the lattice transverse to the direction of current flow. The structures forme by these processes are displayed in Figure 1.1.

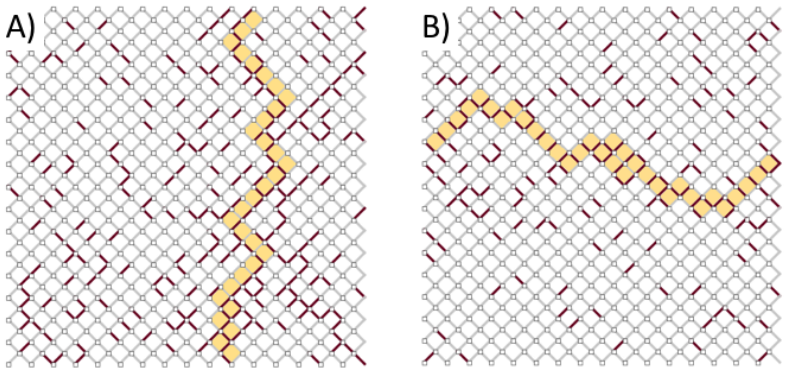

Figure 1.1: Examples of the 'bolt' and 'crack' formations in slowly driven memristive circuits. As the current through the circuit is slowly increased across the top and bottom, a 'bolt' forms bridging the boundaries of the circuit shown in A). When the voltage is reversed, a 'crack' forms, severing the boundaries as in B).

In Chapter 2 we discuss these properties in detail, showing that when g_{on}/g_{off} is small the memristors switch independently, resulting in a percolation process bridging the lattice. For sufficiently large g_{on}/g_{off} avalanches appear leading to the transition like behavior described above. This transition between a non-interacting percolative regime and an interacting regime is described by a mean-field theory and cluster approximation in which the duality between the two switching processes is manifest.

Features of this transition will survive to more complex examples of computing circuits.

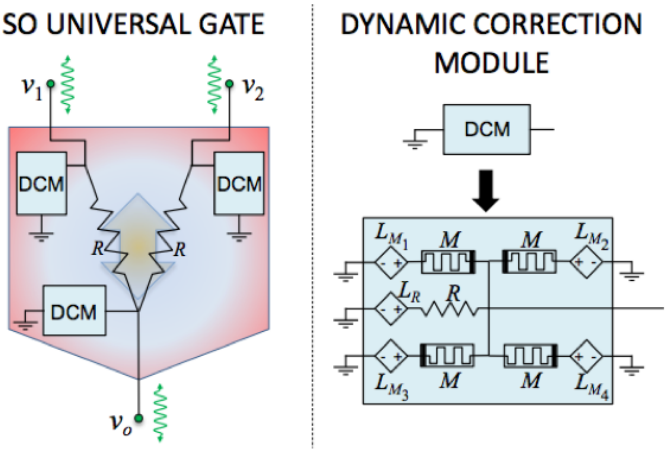

Figure 1.2: Each port of an SOLG is dressed with a *dynamic correction module*, consisting of a number of memristors, resistors, capacitors and voltage/current generators. These are configured to inject or draw current from their ports in order to impose a logical constraint amongst their voltages). [Reproduced from [TDV17].]

Specifically, in the circuit dynamics we examine in Chapters 3 and 4, we will observe a similar set of avalanches leading to a transition. These will arise from a slow driving limit imposed upon the memristive variables in the system. These avalanches however will be associated with navigating the critical points of a cost function whose minimum encodes the solution to a computational task.

1.4 Digital Memcomputing Machines and Satisfiability

Digital Memcomputing Machines (DMMs) [TDV15, TDV17, DT18] are a prescription for transforming a boolean circuit (composed of AND, OR, and NOT gates for example) into an analogous electrical circuit where each logical gate is replaced with a matching dynamical circuit element. These elements are designed to impose the intended logical relationships between their ports and when connected, lead the networks towards a state where all constraints are satisfied. As such, these elements are referred to as *self-organizing logic gates* (SOLGs) and define a *composable* framework for solving logic circuits.

Each port of an SOLG is dressed with a *dynamic correction module*, consisting of a number of memristors, resistors, capacitors and voltage/current generators as shown in Figure 1.2. These are configured to inject or draw current from their ports in order to impose a logical constraint amongst their voltages. In contrast to the networks considered in Chapter 2, these devices include a mixture of active and passive elements whose interplay is essential to their functioning as computing devices.

We can regard the equations governing the network of these elements as a dynamical system which in order to satisfy the criteria of being a DMM must satisfy 2 properties:

1. The stable equilibrium points of the system must correspond to solutions of the boolean circuit.

2. The equations must be *point dissipative* which implies that in the long time limit, they will converge to a global attractor.

In [TDV17] a set of elements satisfying these properties was described, allowing the construction of an arbitrary logic circuit. These properties imply that the resulting circuit should yield the solution in the long-time limit, constituting a hardware solver for this class of combinatorial problem. As the equations governing these networks can be efficiently integrated, the convergence of these solvers can be be assessed through simulation and the original paper also presents simulations on small instances of factoring and subset-sum demonstrating that these do indeed converge to the solution of the circuit.

While intended as a hardware solver, subsequent work has shown that simulations of DMMs can compare favorably to software solvers and thus could constitute an algorithm for the solution of combinatorial problems [TDV17, DT18]. For example, simulations of the subset-sum circuit examining how the convergence time of the circuit scales with instance size demonstrated far superior scaling compared with standard algorithms. This approach, based on the simulation of a continuous dynamical system for solving natively discrete problems occupies a somewhat

unique niche in the algorithmic landscape. Before delving deeper into an examination of the properties of this approach, we examine its dynamics in a more general setting.

In computer science, the solution of a logical circuit is known as CircuitSAT [MM11], and is a particular form of the *boolean satisfiability* problem, or SAT which asks whether a given boolean formula admits an assignment that will evaluate to true. Any boolean formula may be cast in a standard form known as *conjunctive normal form* (CNF) which consists of a conjunction (ANDs denoted by $\wedge$) of a collection of disjunctions (ORs denoted by $\vee$ also known as a *clause*) between *literals* (a variable x or its negation $\overline{x}$) [GJ90]. As an example, the formula

$$(x_1 \vee x_2) \wedge (\overline{x}_1 \vee \overline{x}_2) \wedge (x_1 \vee \overline{x_2} \vee x_3) \wedge (\overline{x}_1 \vee x_2 \vee \overline{x}_3) \wedge (x_1 \vee x_3) \tag{1.12}$$

is satisfied (uniquely) by the assignment $x_1 = 0, x_2 = 1, x_3 = 1$.

If the above formula is further constrained by additional clauses, it may become *unsatisfiable* (sometimes called UNSAT). In this case we can still ask for the maximum number of clauses that can be simultaneously satisfied, known as MAXSAT. This optimization form of satisfiability includes the original problem as a subset and is widely utilized, with yearly competitions between solvers and applications in chip design and scheduling, among others. In Chapter 3 we carry out an extensive benchmarking study, comparing DMMs performance on the most recent MAXSAT competition problems and on a class of random SAT problems we construct and extend to very large sizes. DMMs display remarkable efficacy in solving these problems, comparing favorably on competition problems and displaying far superior scaling on our constructed instances. For the class of instances presented in [TCSD18, SCTD], DMMs are able to converge to a solution in time that scales linearly with instance size, while two competition solvers we test require time that scales exponentially. Additionally, we demonstrate that this scaling advantage holds for instance sizes reaching to millions variables and consumes memory that scales linearly.

1.5 Collective Behavior in DMMs

A careful examination of the dynamics of a DMM reveals a form of collective behavior similar to that seen in the memristive networks discussed in Section 1.3 and Chapter 2 [SDV17]. The evolution of voltages in the circuit proceeds through a series of rapid transitions, with voltages switching between limiting values, separated by periods of relative quiescence. These transitions are interpretable as avalanches in much the same way, however in this case the separation of timescales emerges from the slow motion of memristive variables and a rapid voltage timescale, the reverse of the systems described in [SDV17]. Within these avalanches, correlations calculated from simulations demonstrate a form of long-range order: the correlation functions describing an avalanche can take finite values all the way to the edge of the system where they decay at the fixed boundary conditions [DVTO17]. This indicates that the transitions described by these avalanches can flip an extensive numbers of variables. While these simulations are only carried out for small systems, analytical arguments also support this. Interpreting avalanches as motion between critical points on a manifold, field theoretic arguments argue that correlations along these trajectories should display this same flat structure [DVTO17].

The potential utility of collective behavior may be illustrated by considering the example of solvers based on local search. This class generates an initial state and makes iterative improvements upon it in order to locate the solution. An isolated unsatisfied constraint can sometimes not be satisfied without affecting the states of its neighbors, forming a local minimum in which several variables must be flipped at once in order to escape. For example, in an XORSAT problem (boolean satisfiability with clauses formed by exclusive ORs) flipping a single variable alters the satisfaction of all clauses that include the variable. This sets a minimum step size n that a solver must be able to take in order to escape and continue its search.

The presence of collective transitions capable of escaping these minima thus could aid in the search for a solution *if* the method of pruning the search through possible transitions is

performed efficiently *and* the transitions are actually productive in locating the solution. The deterministic dynamics of DMMs imply that the structure of transitions is determined by past dynamics (through stored memory) and the critical points of the underlying manifold. As a first step towards answering these questions, we attempt to simplify the system as much as possible. In doing so, we show that collective behavior in the dynamics is a useful heuristic in constructing local solvers.

1.6 Nonlinearity and Memory

DMMs were originally posed as hardware solvers and the constraints imposed by this necessarily complicated their implementation. Observing the structure of an SOLG in Figure 1.2 [TDV17] more closely, each port consists of a number of memristors, resistors and capacitors attached to voltage generators. These generators are *local* functions of the other voltages in the SOLG which act to impose the logical constraint. The memristors are tasked with increasing the influence of a particular generator when the constraint it is imposing is not satisfied. Generally, this is accomplished by increasing its conductance when the constraint is unsatisfied and allowing more current to flow through the port.

The primary feature that memristors add to an electrical circuit is a form of integral memory in the dynamical equations. Considering the circuit shown in Fig. 1.3 where v_M is a voltage generator and v is the voltage across the capacitor, Kirchoff's laws give an equation for $\dot{v}$ of the form,

$$C\dot{v} = G(x)(v_M - v) \tag{1.13}$$

$$\dot{x} = f(x, v_M - v). \tag{1.14}$$

These may be recognized an an integral memory term in the dynamical equations by

formally eliminating the conductance,

$$C\dot{v} = \left(\int_0^t dt \, \frac{dG}{dx} f(x, v_M - v) + G_0 \right) (v_M - v) \tag{1.15}$$

which demonstrates that memristors are also a way of adding nonlinearity to the dynamics of the circuit through memory. The precise form of the nonlinearity will depend on the dependence of the conductance on the internal state variables and their evolution with the voltage. If the memristive timescale is sufficiently slow we can interpret the dynamics as overdamped movement in a quadratic potential with a minimum at v_M and with a rate controlled by the conductance.

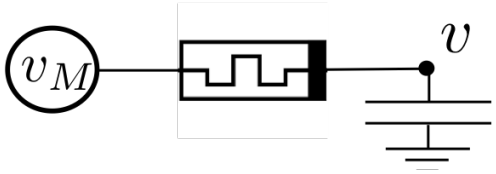

Figure 1.3: A simple memristor based dynamical system. The voltage generator v_M influences the value voltage value at the node v through the memristor. The capacitor-memristor combination introduces a timescale

To accommodate the form of available memristors, the port of an SOLG may contain up to four memristor-voltage generator pairs, each contributing a term of the form (1.15) along with an additional resistor-voltage generator pair designed to divert current when the constraint is satisfied. In order to arrive at a simplified system we amalgamate the contributions of these separate components into a single term with the basic structure borrowed from (1.15),

$$\dot{v} = \Delta g x \Delta V_M + g_R \Delta V_R \tag{1.16}$$

$$\dot{x} = f(x, \Delta V_x). \tag{1.17}$$

In this, Δg and g_R are constants, x is the memristive state variable and ΔV_M, ΔV_R and ΔV_x are functions of the local voltages involved in the constraint, i.e. the voltages at the other ports of the SOLG. Our goal will be to choose functions and constants such that the voltages relax to

a state which fulfills some portion of the constraints on the system, and then the motion of the memristive variables catalyze transitions towards a state where more are satisfied, in a way that manifests long-range behavior. While for a single memristor f usually displays a monotonic dependence on ΔV_M (i.e. memristors are often polar), our 'coarse-graining' of the memristive variables means that f may display a nonlinear dependence on the voltages in the system.

1.7 Hardware Solvers and Spin-Glasses

In order to determine the functions discussed above, we must first decide on a native problem form. In the interest of dynamical simplicity we consider the problem of finding the ground state of an Ising spin glass. This problem is attractive as its constraints only depend on two variables and they are homogeneous throughout the system. Compared to circuitSAT, the complexity of the resulting system is greatly reduced.

The energy of an Ising spin glass is a quadratic form on a set of discrete spin variables $s_i \in \{-1, 1\}$, usually written [FH93, MM09],

$$E(\{s_i\}) = -\sum_{\langle ij \rangle} J_{ij} s_i s_j. \tag{1.18}$$

The sum runs over the edges of a graph which in physical applications is usually taken to be a regular lattice. Spin-glasses are models of systems with quenched disorder such that J_{ij} are random variables described by a distribution. The thermal properties are closely tied to the particular distribution chosen but for many reasonable choices the resulting energy landscape contains exponentially many local minima [FH93, MM09]. The complex structure of the energy landscape obscures a search for the global minimum and to interesting phenomena like ergodicity breaking and slow dynamics which makes them useful models for physical glasses.

While centrally a physics problem, there are close correlates in computer science. The problem of finding the ground state is equivalent to finding the MAXCUT of a graph with edge

weights given by $\{-J_{ij}\}$. This consists of partitioning the nodes/spins into two sets such that edges connecting between the sets have maximum total weight [MM11]. Similarly, interaction terms in the spin glass can be mapped to an equivalent weighted 2-SAT or 2-XORSAT expression which is described in Chapter 4. The resulting weighted MAXSAT problems are equivalent to finding the ground state energy of the spin-glass. This equivalence has been exploited to apply statistical methods developed to study spin-glasses to understanding the average case hardness of a variety of computational problems [MPZ02, BHL$^+$02, RTWZ01, MM09].

Spin-glasses are also receiving modern attention because they are the native instance format for the quantum annealers manufactured by D-Wave [DBI$^+$16, HJA$^+$15, KHZ$^+$15]. These devices utilize a network of physical qubits whose relaxation at low temperature will tend to yield the ground-state [SMTC02]. Efforts to detect a 'quantum speedup' in these devices has produced a lively literature examining features of their dynamics, producing benchmark instances, and comparing their performance to various traditional algorithms. Collective effects have also been implicated in the dynamics of D-Wave annealers in the form of collective quantum tunneling, where groups of qubits can transition their states together as they tunnel through a barrier in the energy landscape [DBI$^+$16]. As dynamical systems are a little explored niche in classical optimization, this opens the possibility of directly comparing classical hardware solvers to their quantum analogues. The instances we utilize in Chapter 4 are derived from benchmarking studies on D-Wave devices.

1.8 Constraints and Rigidity

A natural approach to defining the functions $\Delta V_M, \Delta V_R$, and $f(x, \Delta V_x)$ is to define a continuous energy function and have the voltages follow the gradients with their contributions weighted by a memristive variable. For example, defining the linearized sigmoid function from

-1 to 1,

$$\sigma(x) = \begin{cases} 1 & 1 < x \\ x & -1 < x < 1 \\ -1 & x < -1 \end{cases} \tag{1.19}$$

we can form the energy on the continuous variables $v_i \in \mathbb{R}$,

$$E(\{v_i\}, \{x_{ij}\}) = -\sum_{\langle ij \rangle} J_{ij}\sigma(v_i)\sigma(v_j). \tag{1.20}$$

Having the voltages follow the negative gradient and memristors increase their influence when a constraint is unsatisfied gives,

$$\dot{v}_i = -\frac{\partial E}{\partial v_i} = \sum_j J_{ij}x_{ij}\sigma(v_j)\sigma'(v_i) \tag{1.21}$$

$$\dot{x}_{ij} = x_{ij}(1 - x_{ij})\left(-J_{ij}\sigma(v_i)\sigma(v_j)\right). \tag{1.22}$$

However, the above system fails predictably. Once a constraint is satisfied, its memristive variable x_{ij} decays to 0, removing its influence on the dynamics and the system is then free to violate it, leading to oscillations.

This system also bears a close resemblance to Lagrange programming neural networks (LPNNs) [ZC92, NY96] and the dynamical systems proposed in [ERT11]. In these works a Lagrangian, $\mathcal{L}$ for a constraint satisfaction problem on variables $\{s_i\}$ is formed from a set of positive constraint functions $C_m(\{s_i\}) \geq 0$ with $C_m(\{s_i\}) = 0$ holding only when the constraint is satisfied, and a set of weights for each constraint x_m, $\mathcal{L} = \sum_m x_m C_m(\{s_i\})$. In the case of LPNNs,

the equations of motion of the system are then derived as

$$\dot{s}_i = -\nabla_{s_i}\mathcal{L} = -\sum_m x_m \nabla_{s_i} C_m, \tag{1.23}$$

$$\dot{x}_m = \nabla_{x_m}\mathcal{L} = C_m. \tag{1.24}$$

By construction, with $x_m(0) = 1$ we have $x_m(t) \in [1,\infty)$ which in our formulation (Eqs. (4.2) and (4.3)) would correspond to an *unbounded*, voltage-controlled set of memristors with equal weight. In [ERT11] the equations for the multipliers are altered to $\dot{x}_m = x_m C_m$, which has the effect of making the system hyperbolic, and is analogous to choosing *unbounded* current-controlled memristors in Eq. (4.3). When applied to MAXSAT problems, the presence of unsatisfied clauses leads to an eventual imbalance in the clause weights x_m, leading to oscillations similar to those discussed previously. In addition, the presence of unbounded variables complicates a hardware implementation and we would like a way to naturally embed the algorithm.

These considerations suggest amending the system in (1.21) with terms that maintain some contribution from the constraint when $x_{ij} \to 0$. A possible form for this term is suggested by the desire for collective motion of the spins. An isolated unsatisfied constraint cannot become satisfied without affecting the states of neighboring constraints. However, if multiple variables change states together, it is possible to simultaneously satisfy distant constraints while not affecting the state of their intermediaries. As an example, consider a ferromagnetic spin-glass with $J_{ij} = 1$ but obeying,

$$\dot{v} = \big(\sigma(v_j) - \sigma(v_i)\big)\,\sigma'(v_i). \tag{1.25}$$

If we begin the system in the state $v_i = -1$, then subjecting a single spin to an external field $h > 0$ will draw the entire system to the state $v_i = 1$. This is analogous to *rigidity* in a physical system and the dynamics in (1.25) may be interpreted as being derived from an elastic potential energy that enforces this.

This suggests a system obeying,

$$\dot{v}_i = \sum_j \left[2J_{ij}x_{ij}\sigma(v_j) + (1-x_{ij})|J_{ij}|\left(\sigma v_j - \mathrm{sgn}(J_{ij})\sigma(v_i)\right)\right]\sigma'(v_i) \tag{1.26}$$

$$\dot{x}_{ij} = \beta x_{ij}(1-x_{ij})\left(-J_{ij}\sigma(v_i)\sigma(v_j)\right). \tag{1.27}$$

where the factor of 2 has been added to control the relative contributions of the two terms. The dynamics we hope to achieve will see satisfied constraints relax to $x_{ij} \to 0$ and then transition together.

In Chapter 4 we examine the dynamics of this model in detail, showing that in the limit of slow memristive dynamics it proceeds through a series of avalanches. These form droplets which steadily increase in size, closely mimicking the behavior described in chapter 2. These lead to a transition-like event in which an avalanche can span the entire lattice and we give evidence that this event is capable of flipping an extensive number of spins. Most importantly, the inclusion of these 'rigidity terms' greatly enhances the ability of the system to locate the ground state.

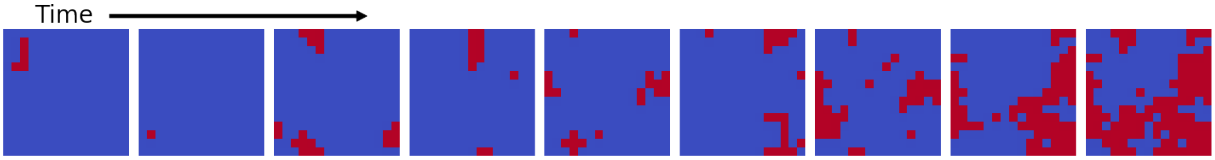

Figure 1.4: A sample of detected avalanches. For the system in (1.26) Over the course of the simulation, the average size of the avalanches grows until it reaches an extensive set of spins which can span the entire lattice. [Reproduced from [STD].]

This system serves as a simple model that allows us to investigate the emergence and effects of long-range order in computation. On large instances however, it can show quasi-periodic or possibly chaotic behavior that affects its ability to locate the solution. The extent to which these behaviors can be eliminated is the subject of current work. We also employ a full implementation of a DMM solver [TDV17] provided by Memcomputing, Inc. in a benchmarking study on a class of instances based on spin-glasses and employed in the quantum annealing community. We

examine the scaling exhibited by this approach in comparison to other local annealing approaches including simulated annealing and parallel tempering and a mixed integer programming solver. On this instance class all other solvers tested display exponential scaling in the time to solution while the memcomputing solver scales polynomially.

Chapter 2

Hors D'Oeuvre: Collective Switching in Memristive Networks

The following chapter was published as,

Forrest C. Sheldon and Massimiliano Di Ventra. Conducting-insulating transition in adiabatic memristive networks. *Phys. Rev. E*, 95:012305, Jan 2017.

Abstract

The development of neuromorphic systems based on memristive elements - resistors with memory - requires a fundamental understanding of their collective dynamics when organized in networks. Here, we study an experimentally inspired model of two-dimensional disordered memristive networks subject to a slowly ramped voltage and show that they undergo a discontinuous transition in the conductivity for sufficiently high values of memory, as quantified by the memristive ON/OFF ratio. We investigate the consequences of this transition for the memristive current-voltage characteristics both through simulation and theory, and demonstrate the role of current-voltage duality in relating forward and reverse switching processes. Our work sheds considerable light on the statistical properties of memristive networks that are presently studied

both for unconventional computing and as models of neural networks.

2.1 Introduction

Although systems that display resistive switching - also referred to as "memristive elements" (resistors with memory) - have been actively studied since the 1960s, they have recently received renewed interest in view of their possible use in computation, both as logic and memory components [PD11]. Of particular note is the tendency of some to display a history-dependent decay constant, allowing a transition between a volatile and non-volatile regime of memory [HOT$^+$10, OHT$^+$11]. The resulting dynamics bear a close resemblance to the short-term and long-term potentiation observed in biological synapses which are thought to be of central importance to learning and plasticity in the brain [Chi10a]. This resemblance has inspired the realization of experimental systems that seek to combine the memory features of biological synapses with the structural complexity of neural tissue [ASMO$^+$12, SA14]. In fact, research is being performed to assess the advantage of using memristive elements in non-von Neumann architectures and is already showing that their networks dynamically organize into the solutions of complex computational problems, thereby performing the computation directly in the memory and avoiding the separation between logic and memory units [VP13, PD13, TRBD15, TD15].

All of these studies however, still lack a fundamental understanding of the role of time non-locality in the dynamics of memristive networks and their statistical properties. For instance, high density ($\sim 10^9$ elements/cm^2) disordered networks of memristive Ag/Ag_2S/Ag atomic switches have been fabricated showing collective switching behavior between a low-resistance (G_{on}) and high-resistance (G_{off}) state, and possible critical states potentially useful in neuromorphic computation [Lan90a, Chi10a, SA12]. Some theoretical work has attempted to reproduce several features observed in the experiments by performing simulations in relatively small networks but an understanding of the dynamics of such systems is far from clear [NS11, SAS$^+$13]. Theoretical

investigations of one-dimensional memristive networks have shown complex temporal dynamics and scale-invariant properties, but have not clarified whether further collective behavior might arise in higher dimensions [DVP13a].

To fill this gap, we study the statistical properties of two-dimensional disordered memristive networks subject to a slowly ramped voltage. In this 'adiabatic' regime, where the applied voltage/current varies much more slowly than the memristance change of individual elements, there is a strong analogy between the behavior of the network and an equilibrium thermal system. Our aim is to understand the dynamics of both the disordered devices being assembled in experiment, and the ordered (but strongly heterogeneous) networks being proposed as novel computational architectures. To this end, we formulate a general model for networks in this limit which is similar to those employed to describe metal-insulator transitions and electrical failure. We thus posit that the transitions identified in these fields should also occur in memristive networks and summarize work done to describe the dynamics in these fields. Through simulations we obtain current-voltage (I-V) curves for various values of the G_{on}/G_{off} ratio and discuss the features implied by the adiabatic limit. The I-V curves found show a duality between forward and reverse switching processes that has also been observed in several experimental systems [IYAT08, SSFM15] and clarifies the role of boundary conditions in the network. These features are captured by a mean-field theory and cluster approximation which clarify the internal dynamics and account for the features of the I-V curves. These results shed considerable light on the statistical and collective properties of networks with memristive elements that are being currently explored for neuromorphic applications.

2.2 Model

As inspiration and as a test bed for disordered memristive networks we consider the $Ag/Ag_2S/Ag$ gapless atomic switches experimentally fabricated in [ASMO$^+$12, SA14, SSA$^+$14].

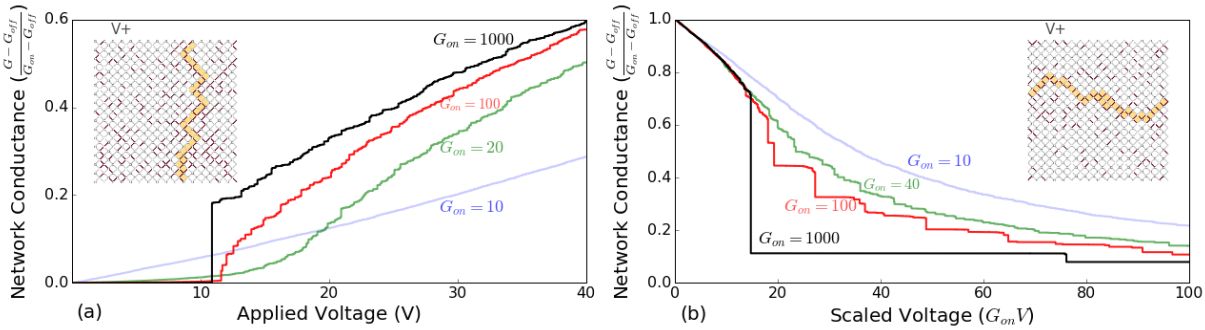

Figure 2.1: The network conductance for several values of G_{on} are plotted against the applied voltage for both (a) $G_{off} \rightarrow G_{on}$ and (b) $G_{on} \rightarrow G_{off}$ processes. The network conductance G has been scaled to vary from 0 to 1 and the range of the applied voltage shortened to focus on the point of transition. The insets show a typical network biased by a voltage V_+ just following the transition where the formation of a (a) conducting backbone and (b) crack may be observed.

The model we consider though, is quite general and can be applied to a variety of other materials and systems [PD11].

A bipolar memristor subject to an external bias will transition to a conductive state (with conductance G_{on}) in one direction and to an insulating state (with conductance G_{off}) when biased in the other. This change is generally induced by the rearrangement of ions in the applied electric field, as in the case of silver sulfide between two silver electrodes, where drifting ions form a filament structure eventually bridging the insulator [XBW$^+$10].

These switching dynamics are typically subject to a threshold in the applied bias, below which the conductance will not vary, or will vary only slowly. When considering the dynamics of a single memristor, such thresholds may be described interchangeably in the applied voltage or current. However when embedded in a network, the two can lead to quite different dynamics. When the conductivity of an element within a network is increased, its current increases while its voltage decreases, and vice versa when its conductivity is decreased. For the electric field driven migration of vacancies or ions [Di 08], we expect that in order to simulate the behavior of actual devices, all elements must be subject to a current rather than voltage threshold. The role of temperature has been emphasized by several studies, especially those focusing on the dissolution of the filament. Such an effect may be taken into account by the explicit inclusion

of a temperature in the model, or by taking a threshold in the power dissipated in an element $P_t = GI_t^2$ which gives a current threshold that changes with the device conductance $I_t = \sqrt{\frac{P_t}{G}}$. In the discretely switching model we consider here, the two choices are identical, but when the full spectrum of memristances is allowed for, as in non-adiabatic regimes, such effects may be important.

The presence of a current threshold has immediate consequences on the dynamics of a network. As the threshold of a device in the insulating state is crossed and its conductivity increases, more current is diverted through the element from neighboring bonds making the transition from G_{off} to G_{on} unstable. On the timescale of the slowly varying applied voltage, elements of the network will appear to switch discretely between the insulating and conducting state. Therefore, we can effectively model the elements as switching discretely from a conductance G_{off} to a conductance G_{on}, when a threshold current, $I_t > 0$, is crossed

$$G(I) = G_{off} + (G_{on} - G_{off})\theta(I - I_t)\,, \tag{2.1}$$

where I is the current through the device, and $\theta(\cdot)$ is the Heaviside function of the argument.

Similar considerations would lead us to conclude that the reverse direction is 'stable', as passing an element's threshold leads to a decrease in the conductance and a corresponding decrease in the current, bringing it back below the threshold. The devices would thus explore the full continuum of memristance between G_{on} and G_{off} leading to a gradual RESET-like behavior on the network level. However, we find evidence both in experimental data and simulations that this effect does not occur or is not significant in describing the behavior of the network for a wide range of parameters. For instance, in the atomic switch networks studied by Stieg *et al.* [SA12] sharp fluctuations in the network conductivity are observed in both directions, thought to be due to the switching of individual elements (see Fig. 3c in Ref. [SA12]). Such sharp behavior may be accounted for by assuming a nonlinear dependence of the conductance on the filament length due

to, for example, the transition from tunneling to ballistic conductance.

Disorder at the network level can similarly render the continuum of memristive values irrelevant to the network dynamics. If the current diverted from a switching element does not cross another's threshold, the increasing current from the boundaries will continue to transition that element to G_{off}. As the conductance and external voltage range in which this occurs is very small relative to the network scale, this is the same as if that element had switched discretely from G_{on} to G_{off}. Simulations of networks in which the full range of memristance was accessible have not shown a significant change in dynamics, and therefore we take the discretely switching model to be appropriate for a wide range of networks. We thus make a similar assumption for the reverse direction, obtaining the equations for an element by exchanging G_{on} with G_{off}, and I with I_t in Eq. (2.1) for a different threshold $I_t < 0$.

Memristive elements are also generally polar. The $Ag/Ag_2S/Ag$ atomic switches formed in atomic switch networks [ASMO$^+$12, SA14] are gapless type devices (see Hasegawa *et al.* [HTTA12] for a review of types of atomic switches and their switching processes). Their symmetric metallic configuration (typically gapless switches have two differing metallic electrodes, e.g., $Ag/Ag_2S/Pt$) suggests that at the point of formation within the network, no preferred direction within the switch has been selected. Polarity is instead instilled through a 'formation' step in which a bias is applied to the network causing filament structures to form in the switches throughout. After a joule-heating assisted dissolution of the thinnest part of the filament, the junctions display bipolar resistive switching. The polarity of the internal switches is thus determined by the direction of current propagation from the boundaries. That this must be true in the experimental systems is evident from the fact that the network undergoes resistive switching as a whole. Without a majority of switching polarities coinciding with the direction of currents from the boundaries, the network would switch between identical states with half the switches in the G_{off} state and half in G_{on} and not display the pinched hysteresis observed in experiment. We thus assign the polarity of elements in the network to coincide with the direction of currents

flowing from the boundaries.

We now turn to a network of these elements. We consider an architecture as depicted in the insets of Fig. 2.1, where the upper and lower boundaries of the network are held to some constant voltage or total current. The diamond lattice is chosen such that all elements participate equally in conduction and networks are periodic in the direction transverse to the current flow to mitigate finite size effects.

Including disorder is most directly done through random pruning of the lattice, however the networks produced are subject to strong finite size effects, requiring the simulation of many instances of large networks to obtain regular results. The random pruning of the lattice imposes a current distribution over the elements that is simply scaled by the external boundary conditions. This distribution determines the order in which elements cross their thresholds and is thus equivalent, at a mean-field level, to assuming a distribution in the thresholds of a fully occupied network. It is also noted in [KBR$^+$88] that this is obtained upon coarse-graining a randomly pruned lattice, and thus may be a better approximation to the thermodynamic limit than simulations on structurally disordered lattices. The use of a disorder distribution also affords us more direct control over the relationship between the disorder and the dynamics, simplifying the search over a large parameter space of possible structural disorders.

It is worth noting that this type of model has been arrived at in several contexts involving the interplay between conduction and disorder in 2D systems. For instance, a similar model was first applied to the study of the random fuse model for electrical failure ($G_{on} \to G_{off} = 0$) [KBR$^+$88]. A uniform distribution of thresholds on the interval $[1-w, 1+w]$ was considered and behavior examined as a function of network size L and w. Brittle and ductile regimes of behavior were identified, both of which culminated in the formation of a lateral 'crack' severing the network. In the brittle, or narrow disorder regime, this occurred as an avalanche upon the first bond failing, while for larger w there was a regime of diffuse failure, causing the networks to progressively deform before the formation of a crack. In the thermodynamic limit, only the

brittle regime survived, except for the case $w \to 1$ when the disorder distribution extended to 0. More recently, in metal-insulator transitions (MIT) [SPZS11] the $G_{off} \to G_{on}$ transition was examined for its G_{on}/G_{off} dependence, finding that a transition occurs in which the network conductivity exhibits a discontinuous jump, corresponding to the formation of a 'bolt'. Similarly, a conducting backbone has been found to form along the direction of current flow in the case of MIT and dielectric breakdown [SPZS11] for sufficiently large values of the G_{on}/G_{off} ratio.

Interest in the dynamics of individual resistive switching (RS) devices has also led to new models such as the Random Circuit Breaker (RCB) model [CLK$^+$08] for unipolar devices, capable of reproducing the conductivity dynamics of a unipolar device in the SET and RESET operations. In the RCB model, elements of a lattice transition to a conductive state when a voltage threshold is crossed, and back to an insulating state when another threshold is crossed in the same direction.

In view of these previous results, we thus expect the transitions observed in MIT and electrical failure studies to occur in memristive networks in the adiabatic limit. The $G_{off} \to G_{on}$ transition will correspond to the formation of a conductive backbone or 'bolt' through the network along the direction of current flow and the $G_{on} \to G_{off}$ transition will correspond to the formation of a 'crack' transverse to the direction of current flow severing the network. In both directions we anticipate a trivial brittle, or narrow disorder regime in which the transition occurs upon the first element switching and all elements transition within a narrow range of the applied voltage, as would be the case if all elements had the same threshold. For broad disorder we expect a ductile regime in which there will be diffuse switching leading up to the transition and activity over a broad range of applied voltages. While the ductile regime does not survive in the thermodynamic limit for electrical failure, the modest size of memristive networks experimentally realized [PD11] suggests the ductile regime is still significant in their dynamics. Of particular interest to us is the influence of such transitions on the I-V curves of the network. From the perspective of the experimenter such effects may be desirable, such as providing strong sensitivity across a small

voltage range or signaling the solution of a computational problem, or undesirable, by reducing the number of internal states accessible to the network.

Investigations of the behavior of memristive networks have been limited to date. In [NS11], the full time integration of a memristive network was undertaken for networks of moderate size. Elements did not include a threshold in their dynamics and the networks were investigated for their dependence on the fraction of memristive to ohmic conductors p, and their AC response. It was found that for poor ohmic conductors ($G = G_{off}$) the networks exhibited pinched hysteresis curves only when $p > p_c = 0.5$, the percolation threshold. For good ohmic conductors ($G = G_{on}$), a strongly memristive phase was observed for $p > p_c$ in which the networks switched abruptly, and for $p < p_c$ a weakly memristive phase was observed, similar to that for poor ohmic conductors and $p > p_c$.

Modeling performed by [SAS+13] of the "atomic switch networks" simulated small networks including volatility and noise in their memristor model, and showed an opening of the I-V curves as the noise term was reduced and $1/f^{\gamma}$ ($0 \leq \gamma \leq 2$) scaling of the power spectral density for the small networks simulated ($L \approx 10$). While such studies reproduce phenomena seen in experiment, little work has been done to analyze the behavior of these models and understand how memristors in networks interact. Here, we instead focus on building an understanding of memristive networks in the adiabatic regime, where analogies with thermodynamic systems are strongest. By moderating the strength of interactions through the G_{on}/G_{off} ratio, we examine the transition that occurs in each direction and its effects on the I-V curves of the network. The features of this transition and the resulting hysteresis curves are captured by a cluster approximation that well approximates the behavior of the network about the point of transition.

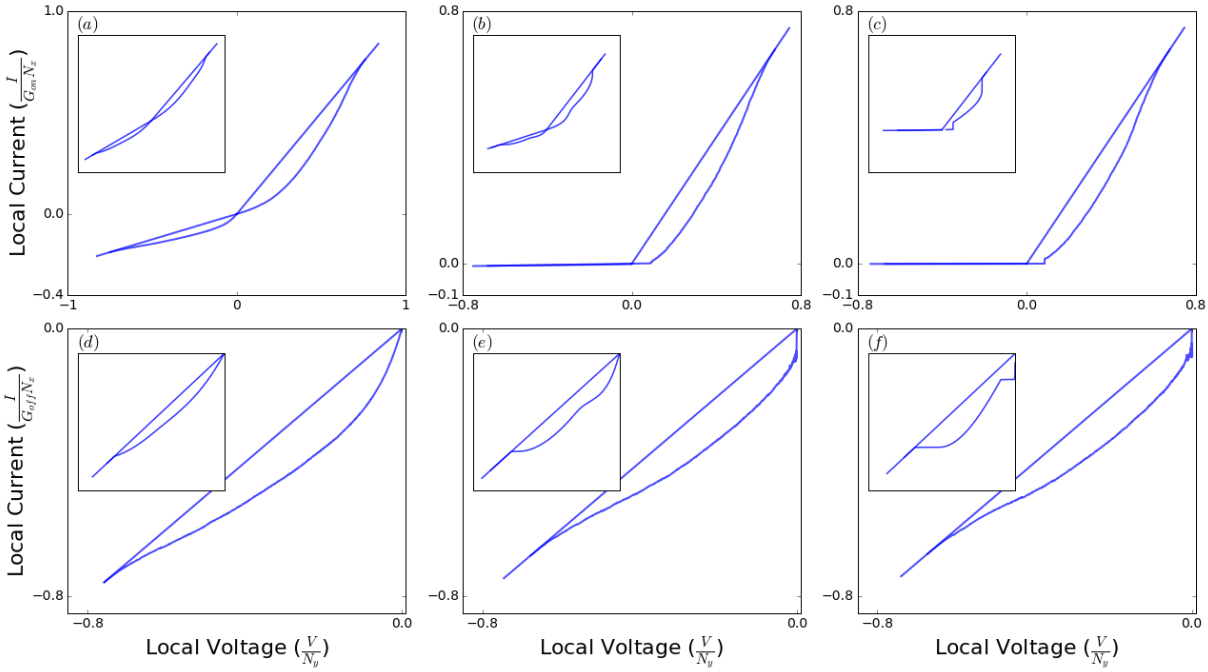

Figure 2.2: Simulated hysteresis curves of memristive networks for $G_{on} =$ (a) 4, (b) 100, and (c) 1000. The discrete jump in conductance becomes evident in the I-V curves for large values of G_{on} The asymmetry of the curves is the result of element thresholds depending on the current, and thus the transition occurring at a factor of $1/G_{on}$ lower voltage than the corresponding forward transition. The insets show analytically reached I-V curves which demonstrate similar asymmetry and the emergence of a jump in the current. The reverse branch has been rescaled and plotted in the second row for $G_{on} =$ (d) 4, (e) 100, and (f) 1000. For large G_{on}, the transition appears as a noisy area near to the y-axis. Here the jumps in conductivity at a fixed voltage give rise to sharp decreases in the current that are opposed by the subsequent increase in voltage.

2.3 Simulations

Simulations were carried out for a diamond lattice at a variety of sizes, G_{on}/G_{off} ratios, and threshold distributions $p(t)$. The network dimensions were chosen such that the conductivity of the network varied from G_{off} to G_{on}. Each element was assigned a current threshold I_t from Uniform$(0,1)$ in each direction, beyond which they transition from $G_{off} \to G_{on}$, or vice versa. Network dimensions were chosen so that the total network conductance varied between G_{off} and G_{on} ($N_x = N_y = 128$). The initial voltage is set to the value required to cross the lowest threshold in the network. Once that element has switched, voltages and currents are recalculated throughout the network with the external voltage held fixed, and all other elements whose currents exceed their thresholds are switched. This is repeated until no currents exceed the thresholds of their elements, at which point the voltage is raised until another threshold is crossed and this process repeated. The forward and reverse protocols are identical aside from the initial state and switching direction of the elements.

In Fig. 2.1 we show the network conductances as a function of applied voltage for various values of G_{on} (setting $G_{off} = 1$), in both forward $G_{off} \to G_{on}$ and backward $G_{on} \to G_{off}$ transitions, and for threshold distribution $p(t) = \text{Uniform}(0,1)$. The displayed networks have a linear size of $N_{x/y} = 128$ memristors which we found large enough to achieve regular results over multiple realizations of the disorder. Network conductances have been scaled to lie on the interval $[0,1]$. We note that for small values of G_{on} in both directions, the conductance is a smooth function of the voltage. As G_{on} is increased, a discontinuity forms in the slope which sharpens, appearing almost continuous until a discontinuous jump appears for large G_{on} analogous to a *first-order phase transition*. Similar behavior was seen for a variety of other distributions of sufficient breadth (not shown) with the point of transition, however, being distribution dependent. In the insulating transition, we have scaled the voltage by G_{on} such that the current densities of all networks are initially equal, bringing the transitions to the same scale in both polarities. While

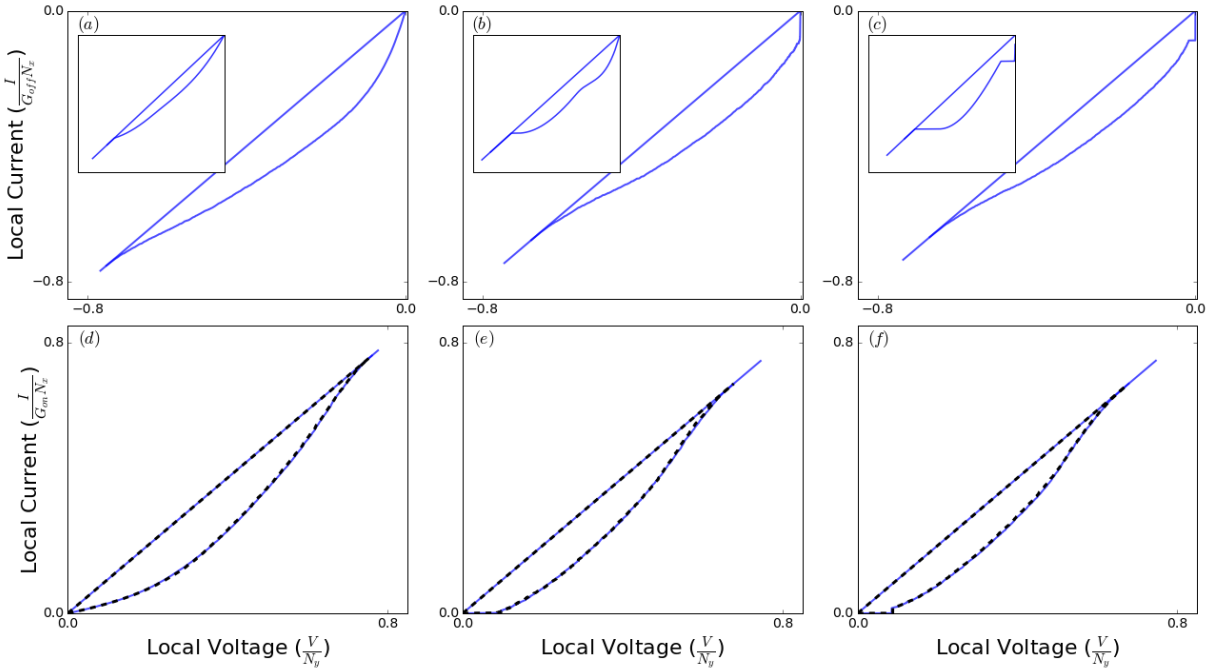

Figure 2.3: A demonstration of duality in forward and backward switching processes. The top row shows hysteresis curves switching from $G_{on} =$ (a) 4, (b) 100, and (c) 1000 to $G_{off} = 1$ in a current controlled network. For a current-controlled network switching from $G_{on} \to G_{off}$ the transition gives a discrete jump, as in the voltage-controlled $G_{off} \to G_{on}$ case. After the mapping in Eqn. (2.2), the reverse branch of the hysteresis curves (dashed) have been plotted over the forward branch for $G_{on} =$ (d) 4, (e) 100, and (f) 1000 demonstrating the duality between the two processes. The insets show analytical IV curves resulting from a cluster approximation, showing identical behavior to the voltage-controlled forward switching case.

the dependence on G_{on}/G_{off} in the forward direction has been shown in MIT [SPZS11], we are not aware of a similar demonstration in the reverse direction, possibly as most work has focused on electrical breakdown in fuse networks in which $G_{off} = 0$.

The insets in Fig. 2.1 each show the configuration of a network just following a transition. In the forward transition (a), this corresponds to the emergence of a conducting backbone spanning the network. Similarly, in the reverse transition (b) a crack forms separating the network transverse to the direction of current flow.

The corresponding I-V curves are shown in Fig. 2.2. While the voltage scale depends only on the geometry of the network, the current scale depends on G_{on}/G_{off} and has been rescaled

so that the axes coincide. The strong asymmetry of the curves is due to the use of the same threshold distribution for $G_{on} \to G_{off}$ and $G_{off} \to G_{on}$ processes. In the G_{on} state, an equivalent current will be reached for a voltage which is a factor $1/G_{on}$ lower. As there is not an obvious physical choice for how to connect the distributions for forward and reverse switching, we have displayed the negative voltage sections of the curves separately, on their own current scales. The discrete jump in the conductance becomes evident for sufficiently large values of G_{on}, but is less apparent than in the conductivity plots due to the long tail following the transition. Past the transition, voltage steps between thresholds become longer as current is diverted into the conducting backbone. The transition thus has an inhibitory effect on the remaining bonds, opening the hysteresis curves and spreading the memristive states over a wider voltage range. Thus, while the transition reduces the number of accessible states, it increases the resolution between those remaining.

In the second row of Fig. 2.2, the scaled reverse branches of the I-V curves are similar to the positive branch, but with the roles of the current and voltage exchanged. This suggests the following mapping of the reverse branch to the forward branch,

$$V \to \frac{IN_y}{N_x G_{on}}, \quad I \to \frac{V G_{off} N_x}{N_y}, \tag{2.2}$$

where N_x and N_y are the lattice dimensions. In the region about the transition however, the jump of the forward branch appears as a fluctuating region in the reverse branch. Here, avalanches of transitioning elements sharply reduce the current, which is then opposed by a subsequent increase of the externally applied voltage. If we instead run the reverse branch with current-controlled boundary conditions, this fluctuating region becomes a discrete jump as seen in the first row of Figure 2.3. In the second row, upon the mapping (2.2), the curves align nearly exactly. This correspondence between the two processes is just the familiar I-V *duality* of electrical circuit theory [Iye85]: the diamond lattice is dual to itself and taking $G \to \frac{1}{G}$ in all links takes a

voltage-controlled insulator-to-metal transition to a current-controlled electrical failure process.

In the context of memristor networks, this indicates that the voltage-controlled I-V curves are dual to the current-controlled I-V curves upon exchanging the roles of voltage and current *and* the direction of the switching process. Running the model in the current controlled setting thus gives nearly identical results, but exchanges the fluctuating region in the reverse direction for the discrete jump in the forward direction.

It is important to note that this connection between the forward and reversed hysteresis loops has been observed experimentally in individual memristive systems [IYAT08, SSFM15] as well. While the applicability of I-V duality to a passive linear network is expected, that it would hold for the dynamic non-linear elements considered here was not obvious. For our model, this occurs because in the adiabatic limit, when elements can be considered to switch discretely, there is no difference between a voltage and current threshold, and thus nothing to break the duality in the dynamics. Microscopically, this leads the structures that accompany the transitions (a backbone forming along the direction of current propagation, and a crack forming transverse to the current flow) to be dual circuits.

In the remainder of the paper, we analytically investigate our model with the aim of understanding the major features of the simulated I-V curves, namely the existence of a transition for sufficiently large G_{on}/G_{off}, and the long tail following the transition leading to the I-V curves displayed in the insets of Figs. 2.2 and 2.3. The duality between the forward and reverse switching processes, will carry through to relate models in each direction.

2.4 Mean-Field Theory

As a first step towards an analysis of the model, we develop its mean-field theory. The method followed is similar to that of Zapperi *et al.* [ZRSV99] employed to analyze random fuse networks. In this form, the central physical quantity considered is the power dissipated by the

network,

$$P = \sum_j g_j v_j^2 = \sum_j \frac{i_j^2}{g_j} \tag{2.3}$$

where g_j is the conductance of an element in the network and $v_j(i_j)$ is its voltage drop (current). We require that the average power dissipated match the power dissipated by the network $G_{net}V^2 = \frac{I^2}{G_{net}}$ and assume that all elements experience a mean-field voltage V_{MF} or current I_{MF} leading to the equations,

$$\left.\begin{matrix} G_{net}V^2 \\ \dfrac{I^2}{G_{net}} \end{matrix}\right\} = \begin{cases} N\langle g\rangle V_{MF}^2 \\ N\langle \frac{1}{g}\rangle I_{MF}^2. \end{cases} \tag{2.4}$$

The choice of the LHS is determined by the boundary conditions applied to the network but the choice of the RHS is not constrained. An interesting form is the 'voltage-voltage' choice, leading to the mean-field voltage

$$V_{MF} = \sqrt{\frac{G_{net}}{\langle g\rangle}} \frac{V}{\sqrt{N}} \tag{2.5}$$

which unlike other choices displays a transition in both directions. To make progress we require the network conductance G_{net}. Below the transition, where switching of elements is primarily driven by the threshold distribution and not by the influence of nearby switched elements, the conductivity of the network may be well approximated by an effective medium theory, giving $G_{net} \approx G_{eff}(f)$ as a function only of the fraction of the devices in the ON state.

The functions $\langle g\rangle(f)$, $G_{eff}(f)$, and $h(f) = \sqrt{\frac{G_{eff}(f)}{\langle g\rangle(f)}}$ are plotted in Figure 2.4. The function $h(f)$ relates the mean-field voltage felt by an individual element to the applied voltage at the boundaries and has the form of the network conductance divided by the average conductance of a single element. We can understand its non-monotonic as arising from competition between switching elements increasing $\langle g\rangle$ and concentrating current away from other elements , and the increasing conductance of the network $G_{eff}(f)$ pulling more current in at the boundaries. For small f, the average conductance of an element is increasing faster than the network conductivity, indicating that current is concentrated away from other elements more quickly than it increases

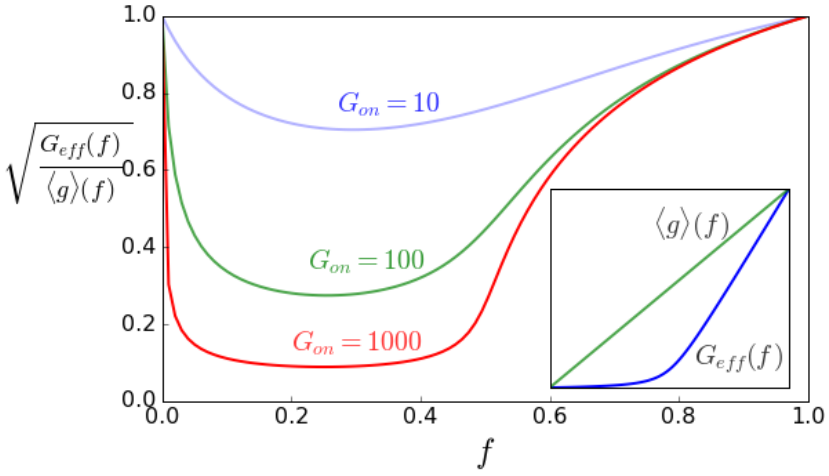

Figure 2.4: The ratio of effective and average conductances in the network, $h(f) = \sqrt{\frac{G(f)}{\langle g\rangle(f)}}$ is plotted for several values of G_{on}. In the inset, $\langle g\rangle(f)$ and $G(f)$ are plotted for $G_{on} = 100$. Note that when the average conductance increases more quickly than the network conductance as for small f, $h(f)$ is decreasing and vice versa for large f.

at the boundaries, and the mean-field voltage decreases. For larger f, the network conductance begins to increase more quickly than the average conductance, pulling in current faster than switching elements can concentrate it, and the mean-field voltage increases. The increasing regime at large f allows for a phase transition to occur from $G_{off} \to G_{on}$, and the decreasing portion at low f allows for the possibility of the reverse transition when the voltage is reversed.

To determine whether a transition occurs for a particular disorder distribution, we derive a self-consistency equation, ensemble-averaging over the number of elements that have switched for a given mean-field voltage. For the transition from G_{off} to G_{on}, the fraction that has switched will approach the average fraction of elements with thresholds below the mean-field voltage,

$$f = \int_0^{h(f)v} p(t)dt. \tag{2.6}$$

Because the applied field enters multiplicatively, the dynamics given by the mean-field theory depend only on the conductance ratio G_{on}/G_{off} and is independent of the length scale of the disorder, both amounting only to a rescaling of the applied local field v. The lhs and rhs of

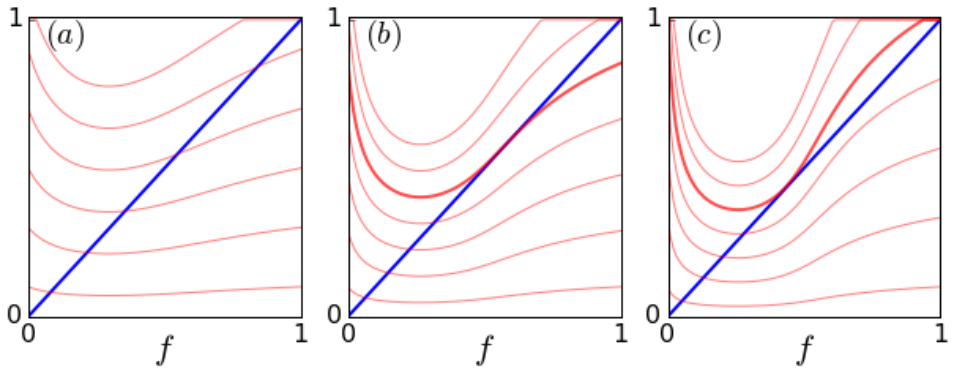

Figure 2.5: Graphical solution of mean field self-consistency. The lhs (blue/dark gray) and rhs (red/light gray) of Eq. (2.6) are plotted for several values of the applied voltage. For low values of G_{on} their intersection gives a solution that is a smooth function of the voltage (panel (a) $G_{on} = 10$). A transition develops for intermediate values that can appear continuous (panel (b) $G_{on} = 30$). For sufficiently large values of G_{on}, a transition occurs where the solution jumps discontinuously (panel (c) $G_{on} = 60$). The transition voltage is highlighted in panels (b) and (c)).

Eq. (2.6) are plotted for several values of the voltage in Figure 2.5. For the chosen distribution and G_{on}/G_{off} ratios a transition is evident at the point

$$1 = p(h(f)v)h'(f)v \quad 0 \leq f \leq 1. \tag{2.7}$$

We also note that the inflection of the curves from the RHS of Eq. (2.6) shows a trend that looks almost like a continuous transition, corresponding to the behavior seen in simulations for intermediate values of G_{on} (see Figure 2.1).

An exactly analogous treatment may be undertaken for the transition from $G_{on} \to G_{off}$. We regard $f_R = 1 - f$ as the fraction of devices in their G_{off} state and $v = \frac{V}{\sqrt{N}}$ as the positive voltage applied in the *reverse* direction. The effective medium conductivity may be obtained from the substitution $f \to 1 - f_R$, $G_{eff,R}(f_R) = G_{eff}(1 - f_R)$ and similarly with the average conductivity $\langle g \rangle_R(f_R) = \langle g \rangle(1 - f_R)$, giving a mean field voltage $V_{MF} = h_R(f_R)v = \sqrt{\frac{G_{eff,R}(f_R)}{\langle g \rangle_R(f_R)}}v$. As the mechanisms for turning ON and OFF within the individual atomic switches are not the same, we take a possibly different probability distribution $p_R(t)$ for the reverse switching

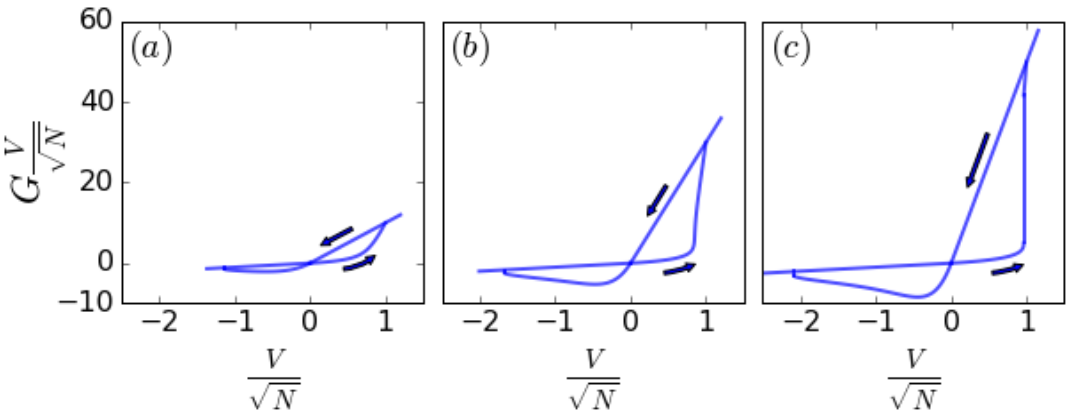

Figure 2.6: Mean field hysteresis curves. Solving the self consistency equations (2.6) and (2.8) for a uniform distribution leads to the hysteresis curves above. Here they are plotted for $G_{on} =$ (a) 10, (b) 30, (c) 50. For small values of G_{on}/G_{off}, the networks are smooth, but as the ratio is increased a discrete jump emerges in the forward direction. A similar jump near the end of the reverse branch is only barely discernible.

thresholds. With these definitions we obtain the self-consistency equation as before,

$$f_R = \int_0^{h_R(f_R)v} p_R(t)dt. \tag{2.8}$$

In Fig. 2.5 this quantity has been plotted for several values of G_{on} and the applied voltage. We observe a first-order phase transition similar to that observed in the $G_{off} \to G_{on}$ branch. However, this transition occurs for much lower values of G_{on} and near the limiting value of the conductance. As we proceed from $f = 1$ to 0, in the vicinity of $f = 0$ the average conductance $\langle g \rangle(f)$ is decreasing more rapidly than the network conductance $G_{eff}(f)$. This leads the internal switches to redistribute current to their neighbors faster than the decrease of the total current at the boundary. This increases the mean-field voltage overall, and hence promotes a transition.

Solving the self-consistency equations gives the mean-field hysteresis curves plotted in Fig. 2.6. While these curves display a qualitative similarity to simulation, several features of the mean-field theory are lacking. As we have already noted, the choice for the form of the mean-field theory (voltage-voltage, current-voltage, etc.) are not prescribed *a priori* and each choice will

give a slightly different account of the dynamics. All of these share the feature that transitions occur due to competition between a changing current at the boundaries and the internal sharing of currents within the network, as summarized by the function $h(f)$ which is some ratio between the network conductance and an average conductance $\langle g \rangle$, $\langle \frac{1}{g} \rangle$. In both directions, the transition will eventually proceed (as $G_{on} \to \infty$) from some critical fraction f_c to a fully switched network $f = 1$ as opposed to the finite jump and long tail seen in simulations.

Having observed the internal form of the transition in simulation, we see that in contrast to thermal transitions, where the phase transition occurs homogenously throughout the system, the conductivity transitions consist only of a $d = 1$ dimensional conducting backbone in the forward direction and a $d = D - 1$ dimensional crack in the reverse. In $D = 2$ both of these correspond to one dimensional subsystems of the network and so the mean-field theory which considers all elements equally cannot model it accurately, especially in the regime following the formation of the backbone. In the following, we consider methods for modeling the formation of the backbone.

2.5 1D Models

The mean-field assumption, that all elements experience either a voltage V_{MF} or current I_{MF}, is equivalent to replacing the network with a 1D parallel or series arrangement of memristors whose boundary conditions are then matched (through an effective medium theory) to the behavior of the original network. In fact, the quantities

$$\langle g \rangle = G_{off} + \frac{n}{N}(G_{on} - G_{off}) \tag{2.9}$$

$$\left\langle \frac{1}{g} \right\rangle = \frac{G_{on} + \frac{n}{N}(G_{off} - G_{on})}{G_{on} G_{off}} \tag{2.10}$$

which appear in the mean-field equations (2.4) are also the conductances G_{net}/N for a series and parallel network of N memristors with n in the ON state. Having seen that the transition is

restricted to a small subset of the network, we do not expect that including the entire network in the backbone will capture the behavior in the vicinity of the transition (where homogeneity, and thus the effective medium theory, fails). We first explore the opposite extreme by ignoring the presence of the rest of the network and considering only those elements involved in the conducting backbone or crack.

In the forward direction, the transitioning elements are a collection of memristors in series of length N_y held at a voltage V. Such an arrangement with a fraction f in the ON state admits a current,

$$I(f) = \frac{G_{on}G_{off}}{G_{on} + f(G_{off} - G_{on})}\frac{V}{N} \tag{2.11}$$

and the fraction of elements in the ON state may be determined self-consistently,

$$f = \int_0^{I(f)} \rho(t)dt. \tag{2.12}$$

The distribution $\rho(t)$ is the distribution of thresholds in the conducting backbone which should be related to the distribution of thresholds across the network. While an explicit calculation of ρ is difficult, a reasonable approximation on the diamond lattice should be $\rho(t) = 2p(t)(1 - F(t))$ where $F(t)$ is the cumulative distribution function of the threshold distribution, such that the current always selects the path with lower threshold. We note that this concentrates the threshold distribution towards its lowest values but does not strongly alter the behavior of the theory. In the interest of simplicity, we thus maintain our use of the Uniform$(0,1)$ distribution in illustrating the features of the following 1D and cluster models. Equation (2.12) is plotted in Fig. 2.7 for several values of G_{on} in which we observe the transition first occurring at $G_{on} = 2$, and then progressing to a jump to $f = 1$ for larger values.

In the reverse direction, we consider a collection of memristors in parallel of length N_x corresponding to the crack that will eventually sever the network. As the 'crack' is separated from the boundaries, the boundary conditions are instead supplied by the network. As every strip

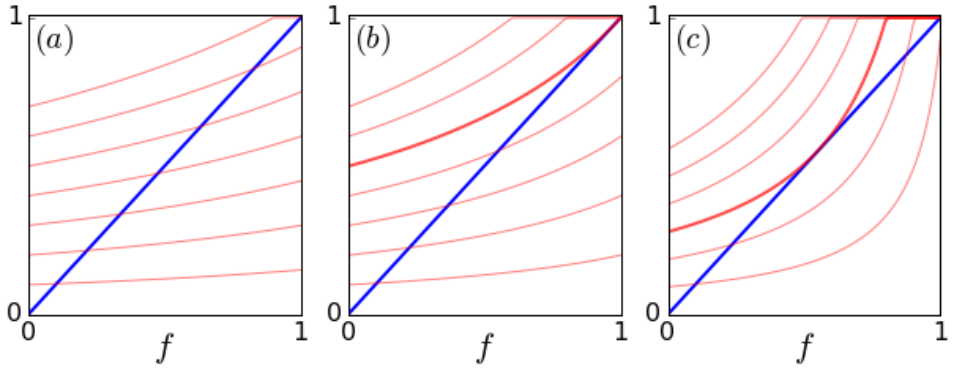

Figure 2.7: Graphical solution of mean field self-consistency for a memristor chain. The lhs (blue/dark grey) and rhs (red/light gray) of Eq. (2.12) are plotted for several values of the applied voltage for the distribution Uniform$(0,1)$. Here they are plotted for $G_{on} =$ (a) 1.5, (b) 2, (c) 10. For a collection of memristors in series, the transition is the completion of the conducting backbone, with all elements in the G_{on} state.

of memristors perpendicular to the direction of current propagation will have a current $G_{net}V$ passing through them, the reverse switching process of a voltage-controlled network should be best described by a current-controlled strip of memristors in parallel. For the moment, we again ignore the presence of the rest of the network and consider an isolated set of elements. The conductance of the 1D strip of memristors in parallel with a fraction f_R in the G_{off} state is

$$N(G_{on} + f_R(G_{off} - G_{on})) \tag{2.13}$$

which gives the current through an element in the ON state,

$$I(f_R) = \frac{G_{on}G_{net}}{G_{on} + f_R(G_{off} - G_{on})}\frac{V}{N}. \tag{2.14}$$

f_R may be similarly found self-consistently

$$f_R = \int_0^{I(f_R)} \rho(t)dt. \tag{2.15}$$

The resulting mean-field theory is just the reverse of that for the forward switching process (taking

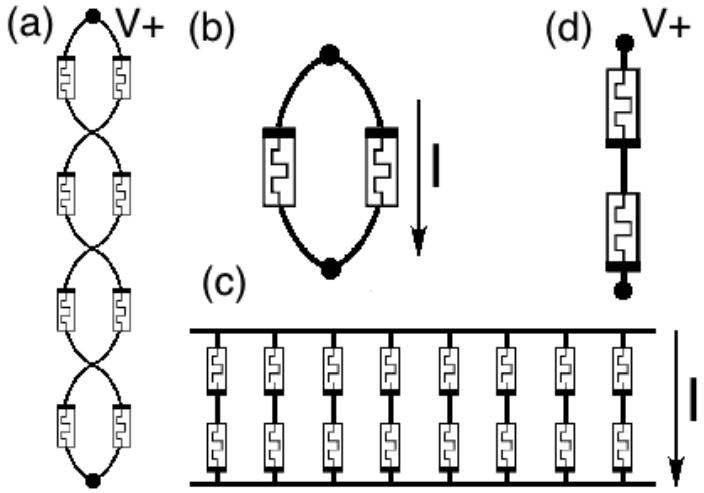

Figure 2.8: The systems and sub-units considered in the cluster approximations. As a model of the conducting backbone embedded in the network we consider (a) a series 'chain' of memristors in parallel subject to a slowly ramped voltage $V+$ composed of sub-units of (b) pairs of memristors in parallel subject to a slowly ramped current. As a model of the crack, we consider (c) a parallel strip of memristors in series subject to a slowly ramped current consisting of sub-units of (d) pairs of memristors in series subject to a slowly ramped voltage.

$G_{net} = G_{on}$) but with a voltage scale smaller by a factor of G_{on}.

Here, physical considerations from the switching processes have led us to two *dual* structures: a series chain of memristors transitioning from $G_{off} \to G_{on}$ subject to a ramped voltage as a model of the conducting backbone, and a parallel strip of memristors transitioning from $G_{on} \to G_{off}$ subject to a ramped current for the crack severing the network. Each of these demonstrates a transition in which the networks proceed from some $f = f_c$ to $f = 1$ at a critical voltage or current.

2.6 Cluster Models

In order to include the influence of the conducting backbone or crack on the rest of the network and thus the behavior of the network past the transition, we use a cluster approach similar to the Bethe-Kikuchi approximation in equilibrium thermodynamics [Bet35, Kik51]. To this end

we replace each memristor in the series chain with two memristors in parallel as in Figure 2.8 A and subject the entire chain to a slowly ramped voltage. Each pair in the chain is thus subject to a current I slowly raised from zero (See Fig. 2.8 (b)). If the threshold of each memristor, t_i is drawn independently from a distribution $p(t)$, the probability distribution for the conductance of the pair $G_{\|}$ is,

$$\begin{aligned}
p(2G_{off};I) &= 2\int_{I/2}^{\infty} dt_1 \int_{t_1}^{\infty} dt_2\, p(t_1)p(t_2) \\
p(G_{off}+G_{on};I) &= 2\int_{\frac{G_{off}I}{G_{off}+G_{on}}}^{I/2} dt_1 \int_{t_1}^{\infty} dt_2\, p(t_1)p(t_2) \\
&\quad +2\int_{0}^{\frac{G_{off}I}{G_{off}+G_{on}}} dt_1 \int_{\frac{G_{off}I}{G_{off}+G_{on}}}^{\infty} dt_2\, p(t_1)p(t_2) \\
p(2G_{on};I) &= 2\int_{0}^{\frac{G_{off}I}{G_{off}+G_{on}}} dt_2 \int_{0}^{t_2} dt_1\, p(t_1)p(t_2).
\end{aligned} \tag{2.16}$$

A long chain of such pairs in series (Fig. 2.8 (a)), each with conductance $G_{\|,i}$ will possess a total conductance close to

$$\left\langle \frac{1}{G_{chain}} \right\rangle = N \left\langle \frac{1}{G_{\|}} \right\rangle \tag{2.17}$$

and we may determine the current through the chain self-consistently as the smallest solution of the equation

$$\frac{V/N_x}{I} = \left\langle \frac{1}{G_{\|}} \right\rangle \tag{2.18}$$

where the I dependence of $\langle \frac{1}{G_{\|}} \rangle$ has entered through the averaging. Using again the distribution Uniform(0,1), the solution to this equation has been plotted in the insets of Figs. 2.2 and 2.3 for several values of G_{on}.

Here, we see the finite jump in conductance observed in simulations followed by the gradual switching of the remaining memristors. This is due to the memristor in the ON state

diverting current away from its neighbor. While a current of only $I = 2$ (in units of $G_{off}V$) is required to switch the first of the pair, a current of $I \approx (G_{on} + G_{off})/G_{off}$ is required to guarantee the switching of the second, which for large values of G_{on}/G_{off} is considerable.

An analogous treatment of the reversed switching process requires replacing each element of the parallel strip of memristors with two elements in series (Fig. 2.8 (c)) subject to a slowly ramped current. Each pair is then subject to a slowly ramped voltage V (Fig. 2.8 (d)). Again drawing the thresholds independently from a distribution $p(t)$, the distribution for the conductance of the pair is

$$\begin{aligned} p(\frac{G_{on}}{2};V) &= 2\int_{G_{on}V/2}^{\infty} dt_1 \int_{t_1}^{\infty} dt_2\, p(t_1)p(t_2) \\ p(\frac{G_{on}G_{off}}{G_{on}+G_{off}};V) &= 2\int_{\frac{G_{on}G_{off}V}{G_{off}+G_{on}}}^{G_{on}V/2} dt_1 \int_{t_1}^{\infty} dt_2\, p(t_1)p(t_2) \\ &\quad +2\int_{0}^{\frac{G_{on}G_{off}V}{G_{off}+G_{on}}} dt_1 \int_{\frac{G_{on}G_{off}V}{G_{off}+G_{on}}}^{\infty} dt_2\, p(t_1)p(t_2) \\ p(\frac{G_{off}}{2};V) &= 2\int_{0}^{\frac{G_{on}G_{off}V}{G_{off}+G_{on}}} dt_2 \int_{0}^{t_2} dt_1\, p(t_1)p(t_2). \end{aligned} \tag{2.19}$$

A strip of N_y of these, each with conductance $G_{series,i}$, in parallel (Fig. 2.8 (c)) will have conductance $N_y\langle G_{series}\rangle$. As discussed above, such a strip embedded within a network will be subject to a current I and thus satisfy

$$N_y\langle G_{series}\rangle V = I, \tag{2.20}$$

where the voltage across the strip is determined self-consistently as the smallest solution of the above equation and the voltage dependence of $\langle G_{series}\rangle$ has entered through the averaging over the voltage dependent distribution above.

The two structures considered above are again *dual* and while the self-consistency equation

may be solved as before, we instead note that the exchange

$$V \to \frac{IN_y}{N_x G_{on}}, \quad I \to \frac{VG_{off}N_x}{N_y} \tag{2.21}$$

takes the above self-consistency equation, to that of the forward switching process (we consider a square network $N_x = N_y = N$ to avoid dimensional factors)

$$N\langle G_{series}\rangle V = I \to \frac{V/N}{I} = \left\langle \frac{1}{G_{\|}} \right\rangle \tag{2.22}$$

The reversed switching process thus maps exactly to the forward switching process upon exchanging the roles of the voltage and current and scaling appropriately, as seen in the simulated I-V curves of Figure 2.3.

2.7 Mean-Field Dynamics

Although the avalanche dynamics of mean-field models akin to the random-field Ising model are well known [SDK$^+$93], we include here a brief discussion in the interest of completeness.

The mean-field theories considered lead to self-consistency equations of the form

$$f = \int_0^{h(f)v} p(t)dt \tag{2.23}$$

where v is an external field and $h(f)$ is some function of the fraction of switched elements f. We consider this relation to be initially satisfied and raise the voltage until the next threshold is passed. This takes $f \to f + \frac{1}{N}$, increasing the limit of (2.23) by $\frac{h'(f)v}{N}$. The probability that n

memristors are switched ON by this increase is given by a Poisson distribution

$$p_n = \frac{\mu^n}{n!} e^{-\mu}, \quad \mu = p(h(f)v)h'(f)v. \tag{2.24}$$

Each of the n memristors will cause a similar increase in the mean-field voltage, and thus will give rise to the same distribution. Therefore, by switching a single memristor gives rise to a poissonian branching process. This may be brought into a more useful form by calculating the total number of memristors switched in a single branching process, or the avalanche size distribution. This leads to the Borel distribution

$$p_S = \frac{(\mu S)^{S-1}}{S!} e^{-\mu S}, \quad S = 1, 2, ... \tag{2.25}$$

This distribution has mean and variance,

$$\langle S \rangle = \frac{1}{1-\mu}, \qquad \sigma_S^2 = \frac{\mu}{(1-\mu)^3}. \tag{2.26}$$

and therefore the mean-field dynamics give avalanches whose size is determined by the parameter μ.

For $\mu < 0$, raising the voltage will cause individual memristors to switch and no avalanches will occur, corresponding to a diffuse regime. For $0 < \mu < 1$, the system will display avalanches of finite size according to the Borel distribution. At $\mu = 1$, the system reaches a critical branching process, at which point the probability of an infinite avalanche begins to grow and the distribution approaches the limiting form $p_S \sim S^{-3/2}$.

The conductance jumps that the system experiences for avalanches in the regime $0 < \mu < 1$ should be approximately $G'_{net}(f)\frac{S}{N}$ and thus, for a particular value of the conductance, conductance jumps in that vicinity should follow a Borel distribution. Such jump avalanche distributions have been well studied both numerically [SPZS11] and in experiment [SRS08] for

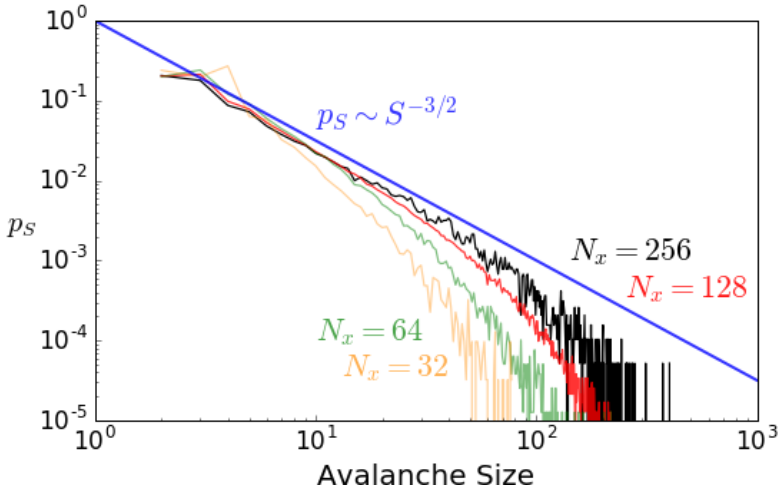

Figure 2.9: Avalanche sizes binned just above the transition for randomly diluted networks ($p = 0.6$, $G_{on} = 100$) of size $N_{x/y} = 32, 64, 128$ (1000 realizations) and $N_{x/y} = 256$ (100 realizations). As the network size increases, the avalanche size distribution approaches the asymptotic form $P(s) \sim s^{-3/2}$ given by the mean-field theory, subject to a finite size cut-off.

individual memristive elements but not yet disordered systems consisting of many memristive elements, such as those of Stieg *et al.* [SA14].

In order to confirm whether this scaling law would be accessible in experiments for physically disordered lattices, we have simulated randomly diluted lattices (by removing bonds above percolation) without threshold disorder as the effect of spatial correlations may modify the behavior. Avalanches were binned in the region surrounding the peak in the avalanche size for various sizes of the networks. The histograms produced are plotted in Figure 2.9. As the system size increases, the histograms approach the form predicted by the mean-field theory, although clearly subject to a finite-size cutoff.

2.8 Conclusions

We have presented a simple model that captures the behavior of a disordered two-dimensional memristive network when subject to bias in the adiabatic limit. As the memristive G_{on}/G_{off} ratio is increased, the conductivity changes from a smooth function of the applied

voltage to displaying a discontinuous jump as in a *first-order phase transition*. Internally, this is due to the formation of a conducting backbone or crack through the network. While the I-V curves demonstrate such a jump, the restriction of the transition to a small subset of the network elements moderates its size to a fraction of the network conductivity. Furthermore, the current diverted from the rest of the network extends the voltage range of the remaining memristors, maintaining the voltage range of the network.

The $G_{on} \leftrightarrow G_{off}$ processes are connected by *IV duality* that maps the hysteresis curves of a voltage-controlled network to those of a current-controlled network in the opposite polarity. A cluster approximation duplicates this behavior and reveals that, in order to fully transition the network, elements of the backbone will need to carry currents a factor G_{on}/G_{off} larger than their neighbors. As filament-type memristive devices have both large G_{on}/G_{off} ratios and are sensitive to the maximum current through them, this may limit the operating voltages of computational devices manufactured from memristors to the neighborhood of the transition. Fortunately, this seems to be the region in which the dynamics of the networks carry the greatest promise for the design of computational devices, as seen in maze and shortest path solvers [PD13] where the transition may correspond to the solution of an optimization problem. We hope this work will provide a foundation to extend the understanding of these networks to the non-adiabatic regime in which their behavior may be substantially more complex and interesting.

2.9 Acknowledgements

We thank G. Pruessner, S. Peotta, F. Caravelli, and F. Traversa for useful discussions and acknowledge partial support from the Center for Memory Recording Research at UCSD.

Chapter 2, in full, is a reprint of the material as it appears in Physical Review E Volume 95. Sheldon, Forrest; Di Ventra, Massimiliano, APS, 2017. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Entremet: Benchmarking Memcomputing

The following chapter has been submitted for publication and is available on ArXiV as,
Forrest Sheldon, Pietro Cicotti, Fabio L. Traversa, Massimiliano Di Ventra. Stress-testing memcomputing on hard combinatorial optimization problems. arXiv:1807.00107.

Abstract

Memcomputing is a novel paradigm of computation that utilizes dynamical elements with memory to both store and process information on the same physical location. Its building blocks can be fabricated in hardware with standard electronic circuits, thus offering a path to its practical realization. In addition, since memcomputing is based on non-quantum elements, the equations of motion describing these machines can be *simulated* efficiently on standard computers. In fact, it was recently realized that memcomputing, and in particular its *digital* (hence scalable) version, when simulated on a classical machine provides a significant speed-up over state-of-the-art algorithms on a variety of non-convex problems. Here, we stress-test the capabilities of this approach on finding approximate solutions to hard combinatorial optimization problems. These fall into a class which is known to require exponentially growing resources in

the worst cases, even to generate approximations. We recently showed that in a region where state of the art algorithms demonstrate this exponential growth, simulations of digital memcomputing machines performed using the Falcon© simulator of MemComputing, Inc. only require time and memory resources that scale *linearly*. These results are extended in a stress-test up to 64×10^6 variables (corresponding to about 1 billion literals), namely the largest case that we could fit on a single node with 128 GB of DRAM. Since memcomputing can be applied to a wide variety of optimization problems, this stress test shows the considerable advantage of non-combinatorial, physics-inspired approaches over standard combinatorial ones.

3.1 Introduction

The increasing demand for computational power and efficiency is driving the scientific and industrial communities to explore new and unconventional ways to compute. In this respect, new ideas and radically different paradigms may be the key to solve or mitigate the computational bottlenecks that affect present computing technology.

A new computing paradigm has been recently proposed called *memcomputing* (which refers to computing *in* and *with* memory) [DVP13b, TDV15], with the potential to increase the computational efficiency of the solution of hard combinatorial/optimization problems. A memcomputer is composed of interconnected dynamical elements called *memprocessors* (processors with memory) whose dynamics and interactions result in evolving the system towards the solution of a computational problem. These components can be fabricated in hardware and thus can offer a substantial increase in computation speed over approaches in software. However, we have recently shown that the *digital* (hence scalable) subclass of memcomputers which we call *digital memcomputing machines* (DMMs) can deliver substantial benefits in the solution of hard combinatorial/optimization problems even when simulated on a standard computer. In this sense, the equations of motion governing the DMM may be regarded as an "algorithm" for the solution

of the computational problem.

We have applied this approach, based on simulating DMMs to a broad array of computational benchmarks and found substantial advantages. For instance, in Ref. [DT18] we have shown that the simulation of DMMs on a classical computer solves the search version of the subset-sum problem in polynomial time for difficult cases. In Ref. [MTD18] we have shown how to accelerate the pre-training of restricted Boltzmann machines using simulations of DMMs, and demonstrated their advantage over quantum-based machines (implemented in hardware), such as those manufactured by D-Wave [AH15], as well as state-of-the-art supervised learning [GBB11, IS15]. In Ref. [TCSD18], we have employed simulations of DMMs and shown substantial advantages over traditional algorithms for a wide variety of optimization problems such as Random 2 Max-SAT, Max-Cut, the Forced Random Binary problem, and Max-Clique [GJ90]. In some cases, the memcomputing approach obtains the solution to problems on which the winners of the 2016 Max-SAT competition failed [TCSD18]. Finally, recent work has applied memcomputing to the solution of integer linear programming problems [TD18] and finding the ground state of the Ising spin-glass [STD], in both cases demonstrating significant speedup over standard methods. Collectively, these works demonstrate the widespread applicability and advantages offered by an approached based on specialized dynamical systems.

We have also performed scalability tests on finding approximate solutions to hard constructed instances of the Max-SAT problem [TCSD18]. Max-SAT possesses an *inapproximability gap* [Hås01], meaning that even calculating an approximate solution beyond a certain fraction of the optimum can require resources which grow exponentially with the input size in the worst case. We observed this exponential growth on the tested instances with some of the best solvers from the 2016 Max-SAT competition [MAX]. Instead, our simulations done for these hard cases, and up to a certain problem size succeed in overcoming that gap in linear time [TCSD18].

In the present paper we approach the distinct question of determining the largest size problem which can be approached using simulations of DMMs on current hardware. This

necessarily requires that attention be paid both to the scaling of the computation time, and to the memory consumed by the simulation since in many applications, memory is the more relevant constraint. We focus again on hard combinatorial optimization problems since these arise in nearly every industrial and scientific discipline. These problems involve the minimization or maximization of a function of many independent variables, often called the cost function, whose value represents the quality of a given solution [CHP98, EKPC13]. In industrial settings this may be the wiring cost of a computer chip layout, or the length of a delivery route [KGV83]. In scientific applications it may be searching for the ground state of a spin system [STD] or proteins [KK06].

We will show here that the simulations of DMMs on hard optimization problems can be pushed to tens of millions of variables, corresponding to about one billion literals, using a commercial and sequential MATLAB code (the Falcon© simulator provided by MemComputing, Inc.) running on a single thread of an Intel Xeon E5-2680 v3 with 128 Gb DRAM shared on 24 threads. Despite the large amount of variables (corresponding to the simulation of a substantial number of coupled ordinary differential equations), the results still scale linearly with problem size both in time and memory used. The latter is ultimately the limiting factor for the simulations. These results show once more the power of physics-inspired approaches to computation over traditional algorithmic methods.

3.2 The problem

In this work, we focus on cost functions over a set of Boolean variables $x_1, \ldots, x_N$, $x_i \in \{0, 1\}$, into which a large number of problems may be cast.

A common formulation of these problems is given by a set of Boolean constraints where the cost function counts the number of constraints satisfied by an assignment. This is often expressed in conjunctive normal form (CNF) where each constraint (also called a *clause*) is

expressed as the disjunction (logical OR denoted by $\vee$) of a set of *literals* (a variable x_i or its negation $\overline{x_i}$) which are then conjoined (logical AND denoted by $\wedge$) together, e.g., in expressions of the type:

$$(x_1 \vee x_2) \wedge (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_1 \vee \overline{x_2} \vee \overline{x_3}) \wedge (\overline{x_1} \vee \overline{x_2} \vee x_3) \\ \wedge (\overline{x_1} \vee \overline{x_3}) \wedge (\overline{x_2} \vee x_3) \wedge (\overline{x_1} \vee x_2 \vee x_3).$$

The CNF representation may be considered general in the sense that any Boolean formula may be expressed in this form [GJ90].

The problem of determining the assignment satisfying the maximum number of clauses is known as Max-SAT and is NP-hard, i.e. any problem in the class non-deterministic polynomial (NP) may be reduced to it in polynomial time [GJ90]. As a result, algorithms for obtaining the solution will, in the worst cases, require a resources that scales exponentially with the size of the instance, thus creating severe bottlenecks in the solution of complex industrial and scientific problems. The ubiquity and importance of this problem is exemplified by the yearly Max-SAT competitions, where state-of-the-art solvers are tested on a variety of benchmarks to stimulate research and innovation [MAX].

In applications where the optimal solution is required *exact* algorithms must be used [GKSS08]. These *complete* solvers typically proceed by first obtaining bounds on the quality of the solution and then using these bounds to prune the search tree in a backtracking search. This systematic approach guarantees that the resulting solution will be the optimum, but typically scales poorly and is impractical for large instances.

In these cases, *heuristic* or *incomplete* algorithms must be used [GKSS08, KSS09, Hro10]. Rather than systematically searching the solution space, these solvers generate an initial assignment and then iteratively improve upon it, using a variety of strategies to boost the efficiency of their local search. After a specified number of steps, the algorithm returns its best assignment.

As randomness is often used to drive the search, this procedure is referred to as *stochastic local search*. While they can no longer guarantee optimatility, these solvers have proven very effective at approximating, and sometimes solving difficult Max-SAT instances. For instance, in the 2016 Max-SAT competition [MAX], incomplete solvers performed 2 orders of magnitude faster than complete solvers on random and crafted benchmarks.

We might hope that if we seek an *approximation*, rather than a solution of a Max-SAT instance, we could avoid the exponential scaling of the run-time. Unfortunately, it turns out that even approximating the solution of many difficult optimization problems is NP-hard. More precisely, for a maximization problem with optimum O defined as the sum of the weights of all satisfied clauses, obtaining an assignment better than fO for the fraction f greater than some critical fraction, f_c, is an NP-hard problem [Fei98, Hås01]. For example, obtaining an assignment for Max-E3SAT (a version of Max-SAT in which every clause has 3 literals) better than $f_c = 7/8$ of the optimum is NP-hard, meaning that we cannot expect a polynomial algorithm to obtain the approximation for any instance of Max-E3SAT unless NP=P. Any algorithm thus must show exponential scaling for a threshold past f_c in the worst case. In [TCSD18] we showed a region in which some of the best algorithms based on stochastic local search show an exponential growth with input size, but where the memcomputing approach based on deterministic dynamical systems requires only linearly growing time to achieve the same threshold.

3.3 Digital Memcomputing Machines

The formal description of memcomputing rests on the concept of Universal Memcomputing Machines (UMMs) [TDV15]. UMMs are a collection of interconnected memprocessors able to process and store information on the same physical location and satisfying specific properties as described in [TDV15, PTDV17]. In particular, they support intrinsic parallelism, i.e., interconnected (mem-)processors act collectively on data using their collective state to perform

computation[TDV15, TDV17, TRBDV15]. Moreover, memprocessors are able to exploit the information available through the *topology* of their connections. Indeed, specific network topologies may be designed to embed problems in a one-to-one mapping. This last property has been named *information overhead* [TDV15].

Here we focus on the *digital* (hence scalable) subclass of UMMs which we call *digital memcomputing machines* (DMMs) [TDV17, TDV15]. DMMs can be implemented in practice as specially-designed dynamical systems whose equilibrium (fixed) points represent the approximations to the computational problem at hand, and which can be realized with standard electrical components and those with memory.

The DMM approach to a specific Boolean problem may be summarized as follows [TDV17, DT18]:

1. The Boolean circuit representing the problem is constructed.

2. Traditional logic gates are replaced with *self-organizing logic gates* (SOLGs) [TDV17] which are circuit elements designed to interact and collectively organize into a logically-consistent state.

3. The resulting *self-organizing logic circuit* (SOLC) is fed voltages along its boundary, and allowed to evolve until it reaches equilibrium where the results of a computation may be read out.

A detailed account of SOLGs may be found in [TDV17] (see also Ref. [DT18]). In essence, they may be understood as dynamical components whose equilibrium points encode the truth table of a logic gate, and which can self-organize into a consistent logical state *irrespective* of the terminal to which a given truth value (a literal) is assigned. Each terminal is equipped with a *dynamic error correcting module* which reads the states of its neighbors, and sets the local current and voltage to enforce logical constraints. These elements are specially designed so that the resulting dynamical system is point dissipative [Hal10] (in the context of functional analysis),

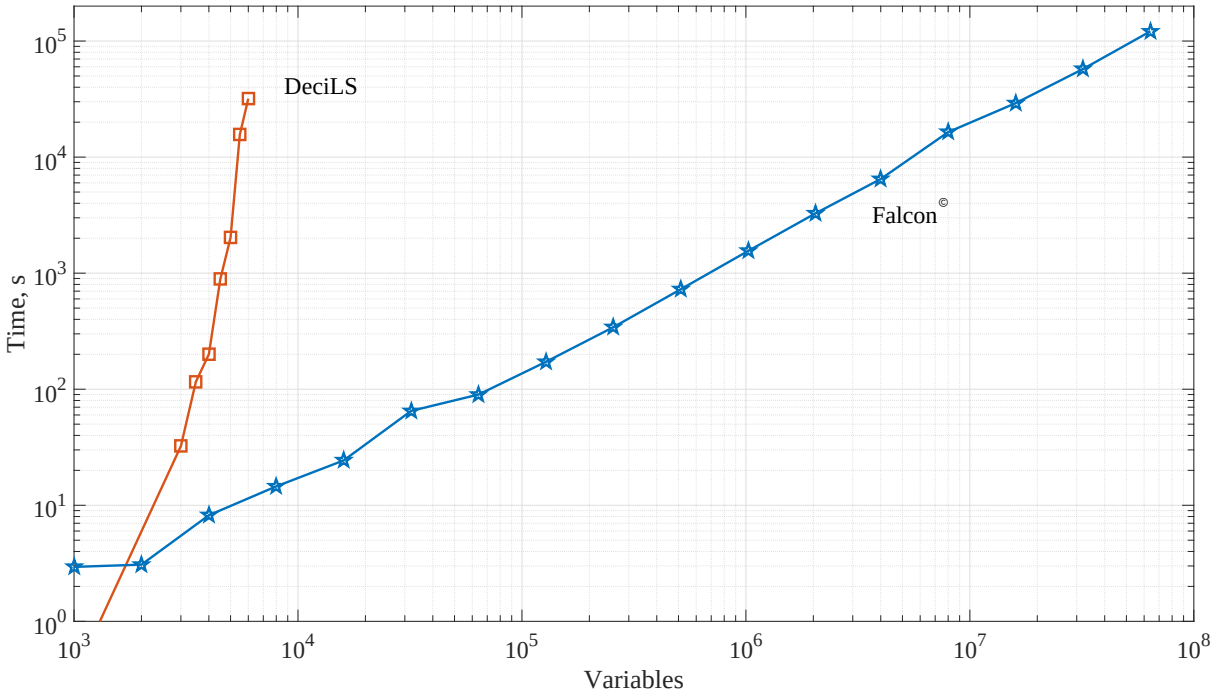

Figure 3.1: Simulation time comparison between the incomplete solver DeciLS against the Falcon© solver of MemComputing, Inc. for the balanced and constrained delta-Max-E3SAT. A threshold of 1.5% of unsatisfiable clauses has been set. We have then tested how long DeciLS and our solver Falcon© take to overcome this limit with increasing number of variables. All calculations have been performed on a single thread of an Intel Xeon E5-2680 v3 with 128 Gb DRAM shared on 24 threads. The local solver requires an exponentially increasing time to reach that limit. Our memcomputing approach instead scales linearly even up to 64×10^6 variables (corresponding to about 1 billion literals), which required a little over than 10^5 seconds on a single thread. We could not go beyond this limit because of memory resources (see text and Fig. 3.2).

avoids periodic orbits [DT17] and the chaotic behavior [DVT17] that is typical of other non-linear dynamical systems, and utilizes components with memory to allow gates to efficiently correlate and *collectively* transition between states [STD, DVTO17] (see also below).

The dynamics of the system are *deterministic* and, since they allow for collective transitions of large numbers of variables, they are also *non-local* in the same sense that complete combinatorial solvers are. Our approach therefore contrasts sharply with those based on a stochastic local search such as simulated annealing [KGV83], and many incomplete solvers [GKSS08, KSS09]. Of particular interest is the fact that the transient dynamics of a DMM proceeds via an *instantonic phase* where the machine rids itself of logical defects [DVTO17, BMTD18]. Instantons are families of trajectories of the non-linear equations of motion of DMMs, and connect different critical points in the phase space that have different indexes, namely different number of unstable directions.

In mathematical terms, if

$$\dot{x}(t) = F(x(t)) \tag{3.1}$$

is the equation of motion describing a DMM, with x the set of elements (e.g., voltages, currents and internal state variables), and F the flow vector field, then instantons are deterministic trajectories

$$\dot{x}_{cl}(t,\sigma) = F(x_{cl}(t,\sigma)); \qquad x_{cl}(\pm\infty,\sigma) = x_{a,b}, \tag{3.2}$$

that connect two arbitrary critical points of F, say x_b and x_a. The parameters σ are the so-called modulii of instantons, and encode their non-locality.

Indeed, the presence of instantons creates long-range order in the system, both in time and in space [DVTO17]. Spatial long-range order means that logic gates can correlate at arbitrary distances from each other. Temporal long-range order implies that the system can follow different paths to obtain the same solution. The structure of instantons manifests this long-range order in

collective transitions of large numbers of variables which can extend across the circuit [STD]. These transitions allow the system to navigate its state space and converge to the solution of the problem more efficiently. In fact, recent work on finding the ground state of spin-glasses [STD] has shown that an approach explicitly incorporating long-range order into the dynamics aids in locating the ground state in a simplified model of a DMM. In addition, since instantons are topological objects, the solution search is robust against noise and structural disorder, a fact that was also shown explicitly in Ref. [BMTD18].

The procedure above may be followed to construct a hardware solver for a wide variety of combinatorial and optimization problems. However, since memcomputing machines are made of non-quantum elements, the resulting behavior is also captured by a set of nonlinear ordinary differential equations describing the circuit (see Ref. [TDV17] for an example of DMMs equations of motion). These equations can be efficiently simulated, constituting an *algorithm* for the same problem. While this approach may seem indirect, as already noted, surprisingly the simulation of these circuits using standard numerical packages [Mat17] is sufficient to outperform the state-of-the-art combinatorial approaches on many benchmarks [TCSD18, MTD18, DT18].

In the subsequent section we demonstrate this and stress test these simulations by showing the results of a direct comparison on the same hardware between the Falcon© solver of Mem-Computing, Inc. and the solver DeciLS, an improved version of one of the best solvers of the 2016 Max-SAT competition [CLZ17]. In all cases, our simulations considerably outperform the other method tested, by orders of magnitude. Our solver indeed scales *linearly* in time and memory compared to the expected exponential scaling of the other solver.

3.4 Max-SAT

As outlined above, we formulate our instances as Max-E3SATs, in which each clause contains exactly 3 literals and which has an inapproximability gap. A particular instance of

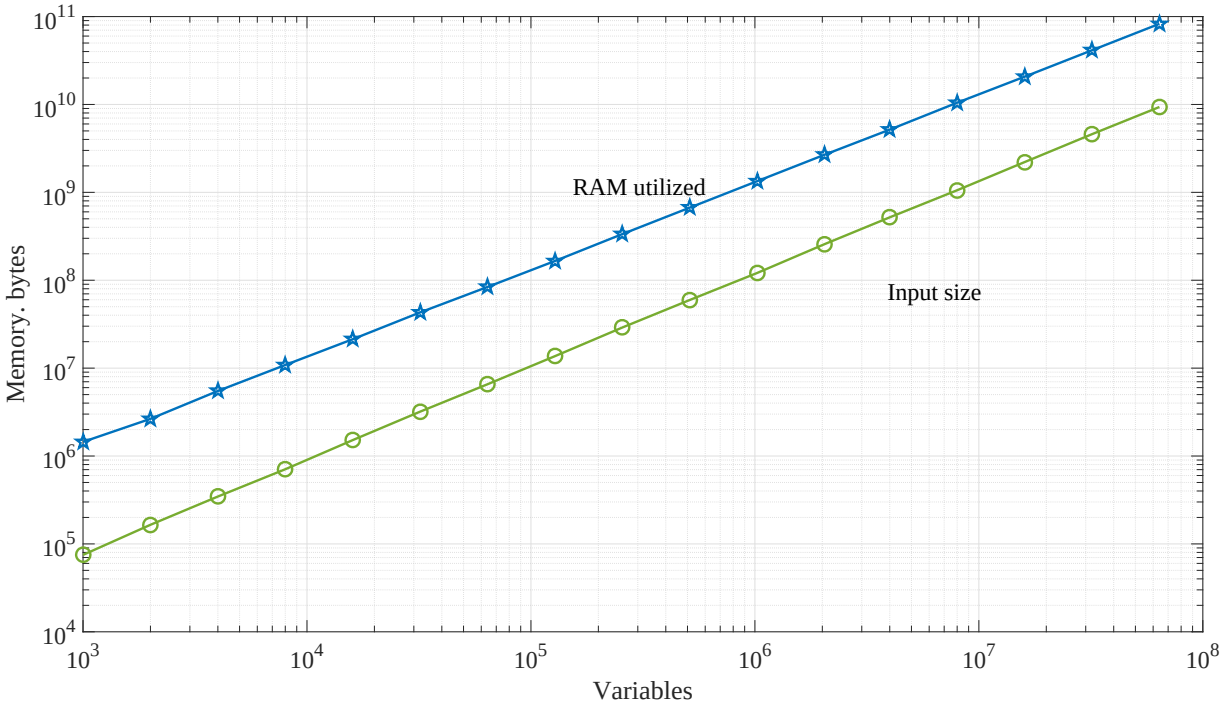

Figure 3.2: Memory requirements of the Falcon© solver as a function of variables for the delta-Max-E3SAT. We provide both the input size memory (open circles) and the RAM used during computation.

Max-E3SAT may be characterized by the number of variables, N, and the number of clauses, M, or alternatively the clause density, $\rho = M/N$. As ρ increases, the relationships between the variables become increasingly constrained, and it becomes less likely that an assignment satisfying all clauses will exists, i.e., the problem is more likely UNSAT.

In the case of Random-3SAT, in which clauses are uniformly drawn from the set of possible clauses amongst N variables, instances undergo a SAT/UNSAT transition at $\rho \approx 4.3$, below which an instance is almost certainly satisfiable, and above which it is almost certainly unsatisfiable (the transition being sharply defined as $N \to \infty$). Different methods of generating instances however, will lead to different transitions: Random-3XORSAT, in which clauses of 3 literals are formed using the exclusive OR, $\oplus$, undergoes a SAT/UNSAT transition at $\rho \approx 0.918$ [RTWZ01] for example.

To stress test the capabilities of simulating DMMs we utilized a set of Max-E3SAT instances based on a variant of Random-3XORSAT introduced in [TCSD18] in which each variable was constrained to occur the same number of times (or as nearly as possible while satisfying the specified N and ρ). This particular set of instances were chosen for their low inter-instance variability in difficulty due to the constraint placed on variable occurrences, and the relevance of MAX-XORSAT to important problems in decoding [RTWZ01, JMS04, CMMS06]. The balanced structure of the resulting SAT makes them difficult for local solvers since, whenever a variable assignment is changed, it flips the state of all XORSAT clauses in which it was included. Thus, in order to find transitions which satisfy a larger number of clauses, the system is forced to flip increasingly large numbers of literals concurrently.

To generate these hard instances, we first generated a random 3XORSAT instance with $\rho = 1.25$ in which each variable was allowed to occur 3 or 4 times. These were then converted to a SAT instance by replacing each XORSAT clause with 4 CNF clauses which reproduce the truth table of the XORSAT clause. The resulting Max-E3SAT instances have a clause density of $\rho = 5$. We call this problem delta-Max-E3SAT [TCSD18].

To clearly show the superior performance of our approach compared to standard algorithms, we have set a threshold of 1.5% of unsatisfiable clauses and tested how long DeciLS and Falcon© take to overcome this limit with increasing number of variables. DeciLS is a compiled code that combines a unit propagation based decimation (Deci) and local search (LS) with restarts [CLZ17]. Past useful assignments are used to break conflicts in the unit propagation and may be considered "messages" from the past.

The solver Falcon© is a sequential MATLAB code that integrates forward the DMMs equations of motion using an explicit Euler method. All calculations have been performed on a single core of an Intel Xeon E5-2680 v3 with 128 GB DRAM. As expected the local solver requires an exponentially increasing time to reach that limit already at a few thousand variables (see squares symbols in Fig. 3.1). Instead, the simulations of DMMs scale linearly in time up to 64×10^6 variables, which at a density of 5, corresponds to 320×10^6 clauses, or about 1 billion literals. The largest case required a little over than 10^5 seconds to complete. In fact, the memory required, rather than the time, was the primary limitation in extending the simulation further.

In Fig. 3.2 we present the memory used in the computation by Falcon© as a function of variables. It is clear that the memory also scales linearly up to 64×10^6 variables. For comparison, we also display the memory size of the input file, showing that the MATLAB implementation of a DMMs equations of motion has about an order of magnitude overhead in memory. Since the memory of the Intel Xeon used for these simulations was only 128 GB DRAM, that limit has set a hard stop to the actual size we could fit on that processor. Of course, different implementations (e.g., using a compiled language rather than an interpreted one), different hardware, etc. may permit execution of even larger instances.

The reason behind the linear scaling of the memory is that, because we are integrating the state of a dynamical system towards a solution of the computational problem, at any time in the computation only the current state of all variables in the integration must be stored. The size of this state is, in general, larger than the input size of the original Max-SAT problem as each

SOLG contains many independent variables. However, since the size of the circuit scales linearly with the input size (the number of self-organizing OR gates is just 4x the number of clauses in the XORSAT) the state which must be stored also scales linearly with the input. This slow scaling of the memory requirements with problem size is a key feature of the approach utilizing simulations of DMMs. Many other approaches to combinatorial optimization such as parallel tempering [WMK15] and integer linear programming [Sch98] will require a faster than linear memory consumption which ultimately limits the maximum size problem to which they can be applied.

3.5 Conclusions

The performance of digital memcomputing machines on the benchmarks presented in this paper demonstrates the substantial advantages of our approach, based on the *simulation* of non-linear dynamical systems, compared to traditional combinatorial ones. In particular, we have stress-tested this approach to hard instances of the Max-SAT and shown that the simulation time *and* memory of the processor used both scale linearly with problem size. Ultimately, we observe that it is the memory of the processor that sets a hard limit on the largest problem size we could address.

While we have focused on the maximum-satisfiability problem, the methods we have illustrated in which a given combinatorial problem is translated to a circuit-based dynamical system readily generalizes to a wide variety of combinatorial optimization problems as discussed in the introduction. In fact, we have already applied the memcomputing approach to several of these problems [DT18, MTD18, TCSD18, TD18, STD], in each case showing a substantial advantage over traditional algorithms. It would then seem that physics-based approaches may offer a lot to the world of computing, and we believe these ideas may form the basis for the next generation of computational devices.

Acknowledgments

We sincerely thank Dr. Shaowei Cai for providing us with the binary compiled code DeciLS. We also thank Haik Manukian and Robert Sinkovits for helpful discussions. M.D. and F.L.T. acknowledge partial support from the Center for Memory Recording Research at UCSD. M.D. and F.S. acknowledge partial support from MemComputing, Inc. All calculations reported here have been performed by one of us (P.C.) on a single processor of the Comet cluster of the San Diego Supercomputer Center, which is an NSF resource funded under award #1341698. The authors would be delighted to provide, upon request, all instances of the constrained delta-Max-E3SAT used to generate Figs. 1 and 2.

Chapter 3, in part, is a reprint of the material which was submitted for publication as, Sheldon, Forrest; Cicotti, Pietro; Traversa, Fabio L.; Di Ventra, Massimiliano. Stress testing memcomputing on hard combinatorial optimization problems. The dissertation author was the primary investigator and author of this paper.

3.6 Appendix

In testing instances to prepare for this paper, we found that the MAX-E3SAT instances we were using were not sufficiently difficult to show any difference between the solvers. The reason for this turns out to be related to an interpretation in terms of spin glasses and is a pleasant bridge to the next chapter. A random SAT instance will lead to fluctuations in the number of variable occurrences and their negations. These fluctuations lead to 'fields' which a solver can follow in order to quickly arrive at the correct solution [JMS04]. These fields appear as the coefficients in a related spin-glass hamiltonian that we will discuss below. This mapping has been utilized to design instances that are intentionally difficult and to study the distribution of solutions of the SAT problem [JMS07, BHL$^+$02, MPZ02, RTWZ01].

3.6.1 The SAT to Spin-Glass Mapping

We wish to encode the solution of a SAT problem as the ground state of a Hamiltonian. The simplest way to do this is to define a clause energy E_μ which is 0 if the clause is satisfied and positive if it is not. The hamiltonian will then be

$$H = \sum_\mu E_\mu. \tag{3.3}$$

The choice of clause energy is not unique, but a choice with the nice property that $E_\mu = 1$ if the clause is unsatisfied is

$$E_\mu = \prod_{r=1}^{k} \frac{1 - J_r^\mu s_{i_r}}{2} \tag{3.4}$$

where the product is across the k variables in clause μ. The indices $i_r \in \{1...N\}$ denote the variable in position r of the clause, and the $J_r^\mu = \pm 1$ are -1 if the variable is negated in clause μ and 1 otherwise. The spins take the usual assignment $s_i = 2x_i - 1$. If any spin agrees in sign with its J_r^μ, the clause is satisfied and this gives no contribution to the energy.

In the case of 3-SAT, we can multiply these factors out

$$E_\mu = \frac{1}{8}\left[1 - \sum_r J_r^\mu s_{i_r} + \sum_{r,r'>r} J_r^\mu J_{r'}^\mu s_{i_r} s_{i'_r} - J_1^\mu J_2^\mu J_3^\mu s_{i_1} s_{i_2} s_{i_3}\right] \tag{3.5}$$

Then summing over clauses we get,

$$H = \frac{M}{8} - \sum_\mu \sum_r J_r^\mu s_{i_r} + \sum_\mu \sum_{r,r'>r} J_r^\mu J_{r'}^\mu s_{i_r} s_{i'_r} - \sum_\mu J_1^\mu J_2^\mu J_3^\mu s_{i_1} s_{i_2} s_{i_3}. \tag{3.6}$$

To transform to sums over spins, we insert resolutions of the identity for each index $\sum_{i=1}^{N} \delta_{i,i_r}$. For the first term

$$\sum_\mu \sum_{r=1}^{3} \sum_{j=1}^{N} \delta_{j,i_r} J_r^\mu S_j = \sum_{j=1}^{N} H_j s_j, \quad H_j = \sum_\mu \sum_{r=1}^{3} \delta_{j,i_r} J_r^\mu = \sum_\mu c_{\mu,j} \tag{3.7}$$

where $c_{\mu,j} = 1$ if x_j appears in clause μ, $c_{\mu,j} = -1$ if $\bar{x}_j$ appears, and 0 otherwise. In other words this gives some global information on variable j, $H_j = n_{j+} - n_{j-}$ which 'pushes' it towards a value like an external magnetic field.

Similarly, the second term becomes

$$\sum_{i,j} T_{ij} s_i s_j, \quad T_{ij} = \sum_{\mu} c_{\mu,i} c_{\mu,j} \tag{3.8}$$

which measures the correlation between the negation of the variables in clauses (sweeping over some index manipulation to symmetrize the coefficients). Note that this term is positive in the energy so the variables which are highly correlated favor being assigned *opposite* values. Assigning them the same value would be redundant in the SAT instance.

Lastly, the 3 spin interaction is

$$\sum_{ijk} J_{ijk} s_i s_j s_k, \quad J_{ijk} = \sum_{\mu} c_{\mu,i} c_{\mu,j} c_{\mu,k} \tag{3.9}$$

which is negative in the Hamiltonian and enforces an XOR like interaction. These fields lead to a Hamiltonian of the form,

$$H(\{S_i\}) = \frac{M}{8} - \sum_i H_i s_i + \sum_{ij} T_{ij} s_i s_j - \sum_{ikj} J_{ijk} s_i s_j s_k. \tag{3.10}$$

3.6.2 Fluctuations in Fields and XORSAT

A clause of length K will contribute to the n spin local field with probability of leading order $\left(\frac{K}{N}\right)^n$ and so the sum determining it will contain $M\left(\frac{K}{N}\right)^n$ nonzero terms each of which is ± 1. So the n-body local field is of order

$$J_n \approx \sqrt{M\left(\frac{K}{N}\right)^n} \tag{3.11}$$

indicating that the size of the fields decays exponentially in order of the interaction. This is simple reason why problems are easy on average. The higher order interactions generally are what make problems most difficult but for most random instances the lower order interactions are dominant. Only rarely will we come across an instance with the 'balanced' structure to cancel out the local fields.

Instead, we can derive the SAT instance from XORSAT clauses, which are equivalent to linear equations modulo 2. For example, the constraint

$$s_k \oplus s_l \oplus s_m = 0 \tag{3.12}$$

is equivalent to an XOR clause with one variable negated, and to the set of SAT clauses,

$$(\bar{s}_k \vee \bar{s}_l \vee \bar{s}_m) \wedge (\bar{s}_k \vee s_l \vee s_m) \wedge (s_k \vee \bar{s}_l \vee s_m) \wedge (s_k \vee s_l \vee \bar{s}_m). \tag{3.13}$$

The contribution to the above mapping from this set will lead to zero contribution for the H_i and T_{ij} fields. The resulting instance has only the 3-spin interaction which is itself equivalent to the XORSAT clause and no smaller order fields for a solver to utilize.

Chapter 4

Main Course: Simplifying Memcomputing

The following chapter has been submitted for publication and is available on ArXiV as,
Forrest Sheldon, Fabio L. Traversa, Massimiliano Di Ventra. Taming a non-convex landscape with dynamical long-range order: memcomputing the Ising spin-glass. arXiv:1810.03712.

Abstract

Recent work on quantum annealing has emphasized the role of collective behavior in solving optimization problems. By enabling transitions of large clusters of variables, such solvers are able to navigate their state space and locate solutions efficiently despite having only local connections between elements. However, collective behavior is not exclusive to quantum annealers, and classical solvers that display collective dynamics should also possess an advantage in navigating a non-convex landscape. Here, we propose a simple model that demonstrates this effect, based on the recently suggested digital memcomputing machines (DMMs), which utilize a collection of dynamical components with memory connected to represent the structure of the underlying optimization problem. This model, when applied to finding the ground state of the Ising spin glass, undergoes a transient phase of avalanches which can span the entire

lattice. We then show that a full implementation of a DMM exhibits superior scaling compared to other methods when tested on the same problem class. These results establish the advantages of computational approaches based on collective dynamics.

Optimization problems draw their difficulty from the non-convexity of their associated landscapes [MM11]. These landscapes are often highly corrugated, dotted with hills, valleys and saddles of varying heights which obscure the search for a lowest (or highest) point. The complexity of this space, combined with the 'curse of dimensionality' yields an exponentially large number of potential solutions which are very difficult to prune down by any systematic method. The innate difficulty and variety displayed by optimization problems, as well as their widespread applications have made their study a continuously active field of research across science and mathematics [CHP98, EKPC13].

The exponential growth of the state space with problem size often renders any exact algorithm for locating the optimum impractical as they require an exponential amount of time to sift through the states. As a result, practitioners must rely on *incomplete* or *approximate* methods which will often generate better solutions in a limited time but are not guaranteed to converge to the exact solution [KSS09, GKSS08].

Early work on approximate methods relied on analogies with the dynamics of physical systems [KGV83] which will minimize their energy as they cool, i.e., during annealing. For example, to find the ground state of the Ising spin glass [FH93],

$$E = -\sum_{\langle ij \rangle} J_{ij} s_i s_j, \;\; s_i \in \{-1, 1\}, \tag{4.1}$$

simulated annealing gradually improves an initial state $\{s_i\}_{i=1}^N$ by stochastically exploring the state space and steadily lowering an effective temperature [CMMS06]. The early success of this approach on combinatorial optimization problems has led to the proliferation of solvers based on a similar stochastic local search and their many variants [SLM$^+$92, Sch99]. Cross pollination with

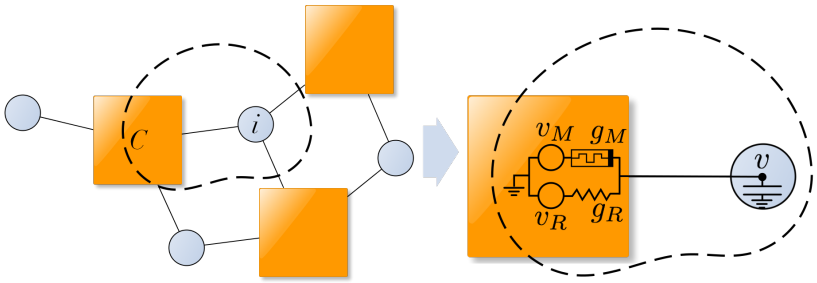

Figure 4.1: Constraint satisfaction problem as electrical circuit. An arbitrary constraint (C) satisfaction problem expressed as a factor graph can be translated into an electrical circuit with memory by considering the effect of each constraint on the site i. v_R is a voltage generator and g_R is the conductance of a standard resistor. v_M is a voltage generator and g_M is the conductance of a resistor with memory.

physics has continued, spawning methods such as parallel tempering [WMK15], and quantum simulated annealing [SMTC02] as well as the analytical characterizations of combinatorial problems [MPZ02] and random energy surfaces [BD07].

Annealing has again jumped to the forefront of modern research in the form of quantum annealing and the machines manufactured by DWave [HJA$^+$15, DBI$^+$16, KHZ$^+$15]. These machines contain 2-state quantum mechanical elements coupled together in a graph realizing a particular energy function. During their relaxation, the quantum dynamics of the system allows for collective tunneling of elements through high, thin barriers in the energy function, which may provide some advantage in the search for the optimum.

Similar ideas in the context of cellular automata, neural networks and neuroscience have already received interest [Lan90b, Chi10b]. These examples substantiate the idea that *collective behavior* would offer an advantage in the convergence of a solver by allowing for a more efficient exploration of the state space. We then expect that classical solvers which incorporate this feature in their dynamics will have an advantage in both the quality of solutions they produce, and their rate of convergence.

The purpose of this work is to explore the presence of collective dynamics in the context of specific deterministic dynamical systems: Digital memcomputing machines (DMMs) [TDV15,

TDV17, DT18]. We show that this collective behavior, in the form of *dynamical long-range order* (DLRO), allows the efficient navigation of a non-convex landscape as the one provided by the prototypical Ising spin glass (4.1).

In DMMs, a combinatorial optimization problem is first transformed into a physical system described by differential equations whose equilibrium points correspond to solutions of the original problem. Theoretical work [DVTO17] and simulations of DMMs [TDV17, DT18, TCSD18, MTD18] have indicated the presence of long-range order in their dynamics. However, as their native problem form involves several distinct dynamical elements, the complexity of the resulting solver obscures the physical principles underlying its design and function. Here, by drawing on the structure of the equations governing a DMM, we propose a simplified model that captures their essential features and can be applied in a setting more familiar to physicists: finding the ground state of an Ising spin glass [FH93].

For our purposes, this problem provides the advantage that it can be expressed in terms of very simple homogeneous constraints leading to a concise set of equations. In addition, its real-space lattice representation allows for a clearer demonstration of DLRO since the real-space distance of the lattice corresponds to the distance in the constraint graph. We draw on a class known as 'frustrated-loop instances' used by DWave and others to benchmark their quantum annealers [HJA$^+$15, KLH15]. These instances are constructed by embedding loops with a single frustrated bond on an underlying graph which we take to be a hypercubic lattice in 2 or 3 dimensions with periodic boundary conditions. These instances are convenient for benchmarking because, by construction, they have a known ground-state energy, and their difficulty is somewhat tunable by choosing the density of embedded loops [HJA$^+$15]. The construction and tuning of these instances is detailed in the Supplemental Information. After showing DLRO in the simplified model of DMMs, we then compare the results obtained by a full-fledged implementation of DMMs with those from a variety of approaches, such as parallel tempering [WMK15], simulated annealing [KGV83], as well as a commercial solver [cpl], for problem instances of increasing

size. We find that DMMs exhibit superior scaling compared to these other methods.

A DMM is constructed in correspondence to the logical circuit it will solve. For example, the subset-sum problem studied in [TDV17] utilizes a circuit with the same structure as one used to add a subset from a group of numbers. Each traditional logic gate is replaced by a *self-organizing logic gate* consisting of a set of interconnected input and output terminals, each of which is dressed with a number of memristors (resistors with memory), resistors, capacitors, and voltage/current generators forming a *dynamic correction module* (DCM) [TDV17]. When voltages are applied to the boundaries of the circuit, the dynamics of these elements are configured to satisfy the constraints enforced by each gate, and lead the circuit to a state where no logical contradictions are present.

We may consider the contribution of constraint C to the dynamics of site i (see Fig. 4.1) [TDV17]. The dynamics of the circuit are constructed such that the voltage generators impose the logical constraint on the voltage v_i at site i. The memristor conductance g_M, sensing a current flowing across it due to an unsatisfied constraint, will alter its value to accelerate the convergence of v_i to the logically-consistent solution. Generally, this is accomplished by increasing the memristor conductance, thus allowing more current to flow into or out of the site. As memristors are polar objects, complex constraints may require several memristors and generators to accomplish this, accounting for the number of memristors in DCMs [TDV17].

A few simplifying assumptions give the general form for the contribution of constraint C to site i as [TDV17],

$$\dot{v}_i = \Delta g_M x \Delta V_M + g_R \Delta V_R, \tag{4.2}$$

$$\dot{x} = h(\Delta V_M, x), \quad x \in [0,1], \tag{4.3}$$

for the voltage v_i representing the variable i and the memory state variable of the memristor x. We can regard the first and second terms on the rhs of Eq. (4.2) as representing the *total*

memrisitive and resistive contributions from the DCM, respectively. These are weighted by the conductances Δg_M and g_R, respectively, into which we have absorbed a capacitive timescale. We regard the memory state variable x and function h in Eq. (4.3) as an effective representation of the state and evolution of all memristors in the DCM, giving us considerable freedom in choosing the form of h. In order to fulfill the requirements of a DMM, the memristor equations chosen must take on *bounded* values and the equations of motion of the whole system must be *point-dissipative* [Hal10], which establishes that trajectories will converge to an invariant set that is uniformly asymptotically stable.

These equations share a close resemblance to those of Lagrange programming neural networks (LPNNs) proposed in [ZC92, NY96] and the dynamical systems proposed in [ERT11]. In these works a Lagrangian, $\mathcal{L}$, for a constraint satisfaction problem on variables $\{s_i\}$ is formed from a set of constraint functions $C_m(\{s_i\})$ which vary from 0 when the constraint is satisfied to 1 when unsatisfied and a set of weights for each constraint x_m, $\mathcal{L} = \sum_m x_m C_m(\{s_i\})$. In the case of LPNNs, the equations of motion of the system are then derived as

$$\dot{s_i} = -\nabla_{s_i}\mathcal{L} = -\sum_m x_m \nabla_{s_i} C_m, \tag{4.4}$$

$$\dot{x}_m = \nabla_{x_m}\mathcal{L} = C_m, \tag{4.5}$$

which in our formulation (Eqs. (4.2) and (4.3)) would correspond to an *unbounded*, voltage-controlled set of memristors with equal weight. In [ERT11] the equations for the multipliers are altered to $\dot{x}_m = x_m C_m$, which has the effect of making the system hyperbolic, and is analogous to choosing *unbounded* current-controlled memristors in Eq. (4.3). The dynamics of both systems are such that the variables s_i of the optimization problem act to minimize the energy, while the weights x_m act to increase it, forming a sort of competitive dynamics which seek out saddle points in the Lagrangian. The weights may be re-expressed as an integral memory term in the s_i equations and so may be interpreted as "memory variables."

The continuous constraint weighting that these Lagrangian methods perform bears a close resemblance to DMMs, but in our investigations we do not observe DLRO in the simulations of these Lagrangian systems nor do they reach the solution of the problem we consider here. Instead, in order to fulfill the properties of a DMM and display the DLRO observed in these, additional terms in the equations of motion are necessary, in particular terms that guarantee the orbits are bounded during dynamics, and that the system manifests a form of "rigidity" in which large groups of variables can transition together (for further discussion of rigidity in a continuous dynamical model of the spin glass, see the Supplemental Information).

When applied to finding the ground state of the Ising spin glass, Eq. (4.1), a simple representation of DMMs (Eqs. (4.2) and (4.3)) that satisfies these requirements is of the form,

$$\dot{v}_i = \frac{1}{2}\sum_{\langle ij\rangle} |J_{ij}|x_{ij}\big(v_i + \mathrm{sgn}(J_{ij})v_j\big) - |J_{ij}|\big(v_i - \mathrm{sgn}\big(J_{ij}\big)v_j\big), \quad v_i \in [-1,1], \tag{4.6}$$

$$\dot{x}_{ij} = \beta x_{ij}(1-x_{ij})\Big(|J_{ij}|\big(1-\mathrm{sgn}(J_{ij})v_i v_j\big) - \gamma\Big), \tag{4.7}$$

in which the memristive and resistive contributions are clearly visible. The voltages v_i are limited to the interval $[-1,1]$. From a state of the dynamical system, the spins of the original Ising spin glass model (4.1) are assigned as $s_i = \mathrm{sgn}(v_i)$ such that the spins of the Ising model undergo the orthant dynamics of the underlying continuous voltages.

The memory state follows the simplest equation for a *bounded*, volatile memristor subject to an effective voltage $|J_{ij}|(1-\mathrm{sgn}(J_{ij})v_i v_j)$. This voltage is the energy with which the constraint is violated, and the constant γ sets a threshold below which x_{ij} will begin to decay. The constant β indicates that the memristive timescale is generally different from the voltage timescale (set by the *RC* constant at the node) which will play an important role in our analysis of the system. While the memory variables do not directly interact, their coupling through the voltages leads to an effective interaction. In [CTDV17, CB18] it was shown that the network topology leads to

pairwise interactions obeying a Lyapunov function. It is useful to re-express (4.6) in terms of the regimes of the memory variable x_{ij},

$$\dot{v}_i = \sum_{\langle ij \rangle} J_{ij} x_{ij} v_j - (1 - x_{ij}) \frac{|J_{ij}|}{2} (v_i - \mathrm{sgn}(J_{ij}) v_j), \tag{4.8}$$

which shows that the x_{ij} interpolate between two different interactions between the voltages. When $x_{ij} \approx 1$ the voltages follow the fields imposed by the neighboring voltages as in an LPNN with $\mathcal{L} = \sum_{\langle ij \rangle} x_{ij} |J_{ij}| (1 - \mathrm{sgn}(J_{ij}) v_i v_j)$, causing them to take the integral values $v_i = \pm 1$. Once the constraint is satisfied, $x_{ij} \to 0$, and the voltages follow the values of their neighbors in a collective manner (see also the Supplemental Information). As a consequence, over the course of the dynamics, voltages form clusters with satisfied constraints that are capable of transitioning together under the influence of neighboring unsatisfied constraints. This has a dramatic effect on the dynamics and inclusion of these "rigidity terms" to the gradient-like first terms in Eq. (4.6) are essential for achieving DLRO in the dynamics and converging to the ground state.

We simulate the system described by Eqs. (4.2) and (4.3) from random initial voltages and $x_{ij}(0) = 0.99$, integrating the equations of motion until the energy (4.1) (calculated from the signs of the voltages) has reached the planted ground state or some maximum time has elapsed. This is typically chosen quite long, such that the system solves an instance with a probability $p > 0.95$ for a given initial condition. For a more detailed discussion of the numerical implementation, see the Supplemental Information. A typical run, showing the voltages, memristances and energy of the system is shown in Fig. 4.2 on a 2-dimensional instance, $L = 15$, where we also show that in the absence of constraint weighting via the memory variables ($\dot{x}_{ij} = 0$) the system is unable to reach the ground state (the red curve in Fig. 4.2(c)). In this case the system undergoes gradient dynamics and converges to a local minimum of $\mathcal{H} = \sum_{\langle ij \rangle} -J_{ij} v_i v_j,\ v_i \in [-1, 1]$. The action of the memory variables may be interpreted as slowly modifying this landscape to destabilize these local minima and push the system into an avalanche. That these avalanches display DRLO, is a

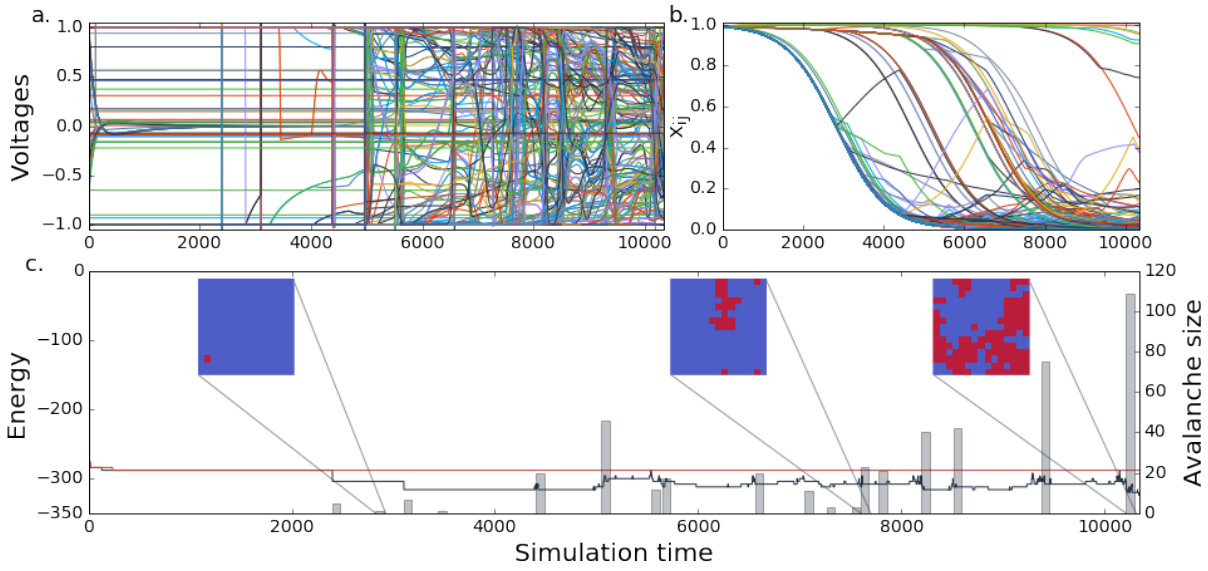

Figure 4.2: Model trajectories. When the model of Eqs. (4.6) and (4.7) is simulated for a 2-dimensional instance ($L = 15$) under a separation of timescales ($\beta = \frac{1}{400}$), the voltage trajectories (a) evolve under a series of sharply defined avalanches due to the slow motion of the memristors (b) modifying the clause weights. In (c) we have plotted the energy (left axis) without the influence of memristors (red, $\beta = 0$) and with them (black, $\beta = \frac{1}{400}$, $\gamma = 0.65$) showing that the motion of the memory variables allows the system to reach a far lower energy, and ultimately the ground state. The sizes of the avalanches (c, right axis) are plotted as gray bars, showing that their size grows over the course of the simulation until a large avalanche brings the system to its ground state. The avalanches are depicted in the inset in red with the rightmost inset corresponding to the largest avalanche in the run.

feature of the added "rigidity terms" in Eq. (4.6).

The discussion of DLRO in continuous dynamical systems is complicated by the continuity of the dynamics, making it difficult to clearly infer causal relationships between changes in variables. However, we can take advantage of the timescales above to separate the dynamics into causally related events. As shown in Fig. 4.2(a) when we slow the memristor timescale β relative to that of the voltages (e.g, by choosing $\beta = 1/400$), after the initial transient the dynamics progress through a series of rapid transitions interpretable as *avalanches* (or instantons [DVTO17]). These are due to the gradient dynamics of the voltages rapidly seeking out saddle points in the energy landscape. Since the memristive dynamics provide the unstable directions of the saddle points [DT18, DVTO17], the system will rapidly shift to a new saddle point. As more constraints become satisfied and transition to a rigid interaction, larger clusters of voltages begin transitioning together (see Fig. 4.2(c)) until, in a manner analogous to a phase transition [DT18, SDV17], extensive clusters of voltages/spins, spanning the *entire* lattice, transition *collectively* and the system converges to the ground state.

We can further emphasize the long-range nature of these clusters by computing correlation functions over the course of an avalanche. In the limit that the timescales become separated (i.e., the slow driving limit) the points at which each avalanche occurs tend towards well defined times as seen in Fig. 4.2(a). For small β these events may be detected as sharp spikes in the voltage derivatives. (See Supplemental Information for a detailed discussion of the method used to extract the structure of the avalanches.) We are interested in the voltages/spins which change sign in the avalanche and thus will affect the energy of the system. We thus define the avalanche configuration as $\Delta_i = 1$ for all spins which change sign during an avalanche, and $\Delta_i = 0$ otherwise. A few typical examples of these avalanches and their sizes occurring during dynamics are plotted in Fig. 4.2(c).

Using the avalanche configurations we are able to compute correlation functions for these events and investigate their decay across the lattice. For each run (defined as generating a unique

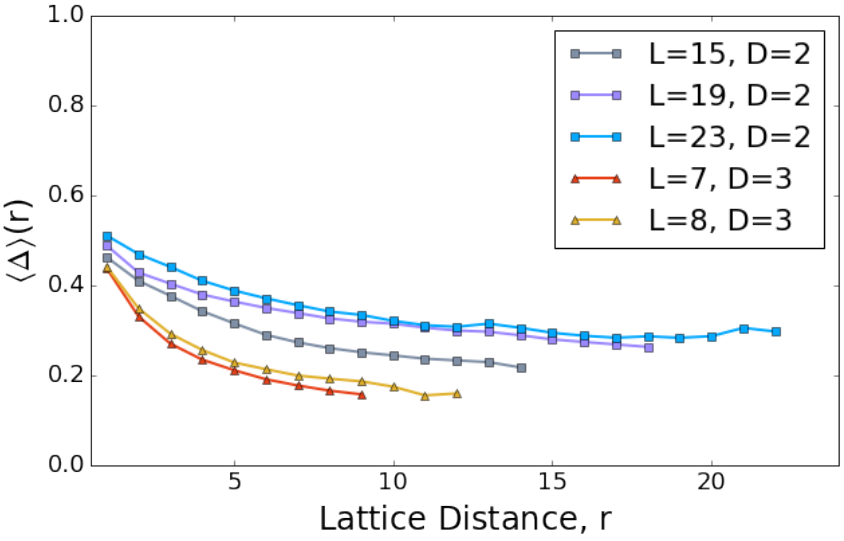

Figure 4.3: Long-range order. Spatial correlations, $\langle\Delta\rangle(r)$, among voltages/spins calculated from the orthant dynamics in the slow driving limit of model (4.6) and (4.7) for the largest avalanches in 2D and 3D and for different lattice sizes. The correlations take a finite value all the way to the lattice edge, indicating that the largest avalanches are extensive. As the system size increases the values appear to saturate to a dimension-dependent value for this instance class.

instance and initial conditions) the system is simulated until it reaches the ground state or a maximum time is reached. If the instance is solved within this interval, the largest avalanche is selected and its configuration and first flipping spin are stored. By averaging across a sample of configurations, suitably shifted so that the initial flipping spins coincide, the probability that a voltage a distance r from the initial spin changes sign, $\langle\Delta\rangle(r)$, may then be calculated. In order to achieve large distances with reasonable simulation times, we calculated these correlations both in 2- ($L = 15, 19, 23$) as well as 3-dimensional ($L = 7, 8$) systems.

As shown in Fig. 4.3, the largest avalanches possess correlations that take finite values all the way to the furthest corner of the lattice, confirming the presence of DLRO. Dimensionally, this requires that the size of the largest avalanche scales as $\sim L^D$ for a system of dimension D, and is thus extensive. We also note that, as the system size increases the correlations appear to saturate to a dimension (and instance class) dependent value.

Having used this simplified model to demonstrate the main physical ingredients promoting

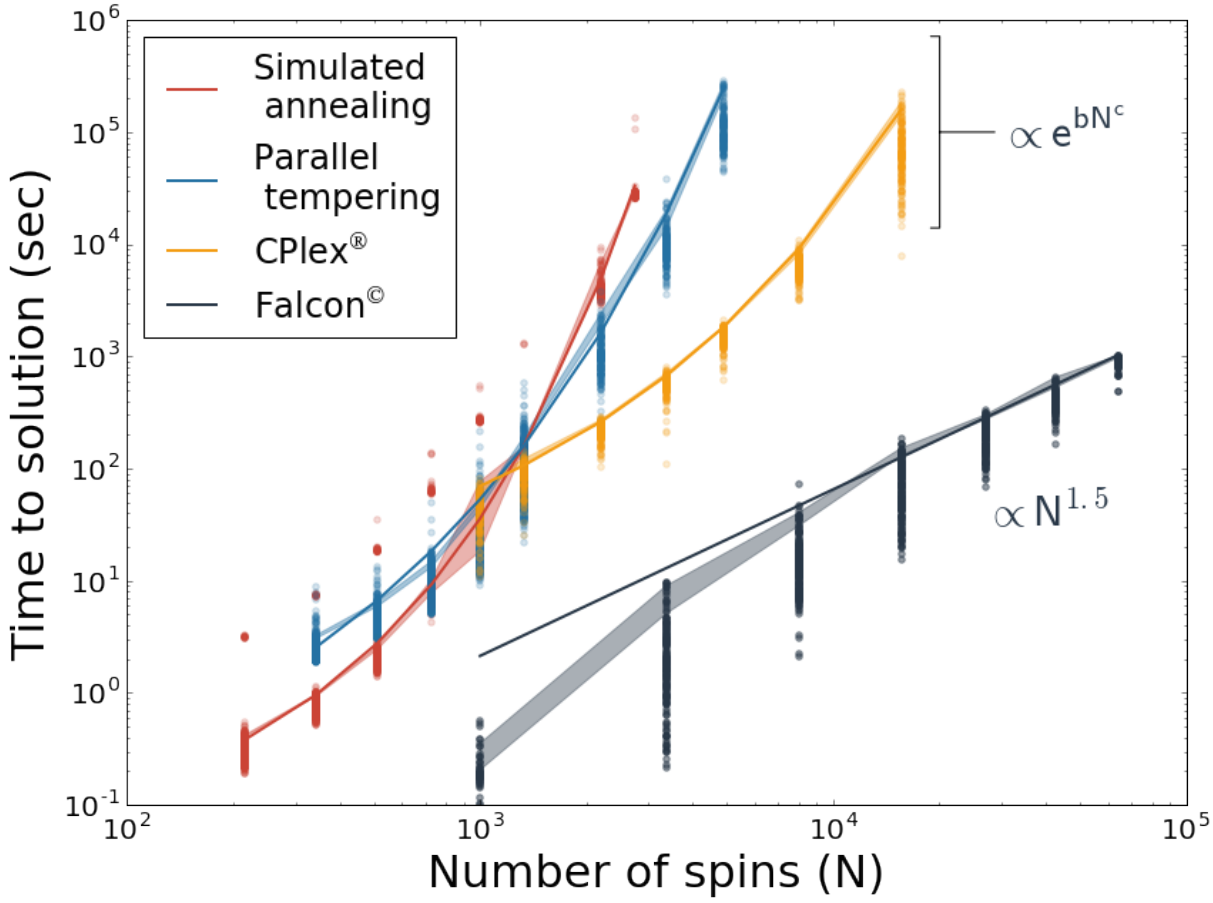

Figure 4.4: Time to solution. Time necessary for different solvers to reach the ground state of the 3D frustrated-loop spin glass as a function of the total number of spins N. The sequential memcomputing solver implemented in MATLAB is dubbed Falcon. Varying numbers of instances were run at each size and solver depending on required computation time (See the supplemental information for details). Comparisons with simulated annealing (SA), parallel tempering (PT), and IBM CPlex are also shown. All calculations were performed on a single core. The solid lines are the best fits of the 95th quantile time to solution for all four solvers. The exponential fits have the following parameters: for IBM CPlex, $b = 0.12$ and $c = 0.46$, for SA, $b = 0.069$ and $c = 0.67$, and for PT $b = 0.32$ and $c = 0.46$.

DLRO, we now show that this feature also leads to superior scaling compared to other methods that rely only on local information. Unlike the model (4.6) and (4.7), DMMs have the additional advantage of having been engineered to have no chaotic behavior or periodic orbits [DT17]. Therefore, for this scalability test, we rely on the full implementation of the dynamical equations of DMMs as in Ref. [TDV17], appropriately modified to handle the Ising spin glass expressed as a maximum satisfiability problem in conjunctive normal form [GJ90] (see the Supplemental Information for a discussion of this transformation). We then utilize a commercial sequential MATLAB solver (dubbed Falcon) that implements such equations. In addition, we have implemented two standard annealing algorithms in Python (simulated annealing (SA) and parallel tempering (PT)), as well as used a well-known commercial mixed-integer programming solver, IBM CPlex [cpl]. Since Falcon was implemented in interpreted MATLAB and the focus was on scaling rather than runtime, we used only the simplest implementation of each solver but performed substantial tuning. Details of the implementation and tuning on the instance class for SA and PT, as well as the configuration for IBM CPlex can be found in the Supplemental Information.

All solvers were run on frustrated-loop instances in 3 dimensions, ranging in size from $L = 6$ (total number of spins $N = 216$) to $L = 40$ ($N = 64,000$). The sizes used for tuning were included for the annealers (SA and PT) while CPlex and Falcon were run on sizes $L \geq 10$. As is clearly visible from Fig. 4.4, Falcon converged to the exact ground state in times orders of magnitude faster than the other methods tested, and for larger sizes than were attainable for other solvers. (CPlex, the next fastest, is predicted to take ≈ 100 years to solve an instance of size $L = 40$.) Most importantly, Falcon displays superior scaling, with the time to solution (TTS) appearing to scale approximately as $TTS \sim N^{1.5}$, while all other instances appear to scale exponentially, $TTS \sim \exp(bN^c)$, with b and c solver-specific constants reported in the supplemental information and shown in Fig. 4.4. Details of the fitting procedure may be found in the Supplemental Information.

Conclusions – In this paper, using the Ising spin glass as a well-known benchmark, we have shown that a solver exploiting dynamical long-range order can navigate a non-convex landscape more efficiently than traditional methods based on annealing despite being composed of only local connections. We have first provided a simple model of DMMs to show how to transform the original problem into a dynamical system in which DLRO emerges naturally. We have then shown results on the 3D Ising spin-glass as obtained by a full implementation of DMMs. The approach based on DMMs exhibits superior scaling in reaching the solution than the other methods tested. The results presented here further reinforce the advantages of employing collective dynamics to compute hard problems efficiently.

Acknowledgments – F.S. and M.D. acknowledge partial support from the Center for Memory and Recording Research at UCSD. The Falcon solver used in the reported simulations has been provided by MemComputing, Inc. http://memcpu.com/. The authors would be delighted to provide, upon request, all instances of the spin-glass problems used in this work.

4.1 Correspondence between statistical physics and combinatorial optimization

The problem of finding the ground state of a system in statistical physics is an optimization problem for which there is an extensive vocabulary in computer science [MM11]. It may be useful for the reader familiar with physics to have some notion of this correspondence and we include a short discussion here to that effect.

Minimizing a Hamiltonian expressed as a sum over interactions between spins may be expressed as a constraint satisfaction problem where each term in the Hamiltonian is regarded as

a constraint on the variables. For example, finding the ground state of the Ising Model

$$E = -\sum_{\langle ij \rangle} J_{ij} s_i s_j, \;\; s_i \in \{-1, 1\}, \tag{4.9}$$

is equivalent to a weighted constraint satisfaction problem, where each interaction is expressed as an exclusive-OR (XOR), or sum modulo-2 between the associated binary variables, $b_i = (s_i + 1)/2$, $j_{ij} = (1 - \mathrm{sgn}(J_{ij}))/2$

$$-J_{ij} s_i s_j \quad \leftrightarrow \quad 2|J_{ij}| \; b_i \oplus b_j = j_{ij} \tag{4.10}$$

where $2|J_{ij}|$ is the weight associated with the constraint. The correspondence between these two can also be easily seen from the fact that flipping the state of any variable changes the state of the interaction in both cases. Transformations between other constraint satisfaction problem types may be undertaken similarly. For example, when transforming to weighted conjunctive normal form (CNF), each interaction may be translated to two OR constraints depending on the sign of the interaction:

$$-J_{ij} s_i s_j \quad \leftrightarrow \quad \begin{cases} 2|J_{ij}| \; b_i \vee \bar{b}_j \\ 2|J_{ij}| \; \bar{b}_i \vee b_j, \end{cases} \quad \mathrm{sgn}(J_{ij}) = 1, \qquad \begin{cases} 2|J_{ij}| \; b_i \vee b_j \\ 2|J_{ij}| \; \bar{b}_i \vee \bar{b}_j, \end{cases} \quad \mathrm{sgn}(J_{ij}) = -1, \tag{4.11}$$

where each constraint carries a weight and negations are indicated with a bar, e.g., $\bar{b}$. In all cases, the factor of 2 may be dropped as a global scaling of the energy.

Constraint satisfaction *instances* (a particular example of the problem) may be described as being either satisfiable (SAT), if there is an assignment of the variables which satisfies every constraint, or unsatisfiable (UNSAT) if there is no satisfying assignment (commonly referred to as *frustrated* in physical treatments).

The corresponding *decision problem* of determining whether such an assignment exists, and, therefore, whether a particular instance is SAT or UNSAT is also referred to as SAT or satisfiablity with context generally determining which meaning is intended. The *optimization*

problem of determining an assignment which satisfies the maximum number of constraints (or maximum total weight) is referred to as MAXSAT. Determining the ground state of a system in statistical physics is thus equivalent to a MAXSAT instance.

Generally, the SAT problem on 2-variable OR and XOR constraints may be trivially solved. In the case of the Ising model, pick the value of any spin to be $+1$ and propagate this throughout the lattice where every spin value will be determined by its neighbor. If a contradiction is reached, the instance is unsatisfiable. If not, this will construct a satisfying assignment and the instance is equivalent to the ferromagnetic Ising model through a gauge transformation.

Despite this, the MAXSAT problem on two variable constraints may be quite difficult, depending on the structure of the instance. It is known that instances on a planar graph may be solved efficiently (in polynomial time) by a perfect matching algorithm [MKT17]. If the graph is non-planar as in the chimera graphs used by DWave [DBI$^+$16] or the 3-dimensional cubic lattice used for benchmarking here, there is no general efficient algorithm known, and the problem of finding an assignment is NP-Hard [Bar82]. This statement, however, only applies to the worst cases and for any individual instance, and especially for classes of randomly generated instances, one might hope that an efficient approach exists. Conversely, despite the fact that an efficient algorithm exists for planar instances, they may still present meaningful difficulty for a solver which only uses local information. Debate over these ideas have surrounded the benchmarking studies for DWave and discussions to this effect may be found in [KHZ$^+$15, MKT17, KLH15, HJA$^+$15].

4.2 'Rigidity' in a continuous dynamical system

The notion of 'rigidity' arises in several areas across physics and here we clarify what our intended meaning is in the context of continuous dynamical systems. The continuity of these systems can give rise to behaviors inaccessible to their discrete counterparts [DVTO17, DT18]. In particular, the presence of a continuous symmetry in the equations and its effective breakdown

can give rise to behavior analogous to zero-modes in statistical physics/field theory [PS95]. As a consequence, along some directions of the phase space the system can respond in a correlated, or 'rigid' manner in which large clusters of variables will transition together [DVTO17].

For example, in a lattice of continuous "spins" obeying,

$$\dot{v}_i = -|J_{ij}|\big(v_i - \mathrm{sgn}(J_{ij})\sigma(v_j)\big), \tag{4.12}$$

$$\sigma(x) = \begin{cases} 1, & x > 1 \\ x, & -1 \leq x \leq 1 \\ -1 & -1 < x. \end{cases} \tag{4.13}$$

the system will exponentially relax to a state in which every variable v_i takes the value $\mathrm{sgn}(J_{ij})v_j$ for all of its neighbors v_j. If the underlying lattice is ferromagnetic ($J_{ij} = 1$), then taking any spin to its limiting value $v_i = \pm 1$ will cause the entire lattice to transition with it in a manner analogous to long-range order. In contrast, for a discrete system in the ferromagnetic state $s_i = 1$, flipping a single spin will not cause the rest of the lattice to transition as it will not change the sign of any local fields. The ability of a local perturbation to flip large clusters of spins might benefit a solver attempting to satisfy a constraint while maintaining the satisfaction of its neighbors. However, the presence of unsatisfiable/frustrated constraints renders this impossible in the model above: in this case all spins will relax to $v_i = 0$ and pulling a single spin to ± 1 will not propagate through the lattice.

Any unsatisfiable spin-glass instance may be associated with one or several satisfiable instances formed by removing any unsatisfied bonds in the ground state (see the preceding Sec. 4.1 for a discussion of these terms in the context of computer science). In the main text we show that the inclusion of memory variables to the dynamical system (4.12) above restores some measure of the long-range order. These variables come in the form of constraint weights which act to isolate the associated satisfiable instance and bear similarities to strategies employed in discrete

constraint satisfaction [Mor93, SK93].

4.3 Generating Instances: Frustrated Loop Instances and Instance Tuning

The problem of benchmarking MAXSAT solvers is generally hindered by the fact that the problems are NP-hard, and, for an arbitrary instance, determining or even confirming a solution will require exponential time [Bar82]. For this reason, planted solution instances are commonly employed in which instances are generated such that they have a known solution [JMS04].

Benchmarking studies on quantum annealing have introduced the class of Frustrated Loop Hamiltonians [HJA$^+$15, KLH15] in which the total Hamiltonian is written as the sum of a set of loops containing a single frustrated bond (see schematic in Fig. 4.5),

$$H = \sum_i H_{FL,i}. \tag{4.14}$$

The loops are formed such that the planted solution minimizes all of the Hamiltonians $H_{FL,i}$ simultaneously, and so minimizes their sum.

In order to generate these instances, we first construct an underlying lattice which we take to be hypercubic in D-dimensions with periodic boundary conditions. Each loop is generated by beginning at a randomly selected site and performing a random walk until it crosses itself. The length, l, of the loop formed is generally required to be above some limit, otherwise it is rejected. For instance, the instances used by DWave use a loop length limit of $l \geq 8$ [HJA$^+$15, DBI$^+$16]. It is also noted that discarding the length limit seems to lead to very difficult instances, although an explanation for this feature is not understood. In our investigations of the instances, we found that discarding the loop length limit leads to instances of widely varying difficulty, and that both the uniformity of the difficulty, and the time to solution (measured with simulated annealing)

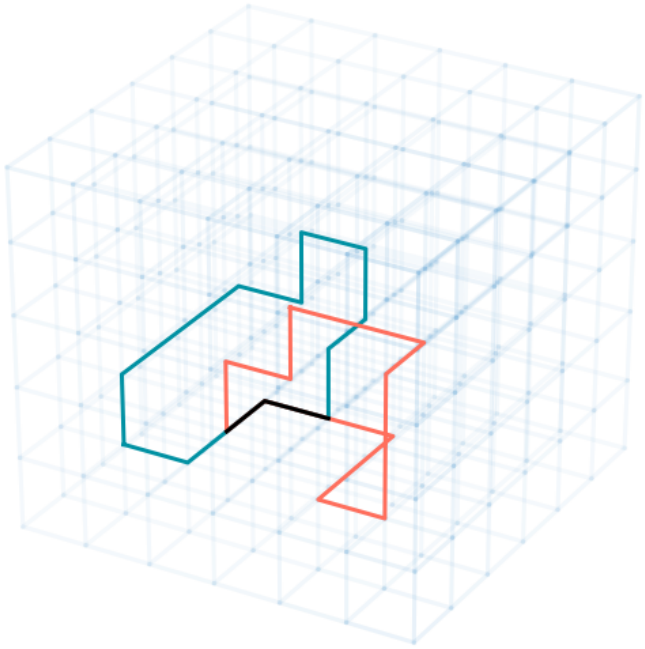

Figure 4.5: A schematic representation of instance creation. Separate frustrated loops (blue and red curves) are generated by random walking around the lattice until the walk crosses itself. Each loop has its own Hamiltonian consisting of $J_{ij} = 1$ for all bonds except one with $J_{kl} = -1$ such that the ground state of the loop will have one unsatisfied bond. When the loops are combined, overlapping bonds (shown in black) have a coupling J_{ij} which is the sum of the contributions from each separate loop.

decreased as the length limit increased. In order to avoid the complications of a widely varying difficulty, while generating the most difficult available instances, we chose a length limit of $l \geq 6$ for our generated instances.

In order to generate a loop, we consider planting the ferromagnetic solution $s_i = 1$. After generating an instance, any other solution may be hidden by means of a gauge transformation. All interactions in the loop are chosen to be ferromagnetic, $J_{ij} = 1$, except one which is selected at random to be anti-ferromagnetic $J_{ij} = -1$. The solution to the loop hamiltonian $H_{FL,i} = -\sum_{\langle ij \rangle \in l_i} J_{ij} s_i s_j$ is thus an assignment with one unsatisfied interaction.

The number of loops, M, generated must be proportional to the number of sites $N = L^D$ and may be characterized by a *density* α such that $M = \alpha N$. These instances are known to demonstrate a *hardness peak* in α such that the most difficult instances are generated when there are neither too few loops, in which case they do not overlap and each may be solved separately, nor too many, in which case the antiferromagnetic interactions tend to be canceled by the more numerous ferromagnetic interactions [HJA$^+$15, DBI$^+$16, KHZ$^+$15]. The value of α at the peak also tends to align with the amount of frustration in the instance, as measured by the number of unsatisfied interactions in the ground state.

In order to generate difficult instances, in $D = 2$ dimensions we used a simulated annealing solver to test instances across a range of α, finding that the most difficult instances lay at $\alpha \approx 0.2$, consistent with the results on the pseudoplanar chimera graphs in [HJA$^+$15]. For $D = 3$ dimensions, the optimal value of α was estimated using the amount of frustration in the instances as suggested in [HJA$^+$15] and found to lie at $\alpha \approx 0.3$.

4.4 Avalanche Detection and Correlation Calculations

As shown in the main text, when the timescale of the memristors is sufficiently separated from the voltage timescale, the system evolves in well defined events which may be interpreted as

avalanches (or instantons [DVTO17]). A set of sample trajectories, calculated on different spin glass instances ($L = 15$, $D = 2$, $\alpha = 0.2$) in the slow driving limit is displayed in Fig. 4.6. Within these trajectories avalanches are clearly discernible, but the timescales governing them seem to grow during the simulation and additional features such as quasi-periodic behavior (as seen in panel 4.6.c at beginning time 8000) also arise. Detecting these events in a robust manner poses a novel problem as we must extract a discrete event from a continuous system. Here we detail the method we used to extract these events and compute correlations during the largest avalanche.

The rate of change of the entire system may be concisely viewed through the magnitude of the voltage derivative vector, $\dot{\vec{v}}(t)$, normalized to the number of spins, $|\dot{\vec{v}}(t)|/N$ shown in Fig. 4.7. Avalanches manifest as sharp spikes in the magnitude of the derivative. However, as the simulation advances and variables begin transitioning together, the slowest timescale in the system tends to increase, making a simple threshold ineffective at separating the later clusters.

Instead, we first find the convex lower envelop of the total derivative shown in Fig. 4.7. This gives an estimate of how the slowest timescale in the system changes. If the slope at the end of the envelope exceeds a bound, its slope is extrapolated from the previous point to avoid errors due to the termination of the integration mid-avalanche.

The time interval of an avalanche is defined as a continuous period in which the magnitude of the derivative $|\dot{\vec{v}}(t)|/N$ exceeds the lower envelop by a threshold. Choosing this threshold is performed through tuning to the specific set of instances and will depend on system size, dimension and memristor timescale β. The threshold value t used for each set of correlation calculations are: $(L = 15, D = 2, t = 1 \times 10^{-4})$, $(L = 19, D = 2, t = 2.15 \times 10^{-5})$, $(L = 23, D = 2, t = 8 \times 10^{-6})$, and $(L = 8, D = 3, t = 4 \times 10^{-6})$. Within an avalanche interval, we define the variables included in the avalanche as those that changed sign and thus can affect the energy calculated from the orthant dynamics. A few of these configurations are shown in the main text and a more complete selection is displayed in Fig. 4.8.

Once a set of avalanches has been extracted from a simulation, we investigate the structure

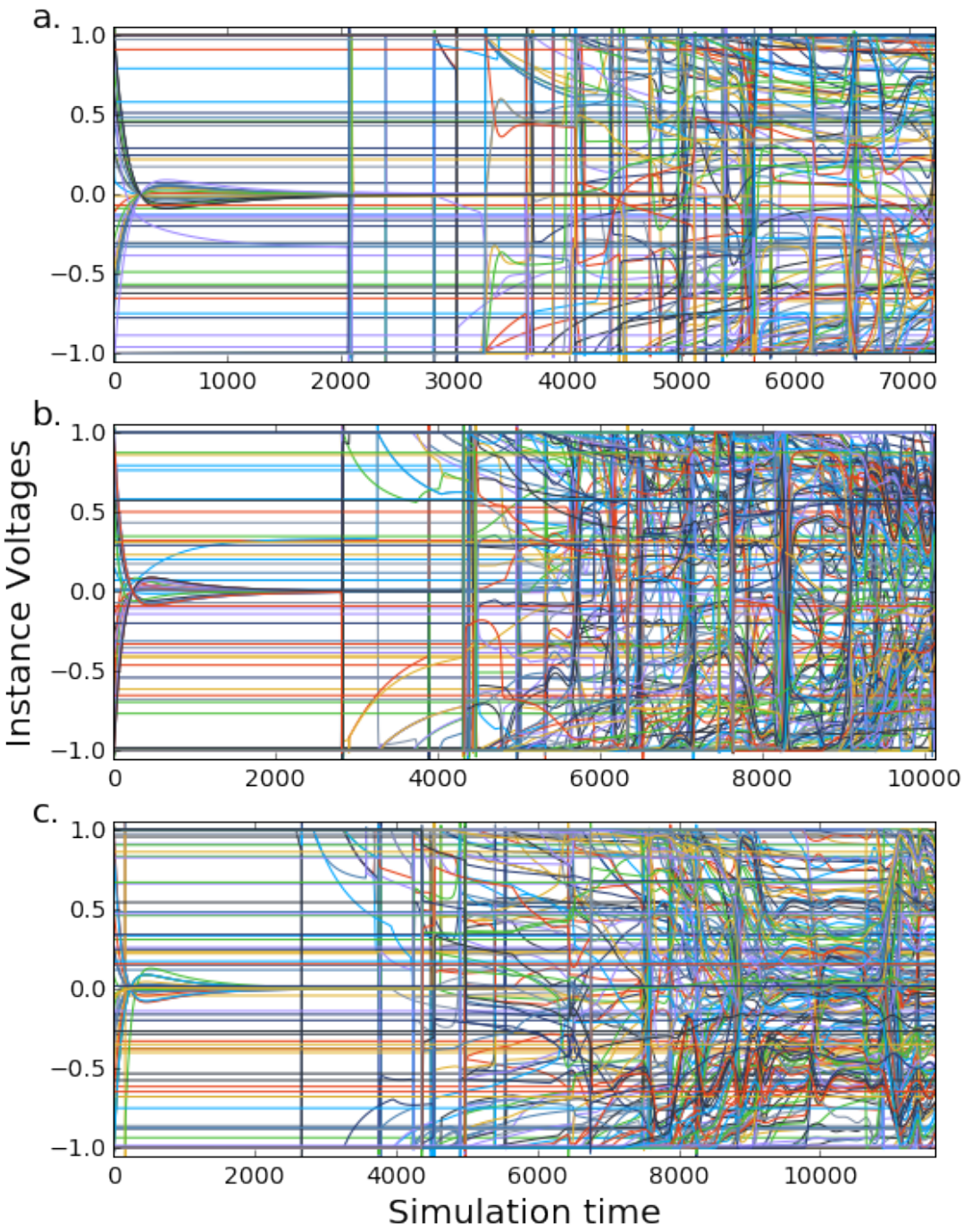

Figure 4.6: A sample of trajectories. Here we show the simulation results for the same initial conditions on three different instances ($L = 15$, $D = 2$, $\alpha = 0.2$). In (a.) we see well separated events that maintain their separation until the instance is solved. In (b.) the longer run results in lower memristor values and a slower voltage timescale, causing the width of the avalanches to grow. In (c.) we see markers of quasi-periodic behavior extending from ≈ 8000 to 10000. A scheme to detect avalanches must be robust to these effects in order to be accurate.

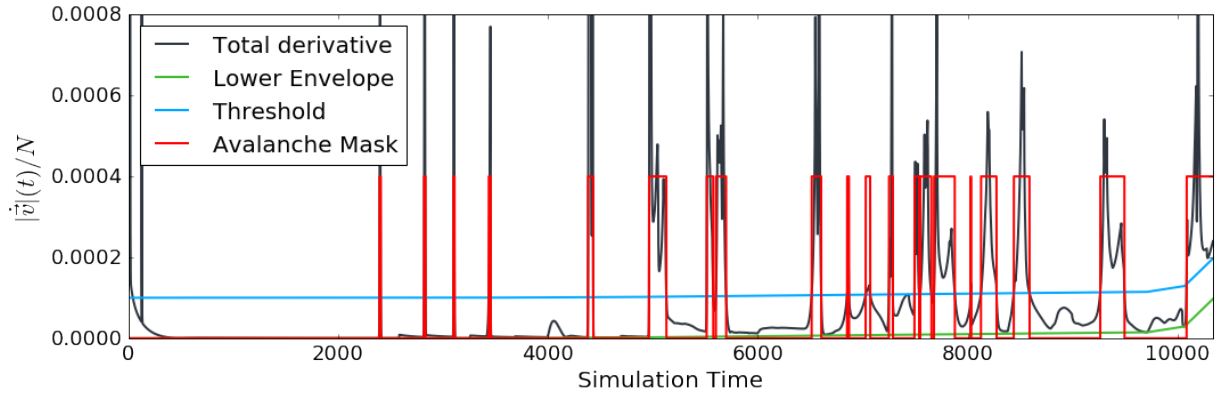

Figure 4.7: Extracting avalanches from a continuous trajectory. Avalanches display as sharp spikes in the total voltage derivative vector of the system. Here, we have displayed the magnitude of the total voltage derivative of the system normalized to the number of spins N for the trajectory displayed in the main text ($L = 15, D = 2\alpha = 0.2$). As the timescale of the avalanches can slow, sometimes dramatically, over the course of the simulation, the convex lower bound of the derivative is first calculated. An avalanche interval is defined as a continuous period in which the system exceeds a threshold above this envelope (here chosen as 0.0001). Voltages that change sign during an interval are included in the avalanche configuration as shown in Fig. 4.8.

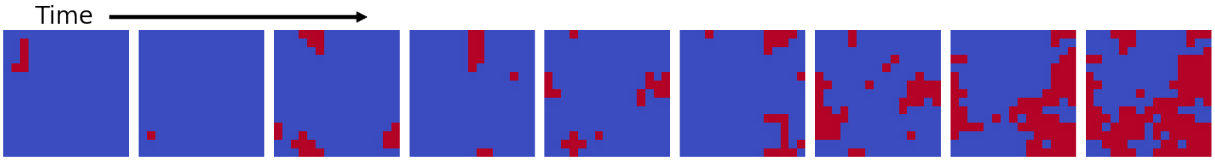

Figure 4.8: A sample of detected avalanches. For the trajectory shown in the main text and the clusters detected in Fig. 4.7 a representative sample of the detected avalanche configurations are displayed, including the first and last avalanches in the trajectory. Over the course of the simulation, the average size of the avalanches grows until it reaches an extensive set of spins which can span the entire lattice.

of the largest avalanche by calculating the probability that a spin lying a distance r away (measured in terms of lattice steps) from the first spin to flip is included within the avalanche. To this end, we define a cluster configuration as being $v_i = 1$ if the spin is included in the cluster and 0 otherwise. This acts as indicator variable which for an individual cluster allow us to calculate the probability that a spin a distance r away was flipped (recall that the lattices we generate are periodic and this distance is calculated as the minimum distance of a path between the two sites). This is then averaged across randomly generated instances and initial conditions of the solver.

The probability obtained may be interpreted as a correlation in the slow driving or instantonic limit in which the avalanche may be regarded as occurring at an instant in time, and calculated on the orthant dynamics of the system. As shown in the main text, the largest avalanche gives a finite probability for a spin anywhere in the lattice to change sign and is thus extensive.

4.5 Simulations and Solver Tuning

With the exception of IBM CPlex and Falcon, solvers were implemented in Python 2.7 using the NumPy and SciPy libraries [JOP$^+$]. Simulated Annealing, Parallel Tempering, CPlex and the model in the main text were run at UCSD on a single core of an Intel Xeon E5430 with 16 Gb RAM.

4.5.1 Model Simulations

In order to limit the voltages and memristors to the allowed regions and make them robust to numerical errors, the equations simulated were,

$$\dot{v}_i = B_{v_i,(-1,1)}\Bigg(\sum_{\langle ij\rangle} J_{ij}x_{ij}v_j - (1-x_{ij})\frac{|J_{ij}|}{2}\big(v_i - \mathrm{sgn}(J_{ij})v_j\big)\Bigg) \tag{4.15}$$

$$\dot{x}_{ij} = \beta B_{x_{ij},(0,1)}\Bigg(x_{ij}\big(1-x_{ij}\big) \times \Big(\frac{|J_{ij}|}{2}\big(1-\mathrm{sgn}(J_{ij})v_iv_j\big) - \gamma\Big)\Bigg) \tag{4.16}$$

where $B_{x,(l,h)}(\cdot)$ implements the bounds to ensure integration steps that leave the region return to the fixed points as

$$B_{x,(l,u)}(f) = \begin{cases} (u-x), & x > u, f > 0 \\ (l-x), & x < l, f < 0 \\ f, & \text{otherwise.} \end{cases} \tag{4.17}$$

Integrations are carried out using forward Euler with parameters tuned for each set of instances to maintain the slow driving limit, hence allowing for an easy identification of avalanches. This tuning is *not* required to *solve* an instance, but it is in order to detect clearly defined avalanches. First, it was determined through tuning that $\beta = \frac{1}{400}$ with a maximum time of $t_{max} = 25,000$ gave well defined avalanches for $L = 15$, $D = 2$ and usually solved instances near to $t = t_{max}/2$. In order to maintain this limit for larger instances, the memristor timescale was slowed, scaling with the inverse square of the number of spins. Slowing the memristor timescale requires increasing the maximum simulation time in the same way, such that t_{max} was scaled with the square of the number of spins. For $L = 15$, $\gamma = 0.65$ was found to give a well defined transition, but as instance size increased this value would lead to quasi-periodic behavior more often and γ was increased to

$L = 19, \gamma = 0.75$, $L = 23, \gamma = 0.85$ and for $L = 7$, $D = 3$ and $L = 8, D = 3, \gamma = 0.85$.

4.5.2 Simulated Annealing

Simulated annealing was implemented using a linear schedule in $\tilde{\beta}$ (the inverse temperature) from $\tilde{\beta}_i = 0.01$ to $\tilde{\beta}_f = \ln(N)$ where the low temperature was scaled such that excited states were suppressed as $\frac{1}{N}$ [Whi84]. At each step in the annealing a sweep of metropolis samples across the entire lattice was performed. For annealers to be most effective in a given computational time, the number of cooling steps (i.e., number of metropolis sweeps) performed on a single initial condition versus the number of initial conditions attempted must be optimized. As we are working with planted solution instances, runs consisted of continued repetitions at a fixed number of metropolis sweeps until the solution energy was encountered or some maximum allowed time was reached. Tuning was performed by running each instance with a varying number of temperature steps in the cooling schedule until the optimum was reached. This was performed on 1000 frustrated loop instances ($d = 3$, $\alpha = 0.3$, $l \geq 6$) for $L = 6$ through 11 and the time to solution for each run was recorded. The results of these runs are plotted in Fig. 4.9.

The time to solution for a given instance size and number of sweeps exhibited a bimodal distribution with a secondary peak of smaller measure approximately one order of magnitude above the primary peak. In order to estimate the optimal scaling, the 95th quantile of each sample distribution was found and the error estimated with a bootstrapping procedure (30,000 samples). As the peaks of the distributions were approximately lognormal, these were both calculated in logspace and the error estimated as 2σ of the sample distribution. As can be seen in Fig. 4.9, the presence of a secondary peak leads to large uncertainties in the region surrounding the cross-over of the 95th quantile. The optimal sweeps were found as the minimum of the upper bound uncertainty estimate.

Since the intention of this work is to examine the scaling for very large sizes, this analysis cannot be repeated for all sizes we intended to run. Instead, we used the values at small N to

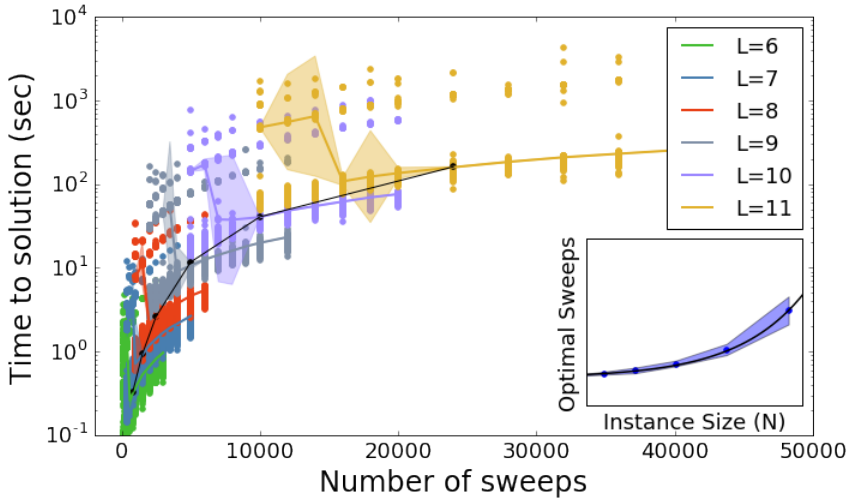

Figure 4.9: Tuning for simulated annealing. Simulated annealing runs were performed on 1000 3-D frustrated loop instances for $L = 6,7,8,9,10,11$ at varying number of metropolis sweeps. As can be seen, if the number of sweeps is too few the ground state is only rarely encountered. Beyond a certain amount, more sweeps will have a slightly negative effect on the runtime. We estimated the location of this crossover across lattice sizes with the dark curve, and extracted the scaling of the optimal runtime (inset) which was well fit by an exponential.

estimate the scaling of the optimal number of sweeps per repetition as shown by the inset curve in Fig. 4.9. This was well fit by an exponential dependence as,

$$\#sweeps_{opt} = (604)\exp(2.78 \times 10^{-3} N) \tag{4.18}$$

which was used to estimate the optimal number of sweeps for the scaling figure in the main text.

4.5.3 Parallel Tempering

The parallel tempering (PT) algorithm we employed utilized a ladder of temperatures geometrically spaced in T from $T_h = 10$ to $T_l = 1/\log N$ [Whi84, ED05]. One step of the algorithm consisted of a single sweep of metropolis sampling over all replicas, followed by a single proposed exchange where replicas at neighboring temperatures had their configurations

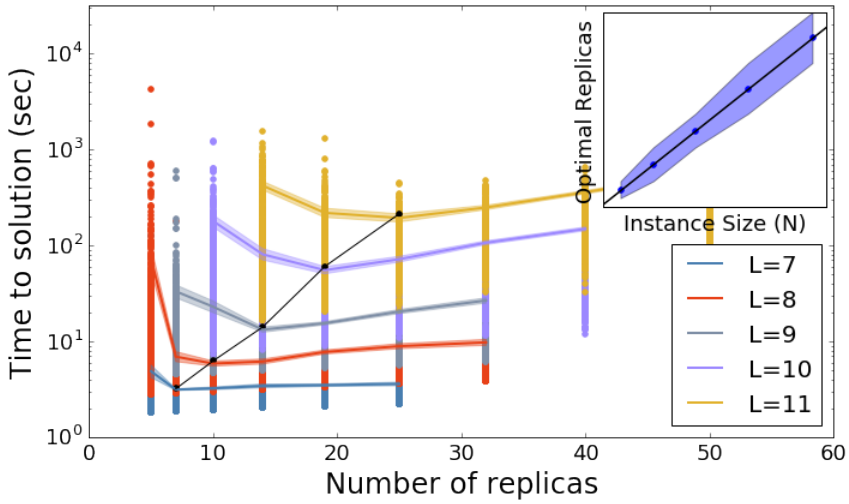

Figure 4.10: Tuning for parallel tempering. Parallel tempering runs were performed on 1000 frustrated loop instances for $L = 6,7,8,9,10,11$ with varying numbers of replicas. For sufficiently large sizes ($L > 6$) a clear optimum was observed in the time to solution, with the number of replicas at the optimum growing with the number of spins. The location of these optima were used to estimate a scaling law for the optimal number of replicas at larger sizes.

switched according to the probability,

$$P_{exchange} = \min\{1, \exp\left((\tilde{\beta} - \tilde{\beta}')(E - E')\right)\}, \tag{4.19}$$

where E and E' are energies of the instances. This cycle of metropolis sweeps and exchanges was repeated until the solution was reached. As the system size increases, the extensivity of the energy will cause exchanges to become less likely and so the density of temperatures simulated should be increased such that the optimal temperature set will balance the extra computational work of performing metropolis sweeps over the replicas with the diffusion of replicas across temperatures. In order to determine this point, the time to solution was found for 1000 frustrated loop instances (dimensions $D = 3$, density $\alpha = 0.3$, loop length $l \geq 6$) for $L = 6$ through 11 across a range of different numbers of replicas, as shown in Fig. 4.10. At each number of replicas and size, the 95th quantile of the sample was found and the error in this statistic was estimated from a bootstrap in logspace using a 2σ confidence interval. The optimal number of replicas was found as the

minimum of the upper bound error estimate. Across sizes these were found to be very well fit by a linear dependence,

$$\#replicas_{opt} = 0.0183N + 0.7. \tag{4.20}$$

This leads to a quadratic iscaling in the memory footprint of the algorithm.

4.5.4 CPlex

CPlex was run using the python API within the IBM ILOG CPLEX Optimization Studio version 12.7.1.0 under an academic license [cpl]. The QUBO (quadratic unconstrained binary optimization) form for the associated frustrated loop instance was found through the transformation to binary variables, $s_i = 2b_i - 1$ which leads to the correspondence,

$$E_{SG} = -\tfrac{1}{2}\textstyle\sum_{ij} J_{ij} s_i s_j \tag{4.21}$$

$$= \tfrac{1}{2}\textstyle\sum_{ij} Q_{ij} b_i b_j + C \tag{4.22}$$

$$Q_{ij} = \begin{cases} -4J_{ij}, & i \neq j \\ 4\sum_j J_{ij}, & i = j \end{cases} \tag{4.23}$$

$$C = -4\textstyle\sum_{ij} J_{ij}. \tag{4.24}$$

Within CPlex, problems in this form are first transformed to a mixed integer programming (MIP) form. Unlike the other solvers in this work, CPlex is a complete solver and will attempt a proof of optimality along with solving the instance. To prevent this, a callback was employed that terminates the search once the planted solution energy was found. Cuts were set to balance optimality and feasibility in the search.

4.5.5 Falcon

The memcomputing solver, Falcon, was implemented with MATLAB as specified in [TDV17] as a Boolean satisfiability solver which accepts instances in conjunctive normal form [GJ90]. The transformation to CNF shown in section I has been performed on the frustrated loop instances and simulations were carried out on a single core of an Intel Xeon 6148 with 192 GB RAM. The Falcon parameters were first tuned for the smallest size instances, and then the same parameters have been employed to solve all instances reported in Fig. 4 of the main text. Since Falcon integrates differential equations numerically (using forward Euler), it employs memory that scales linearly with problem size [SCTD].

4.6 Fitting

For each solver, instances were run at a variety of sizes in order to estimate the scaling of the 95th quantile time to solution. The annealers, SA and PT were run on 1000 instances for lattices $L \leq 11$ and 200 for $L > 11$. CPlex was run on 200 instances from $L = 10$ to $L = 25$ and Falcon was run on 200 instances for sizes from $L = 10$ to $L = 30$ and 100 for $L = 35, 40$. The error in the sample 95th quantile estimate was calculated with a 2σ estimate from a bootstrap (30,000 samples) performed in log space to account for the lognormal distribution of the runtimes.

Fitting of each solver was performed with the SciPy optimization library's curve fit and polyfit functions [JOP$^+$], which utilize least squares to fit a particular function. The time to solution for instances was found to be approximately log-normally distributed, as may be observed from the consistent standard deviations (0.5-1.5 orders of magnitude) found in log-space in Fig. 4.11. Fitting was thus performed in log-space on the sample 95th quantile time to solution. For each solver, fits were performed for a polynomial, $TTS \approx AN^B$ and exponential $TTS \approx ae^{bN^c}$ trend as is displayed in Fig. 4.11. For Falcon, an exponential fit did not converge and so only a polynomial fit is shown following $TTS \approx (7.5 \times 10^{-5}\, sec) N^{1.5}$.

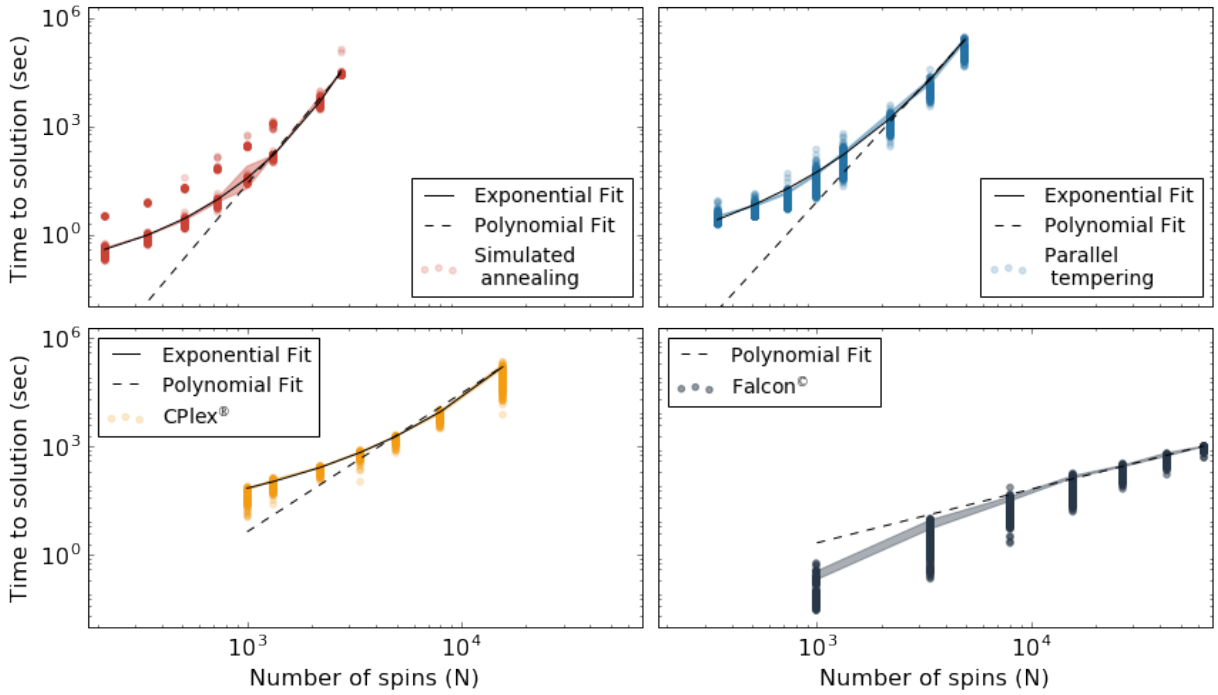

Figure 4.11: Fitting the Time to Solution. For each solver, the geometric average complexity of the instance class was estimated by fitting the 95th quantile time to solution for a set of frustrated loop spin glasses ranging in sizes across several orders of magnitude. Polynomial ($TTS \approx AN^B$) and exponential ($TTS \approx ae^{bN^c}$) trends were fit to the data to assess the performance of each approach. For Falcon, only the polynomial fit was found to converge and it demonstrated the lowest exponent of all fits ($B = 1.5$). All other solvers were found to scale exponentially.

For the other solvers, both a polynomial and exponential fit was found to converge but in each case the exponential fit is clearly favored. For SA, the time to solution was found to follow $TTS \approx (0.030\,\text{sec})e^{0.069N^{0.67}}$, for PT $TTS \approx (0.022\,\text{sec})e^{0.32N^{0.46}}$, and for CPlex $TTS \approx (3.5\,\text{sec})e^{0.12N^{0.46}}$. It is interesting to note that the exponential coefficients for each solver are quite close to rational values, with SA scaling close to $e^{bN^{2/3}}$ and PT and CPlex scaling close to $e^{bN^{1/2}}$.

Chapter 4, in full, is a reprint of the material which was submitted for publication as, Sheldon, Forrest; Traversa, Fabio L.; Di Ventra, Massimiliano. Taming a non-convex landscape with dynamical long-range order: Memcomputing the Ising spin-glass. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Dessert: Concluding Remarks and Future Directions

A primary aim of this research is to eventually leverage these systems to make interesting statements about complexity and vice versa. At the moment, this remains an ambition, but here I lay out an approximate roadmap towards its fulfillment. Throughout this thesis I have paid rather incomplete attention to the notion of complexity and I will now attempt to remedy this. For a brief review, I draw from the excellent text by Moore and Mertens [MM11] which can be consulted for further detail.

We regard the complexity of an *algorithm* as the leading order scaling of the solution time with the input size in the worst case (i.e. on the instances of the problem that scale most slowly). We can utilize this to define the *intrinsic complexity* of a *problem* as the complexity of the most efficient algorithm that solves it, where a *problem* consists of all possible instances of the specified form (e.g. SAT instances in conjunctive normal form). Any known algorithm provides an upper bound on this complexity but we do not generally know the most efficient possible algorithm.

We can then define the complexity class P, as consisting of those problems for which an

algorithm exists with complexity $O(n^c)$ for input size n and some constant c (i.e. whose solution time scales polynomially). We can also define the class NP (or *nondeterministic polynomial*) as those problem for which a proposed solution can be checked in time that scales polynomially. A problem in P is necessarily in NP as the algorithm constitutes a check. However, for a problem like SAT, any proposed solution can be checked in linear time, but we do not know any algorithm that exhibits polynomial complexity.

It was also noted that for many problems in the class NP there is a polynomial time algorithm to translate or *reduce* it to another problem in NP. This lead to the notion of NP-completeness which requires for a problem $A \in \mathsf{NP}$, that any problem $B \in \mathsf{NP}$ can be reduced to it in polynomial time. The problems in this class have equivalent complexity in the sense that, if a polynomial time algorithm exists for any of them, then it must exist for all of them. In this case, the classes P and NP would be equal. Similarly, if one of these problems exhibits exponential complexity, then they all do and $\mathsf{P} \neq \mathsf{NP}$. As discussed in Chapter 3, these notions can be generalized to the approximation of problems.

The definitions above conspire to make these notions almost entirely irrelevant in practical benchmarking studies. In particular, there is no notion of an NP-complete *instance*, although many times the words problem and instance are used interchangeably in the literature (though usually not in computer science). The notion of NP-completeness is better understood as a measure of *expressiveness* rather than difficulty. In particular, if a problem is NP-complete than many both 'easy' and 'hard' instances of the problem must exist.

In benchmarking, we almost never know what the worst case (i.e. hard) instances for our algorithms are, and in order to carry out testing, we must define some measure over the space of possible instances and draw from that (e.g. in random 3-SAT, we uniformly draw from the space of length 3 clauses over N variables). Our results are thus not relevant to the full problem (e.g. the complexity of 3-SAT). As discussed in Chapter 3, the structure of this measure can profoundly affect the difficulty of instances. If we are not careful, a random instance class for an

NP-complete problem is possibly quite easy to solve and thus not a useful benchmark.

A useful notion in these cases is *average-case complexity* which characterizes the average complexity of the most efficient algorithm over some class of instances defined by a measure. This would be a useful tool in constructing benchmarking instances, and is where the closest connections to physics have been found. While the most efficient algorithm for a class is still usually unknown, there are some indications that statistical features of the cost function may correlate to aspects of the behavior of algorithms. In some cases these are thought to be universal across all algorithms and thus may be relevant to the intrinsic average-case complexity of the instance class.

The current recipe for studying optimization problems using the machinery of statistical mechanics [MM09] is, for problem defined by a configuration $C \in \chi$ and cost function $E(C) \in \mathbb{R}$, to study the properties of the Boltzmann measure,

$$\mu_\beta(C) = \frac{e^{-\beta E(C)}}{Z(\beta)}. \tag{5.1}$$

In the limit $\beta \to \infty$, μ_β concentrates in the configurations that minimize E, from which information about the number and distribution of solutions may be recovered. At finite β, information about the free energy landscape can also yield the structure of local minima and critical points in the cost function.

A complementary and relatively recent approach is through applications of the Kac-Rice formula [BD07, AAČ13] which calculate the density of critical points of a cost function, $\rho(e, \alpha)$ at value e and with index α. Generally, these calculations show that for many problems, at high values of the cost function e, saddle points dominate the critical points until at a transition e_c, all directions become stable and local minima emerge.

In both cases, computation hardness (interpreted as requiring $O(e^{an^b})$ time) is attributed to the proliferation of critical points in the underlying cost function, and the behavior of algorithms

may be tied to their stability. For algorithms that operate directly on the cost function and Boltzmann measure, this argument is reasonable. However, for DMM type solvers the landscape they traverse is not $E(\{v_i\})$, but $E'(\{v_i\},\{x_\alpha\})$ that also includes the memory variables. Alternatively, we can view their motion as on a cost function $E_x(\{v_i\})$ similar $E(\{v_i\})$ but where the indices of the critical points are transformed by the motion of the memory variables x_α.

That we can write down a dynamical system for the same problem but with a different structure of its critical points calls into question whether such things may be regarded as intrinsic to the problem, or merely to the particular description (i.e. choice of cost function). In addition, if we trust that $\mathsf{P} \neq \mathsf{NP}$, then there must be some limitation to how far we can transform the structure of the cost function while maintaining the features that give rise to computational hardness. While I believe that this is the case, I see no reason for such a limitation in the language of dynamical systems. It is my hope that by understanding the limitations of DMMs, we can also understand how $\mathsf{P} \neq \mathsf{NP}$ manifests in the context of dynamical systems and the structure of critical points on their manifolds.

A further complication in applying these ideas to continuous dynamical systems is in defining how complexity is measured. For a continuous set of differential equations, we can rescale the time by an exponential and thus arbitrarily change the computation time. Reliable measures of complexity thus must take into account other features of the trajectory such as the arc-length [BGP17] and these may also manifest as chaos and instability in integration [ERT11]. A notion of complexity drawn from the distribution of critical points on a manifold would inherently incorporate these features.

In some sense, I think this is analogous to the position that thermodynamics was in during the debate over Maxwell's demon. $\mathsf{P} \neq \mathsf{NP}$ has a similar feeling of inviolability to the second law (at least to me), and in order to find an avenue of attack, Szilard posed simple devices that would *seem* to break the second law [Szi64]. In order to save it, he was forced into the beautiful idea of information entropy which has now become widely successful. His approach to this problem was

quite unique in physics and leveraged the freedom to engineer any imaginable system to strike at the proper idea.

I see some similarity in our freedom to define any equation of motion for a dynamical system navigating a cost function and the guarantee that such a system will always require exponential time in some cases. It is my hope that the structure of equations we have discussed in this thesis will be a starting point for this work going forward, and can now admit that my dedication to Szilard, who spent the end of his life around UCSD, was motivated by more than just a shared love of pastry.

Appendix A

Phase-dependent noise in Josephson junctions

The following chapter was published as:

Forrest Sheldon, Sebastiano Peotta and Massimiliano Di Ventra. Phase-dependent noise in Josephson junctions. *Eur. Phys. J. App.*, 81 (1), 10601 (2018)

and is included with the kind permission of The European Physical Journal (EPJ).

Abstract

In addition to the usual superconducting current, Josephson junctions (JJs) support a phase-dependent conductance related to the retardation effect of tunneling quasi-particles. This introduces a dissipative current with a memory-resistive (memristive) character that should also affect the current noise. By means of the microscopic theory of tunnel junctions we compute the complete current autocorrelation function of a Josephson tunnel junction and show that this memristive component gives rise to both a previously noted phase-dependent thermal noise, and an undescribed non-stationary, phase-dependent dynamic noise. As experiments are approaching ranges in which these effects may be observed, we examine the form and magnitude of these processes. Their phase dependence can be realized experimentally as a hysteresis effect and

may be used to probe defects present in JJ based qubits and in other superconducting electronics applications.

A.1 Introduction

The Josephson junction (JJ) [Jos62, Jos65, BP82] is the basic circuit element of superconducting electronics. Formed by a tunneling barrier between two superconductors, its primary feature is the nondissipative supercurrent $I_S = I_c \sin\gamma(t)$ (Josephson current), where I_c is the critical current and $\gamma(t)$ is the gauge-invariant phase difference across the barrier [Tin96]. In Josephson's original work [Jos62] it was shown that a JJ also supports a phase-dependent and dissipative current $I_M = G(\gamma)V$ with V the voltage drop across the junction.

This phase-dependent conductance (PDC) $G(\gamma) \propto \cos\gamma$ is often referred to as the 'cos' term and arises from the imaginary part of the superconducting response function [Jos62, BP82, Lik86]. As such, it has been interpreted as a consequence of the retarded phase-current response [Har74, Har75, Har76, ZKLS79], or as an interference effect between quasiparticle and Cooper pair currents [Tin96, Jos62]. It is illustrated in fig. A.1 as the breaking of pairs into quasiparticles which tunnel through the barrier and re-form [Ste69, BP82]. Irrespective of interpretation, it is a memory resistive (memristive) component since it gives rise to hysteretic behavior under specific driving conditions [PDV14b].

Initial interest in the PDC was driven by a discrepancy between the value given by BCS theory [BCS57] and that measured in experiments [PFL72, SMP77, VD74, RD76, NW75, FPT73, Lik79]. Several effects may account for the discrepancy (see, *e.g.*, ref. [ZKLS79]) and most attribute its value to a dependence on microscopic details of the junction such as the presence of impurities.

Recently, the PDC has again attracted interest for its role in the quasiparticle-induced decoherence of superconducting qubits [MOL$^+$99, LG07, MSS01, MAA09, LWB$^+$11, CSDG11,

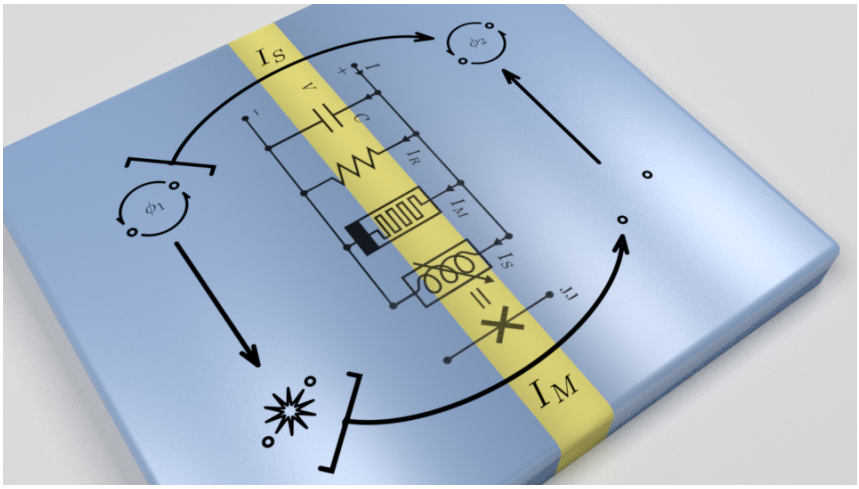

Figure A.1: A schematic representation of a JJ illustrating the two pair tunneling processes and the RSJ model equivalent circuit. Pairs in the left electrode with phase ϕ_1 may tunnel to the second either directly as shown by the upper arrow (I_S) or by breaking, tunneling as separate quasiparticles and re-pairing with phase ϕ_2, as shown by the lower arrow (I_M). In the low voltage/frequency limit, the TJM model is well approximated by the RSJ model given in equation (A.6) and shown across the tunneling barrier (yellow). This consists of a displacement current, $C\frac{dV}{dt}$, resistive current I_R, supercurrent I_S and a memristive component $I_M = \varepsilon G_L V \cos\gamma$.

YN05, CW08, DS13]. This has been discussed theoretically in refs. [CKF$^+$11, LMS11] and in an experiment on fluxonium qubits [PGC$^+$14]. In a recent publication [PDV14b], two of us (SP and MD) have proposed a two-junction interferometer to isolate the PDC from the nondissipative pair current, allowing for a more detailed study of its properties and extraction of its hysteretic features. An analogous phase-dependent term was also recently measured in the thermal current through a JJ [GMP12].

The fluctuation-dissipation theorem suggests that the PDC should lead to a similar phase-dependent current noise. With the possibility of isolating the PDC and the accessibility of electronic measurements reaching into the Josephson frequency range, such non-stationary noise processes may become important for the interpretation of experiments and technological applications. For Josephson junctions used as detectors, the action of thermal and non-Gaussian noise sources is an area of active research [AVS10, GVS15] and the introduction of non-stationary noise processes may have interesting dynamical consequences on the lifetime of the supercon-

ducting state. Solitons forming in long junctions [VGS14, SGM$^+$17] may also be affected due to the non-uniformity of fluctuations in the vicinity of the kink/anti-kink. In the interferometer mentioned above, measurement of phase-dependent noise would allow the determination of the junction phase without access to the supercurrent or applying a voltage, thus providing an avenue for nondestructive readout of the device state. As mentioned above, the magnitude of the PDC and the resulting noise is influenced by the presence of defects in the junction. Measurement of these effects can thus form a probe into the detailed structure of the junction.

In this article, by means of the microscopic theory of tunnel junctions we compute the complete current autocorrelation function of a Josephson tunnel junction. The resulting function contains a modulation which, in appropriate limits, takes a form $\propto \cos\gamma(t)$. We pay particular attention to the effects of corrections to the BCS theory on the subgap current response, and demonstrate that the expected noise variation due to the phase-dependent dissipative current is comparable to the averaged noise present at frequencies below the gap frequency $\omega_g = \frac{2\Delta}{\hbar}$ and thus we expect it to be detectable in experiments.

That such a modulation exists has been noted previously in ref. [RS74] but in subsequent considerations the time average of the spectrum has been taken, and the phase dependence was assumed to vanish. However, in the thermal limit the phase of the junction may be kept approximately stationary and does not necessarily vanish. Phase-dependent thermal noise has also been predicted in quantum point contacts [MRLYGV96] although the form is quite different from that expected in tunnel junctions as the dominant contribution comes from bound states whose energies lie within the gap. We thus suspect a similar role may be played by the sub-gap currents due to impurities in a tunnel junction. While our treatment of phase dependent noise is restricted to the tunneling regime, we expect the phenomenon to be generic across a wide range of JJ devices, differing somewhat in form. For example, the low transparency limit of the thermal conductivity of a weak-link shows a $\cos\gamma$ dependence with the addition of a secondary phase dependent term [ZLS03].

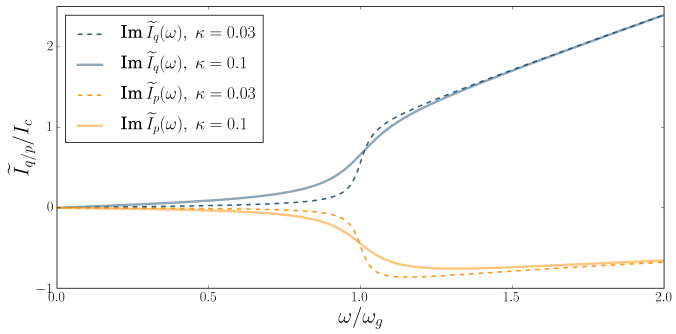

Figure A.2: The Fourier transforms $\mathrm{Im}\widetilde{I}_p(\omega)$ and $\mathrm{Im}\widetilde{I}_q(\omega)$ of the pair and quasiparticle response functions at $T = 0$ given by equations (A.3) and (A.4). They are plotted in units of the gap frequency $\omega_g = \frac{2\Delta}{\hbar}$ and the critical current I_c for $\kappa = \frac{\tau_g}{2\tau_r} = 0.03, 0.1$. Note that the current response below the gap frequency is enhanced with κ.

In biased junctions, due to the unavoidable time averaging, a phase-dependent power spectrum cannot be defined for frequencies less than the Josephson frequency, $\omega_J = \frac{2e}{\hbar}V$. We demonstrate instead that a phase dependence can still be expected in the limit $\omega > \omega_J$. Here we distinguish between the subgap, $\omega < \omega_g$, and quantum noise, $\omega > \omega_g$, regimes and calculate the expected phase dependence in each. As the Josephson frequency is in the GHz range, measurement of this phase dependence will require both high-frequency and short-time resolution. Experimental systems designed for probing the quantum noise limit of mesoscopic systems have reached frequency-resolved measurements on the order of 100 GHz [DOGK03] and could potentially be adapted to the detection of such non-stationary processes.

A.2 Phase-dependent conductance

In order to demonstrate the existence of phase-dependent noise in JJs we make use of the tunnel junction microscopic (TJM) model [Lik86, Har74, Ste69] with a phenomenological factor which smoothes the energy gap edge. This correction may be attributed to the presence of small metallic layers within the junction and accounts for several deviations from the BCS approximation observed in experiments.

The dynamics of a generic low-transparency JJ are well described by second-order perturbation theory in the tunneling matrix elements resulting in the tunnel junction microscopic (TJM) model [Lik86]. The total current $I = I_{pair} + I_{qp}$ is the sum of the pair current I_{pair} and quasiparticle current I_{qp} given by,

$$I_{pair}(t) = \int_{-\infty}^{t} dt' I_p(t-t') \sin\left(\frac{\gamma(t)+\gamma(t')}{2}\right) \tag{A.1}$$

$$I_{qp}(t) = \int_{-\infty}^{t} dt' I_q(t-t') \sin\left(\frac{\gamma(t)-\gamma(t')}{2}\right). \tag{A.2}$$

The time-dependent phase $\gamma(t)$ results from a voltage drop across the junction, $\frac{d\gamma}{dt} = \frac{2e}{\hbar}V(t)$.

The material properties of the junction and superconducting electrodes are represented in the pair and quasiparticle response functions $I_p(t)$ and $I_q(t)$. We choose the form given by BCS theory at $T = 0$ and with superconducting energy gaps $\Delta_1 = \Delta_2 = \Delta$. In this case, the response functions have closed form,

$$I_p(t) = -\frac{2I_c}{\tau_g} J_0\left(\frac{t}{\tau_g}\right) Y_0\left(\frac{t}{\tau_g}\right) \exp\left(-\frac{t}{\tau_r}\right) \tag{A.3}$$

$$I_q(t) = \frac{2I_c}{\tau_g} J_1\left(\frac{t}{\tau_g}\right) Y_1\left(\frac{t}{\tau_g}\right) \exp\left(-\frac{t}{\tau_r}\right) - \frac{\hbar G_N}{e}\delta'(t). \tag{A.4}$$

where J_n, Y_n are the Bessel functions of the first and second kind, δ' is the derivative of the delta function, $\tau_g = \frac{\hbar}{\Delta}$ the gap timescale, G_N is the normal junction conductance, and we have included the phenomenological factor $\exp(-t/\tau_r)$ mentioned above which cuts off the algebraic decay of the Bessel functions at $t > \tau_r$ [Lik86]. This is also known as a Dynes parameter [DNG78] and is a standard way of introducing a cutoff. Further remarks on the introduction of the cutoff may be found in [PDV14b].

For our purposes it will be sufficient to consider a constant voltage so that the presence of the PDC $\propto \cos\gamma(t)$ may be explicitly shown. The phase advances linearly in time, $\gamma(t) = \omega_J t + \gamma_0$ where we have defined the Josephson frequency $\omega_J = \frac{2eV}{\hbar}$. The junction current may then be

written as

$$I(t) = \mathrm{Re}\,\widetilde{I}_p\left(\frac{\omega_J}{2}\right)\sin\gamma(t) - \mathrm{Im}\,\widetilde{I}_p\left(\frac{\omega_J}{2}\right)\cos\gamma(t) + \mathrm{Im}\,\widetilde{I}_q\left(\frac{\omega_J}{2}\right) \tag{A.5}$$

where $\widetilde{I}_{p/q}$ are the Fourier transforms of the response functions. These may be evaluated in terms of elliptic integrals and have been plotted in fig. A.2 for several values of $\kappa = \frac{\tau_g}{2\tau_r}$. The range $\kappa = 0.03 - 0.1$ gives reasonable values of the peak broadening [Lik86] and we note that the response for $\omega < \omega_g$ of $\mathrm{Im}\widetilde{I}_{p/q}$ increases with κ. It is convenient to define the conductances $\mathrm{Im}\widetilde{I}_q(\frac{\omega_J}{2}) = \sigma_0(V,T)V$ and $\mathrm{Im}\widetilde{I}_p(\frac{\omega_J}{2}) = \sigma_1(V,T)V$ from which we define the leakage conductance $G_L = \lim_{V\to 0}\sigma_0(V,T)$ and the ratio $\varepsilon = \lim_{V\to 0} -\frac{\sigma_1(V,T)}{\sigma_0(V,T)}$. Including the effects of a finite capacitance and fluctuations, the total junction current may then be written as,

$$I(t) = C\frac{dV}{dt} + G_L V(1 + \varepsilon\cos\gamma(t)) + I_c\sin\gamma(t) + I_F(t). \tag{A.6}$$

The equivalent circuit to equation (A.6) is given in fig. A.1 and the TJM model is thus well approximated by the resistively shunted junction model (RSJ) [Lik86] with the phase dependent conductance $G_L(1 + \varepsilon\cos\gamma)$.

The ratio ε has been investigated in a number of experiments on tunnel junctions [PFL72, SMP77], weak links [VD74, RD76], and point contacts [NW75, FPT73], consistently finding $\varepsilon \sim -1$ at low temperatures in disagreement with BCS theory which predicts $\varepsilon > 0$. This discrepancy may be accounted for by including frequency broadening in the BCS result [Lik86] as we have done with the exponential factors in equations (A.3) and (A.4). This may be physically attributed to a finite quasiparticle lifetime, gap anisotropy and renormalization [BP82]. The resulting sign and magnitude of ε is regularization dependent, varying from -1 to 1, indicating that the specific microscopic details leading to the broadening exert a strong influence on the

quasiparticle current, as can be seen in the enhanced subgap response in fig. A.2. While the particular sign and value of ε are not essential for our results, we emphasize that both theory and experiment place the magnitude of the phase dependence to be on par with the dissipative current itself.

A.3 Fluctuations

The PDC should provide a contribution to the current fluctuations according to the fluctuation-dissipation theorem. The autocorrelation function of the noise current $I_F(t)$ can be calculated from the microscopic theory in the case of an arbitrary phase dynamics $\gamma(t)$ [Zor81, Lik86]. In the simple case of DC voltage bias the autocorrelation function of the noise current reads

$$\begin{aligned}\langle I_F(t)I_F(t')\rangle_S = \frac{e}{4\pi}\int_{-\infty}^{+\infty} d\omega e^{i\omega(t-t')}\coth\frac{\hbar(\omega+\frac{\omega_J}{2})}{2k_BT} \\ \times\left[\mathrm{Im}\widetilde{I}_q(\omega+\tfrac{\omega_J}{2})+e^{-i(\gamma_0+\omega_J t')}\mathrm{Im}\widetilde{I}_p(\omega+\tfrac{\omega_J}{2})\right] \\ +\left\{\begin{array}{c}\gamma_0\to-\gamma_0\\ \omega_J\to-\omega_J\end{array}\right\}.\end{aligned} \tag{A.7}$$

where we denote the symmetrized autocorrelation function,

$$\langle I_F(t)I_F(t')\rangle_S = \frac{1}{2}\left\langle I_F(t)I_F(t')+I_F(t')I_F(t)\right\rangle. \tag{A.8}$$

As expected, the PDC provides a contribution to the fluctuations given by the term proportional to the Fourier transform $\mathrm{Im}\widetilde{I}_p(\omega)$ (see fig. A.2) modulated by the phase factor $e^{-i(\gamma_0+\omega_J t')}$. Due to the modulating factor, the autocorrelation function is not simply a function of the time difference $t-t'$, which means that *the fluctuating current is not a stationary stochastic process* [vK07]. The autocorrelation function is only invariant under discrete time translations $\langle I_F(t)I_F(t')\rangle_S =$

$\langle I_F(t+2\pi/\omega_J) I_F(t'+2\pi/\omega_J)\rangle_S$.

As the noise is non-stationary, it cannot be characterized by a power spectrum. We introduce a quadratic time frequency representation (TFR) defined by

$$TFR(\omega,t) = \int_{-\infty}^{\infty} \frac{d\tau}{2\pi} e^{-i\omega\tau} \langle x(t+\tau) x^*(t)\rangle_S \tag{A.9}$$

which respects time and frequency shift covariance [HBB92]. The averaging of the correlation function is performed for a fixed time t over the thermal and quantum ensemble of states of the electrodes. The nonstationarity of the noise is reflected in the time dependence t about which the decay of the correlation function localizes the noise statistics. This is similar to the introduction of a windowing function in a spectrogram, except that the windowing is naturally supplied by the correlation function and will determine in which regimes we can observe a phase dependent noise (see sect. A.3.2). The resulting spectrum should display a phase dependence $\propto \cos\gamma(t)$ for frequencies larger than the Josephson frequency ω_J.

From (A.7) the TFR reads

$$\begin{aligned} TFR(\omega,t) = {} & \frac{e}{4\pi} \coth \frac{\hbar(\omega+\frac{\omega_J}{2})}{2k_B T} \\ & \times \left[\mathrm{Im}\widetilde{I}_q(\omega+\tfrac{\omega_J}{2}) + e^{-i(\gamma_0+\omega_J t)} \mathrm{Im}\widetilde{I}_p(\omega+\tfrac{\omega_J}{2}) \right] \\ & + \left\{ \begin{array}{c} \gamma_0 \to -\gamma_0 \\ \omega_J \to -\omega_J \end{array} \right\}. \end{aligned} \tag{A.10}$$

From here we examine the thermal and dynamic ("shot") noise limits of this expression.

A.3.1 Thermal noise

The contribution to the noise power spectrum at *zero bias* given by the PDC is discussed in ref. [RS74]. However, after its introduction the *time-averaged* autocorrelation function was

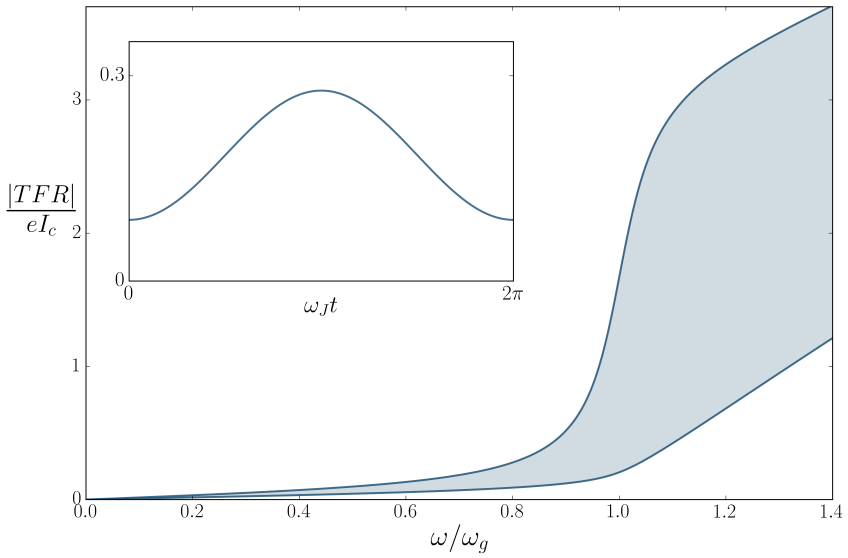

Figure A.3: $TFR(\omega,t)$ plotted across a period of the Josephson current. The Josephson frequency and κ have been chosen for the intermediate values $\omega_J = 0.005\omega_g$, $\kappa = 0.05$. The spectrum is plotted in units of the gap frequency ω_g and the critical current times the electron charge eI_c. The inset shows the variation of the spectrum across a period at $\omega = 0.05\omega_g$, in which the $\cos\gamma(t)$ dependence is clearly visible. The functional form derived in eq. (A.13) applies for $\omega \ll \omega_g$. At frequencies above the gap, the phase dependence is strongly enhanced.

considered and, because of the modulating factor, the phase-dependent noise averages to zero at finite bias. From our expression (A.10), we consider the limit $\hbar\omega$, $eV \ll k_B T$,

$$TFR(\omega,t) = \frac{1}{2\pi} 2k_B T G_L \left[1 + \varepsilon \cos\gamma(t)\right] \tag{A.11}$$

which leads to the autocorrelation function

$$\left\langle I_F(t) I_F(t') \right\rangle_S = 2k_B T G_L (1 + \varepsilon \cos\gamma(t')) \delta(t - t') \,. \tag{A.12}$$

This limit is appropriate for $|t - t'| \gg \frac{\hbar}{k_B T}$. We thus assume eq. (A.12) to be valid for a sufficiently slowly varying function $\gamma(t)$. This is justified since in this case the system is never driven too far away from thermal equilibrium and the quasiparticle relaxation occurs on a small time scale of the order of $\tau_g = \frac{\hbar}{\Delta} \sim \frac{\hbar}{k_B T}$.

A.3.2 Dynamic noise

We now consider the noise across a biased junction. In the thermal noise limit, the junction retains some memory of the initial conditions but over long times, the stationary phase leads to stationary noise. In the biased case, the averaging of the noise at low frequencies limits the range in which we can observe a phase dependence. In particular, if we consider the typical shot-noise limit $\hbar\omega_J \gg k_B T, \hbar\omega$ we find $TFR(\omega,t) = \frac{1}{2\pi} eVG_L\left[1+\varepsilon\cos\gamma(t)\right]$ however, the time averaging of the noise for $\omega \ll \omega_J$ makes the measurement of the phase dependence impossible at present. We instead consider the frequency range $\omega_J \ll \omega \ll \omega_g$ where the approximate values of $\omega_J \approx 10^9$Hz, and $\omega_g \approx 10^{12}$Hz place ω in the region $10^{10} - 10^{11}$Hz. Considering the limit in which $k_B T \to 0$ we may write the TFR to lowest order in ω_J as

$$TFR(\omega,t) = \frac{1}{2\pi}\hbar\omega G_L\left[1+\varepsilon\cos\gamma(t)\right]. \tag{A.13}$$

While this resembles a phase dependent quantum noise, it applies in the limit $\omega \ll \omega_g$ where the current response is due to the regularization and this term should be reserved for the behavior at $\omega \gg \omega_g$ where the spectral function indeed increases linearly with $\hbar\omega$ as we would expect for the quantum noise limit. The full spectrum has been plotted in fig. A.3 across a period of the Josephson current. Intermediate values have been chosen for the Josephson frequency $\omega_J = 0.005\omega_g$, and the regularization $\kappa = 0.05$. In the inset, its variation with time shows the $\cos\gamma(t)$ dependence of eq. (A.13). The subgap noise amplitude shows a linear dependence on κ in the range of interest. While the form in (A.13) applies for $\omega \ll \omega_g$, fig. A.3 shows that in the true quantum noise limit $\omega > \omega_g$ the phase dependence of the noise is strongly enhanced. In the conclusion, we describe an experimental arrangement to detect the phase dependence in this regime.

A.4 Conclusions

We have shown that the dissipative $\cos\gamma(t)$ term in the Josephson response gives rise to non-stationary noise whose magnitude displays a similar $\cos\gamma(t)$ dependence in both a previously noted thermal regime and a previously undescribed dynamic regime. In an unbiased junction, this gives rise to a modulation of the normal thermal noise present at low frequencies with a magnitude of variation comparable to that of the thermal noise itself. The presence of phase-dependent thermal noise in the junction has not yet been observed experimentally.

In a biased junction, experimental confirmation requires fluctuation measurements at very high frequency $\omega \sim 10^{10} - 10^{11}$Hz with a temporal resolution less than a period of the Josephson phase $\Delta t \sim 10^{-9}$s. While these conditions pose an experimental challenge we do not consider them unreasonable and sketch a possible experimental approach towards their measurement. To this end, we propose an arrangement along the lines of ref. [DOGK03] which utilizes a superconductor-insulator-superconductor (SIS) junction as an on-chip spectrum analyzer when capacitively coupled to the component of interest. A JJ which intentionally contains impurities and with $\Delta_{JJ} < \Delta_{SIS}$, will exhibit fluctuations that dominate those of the SIS junction and which are modulated by the $\cos\gamma(t)$ dependence. Through the capacitive coupling, these fluctuations in the JJ would then subject the SIS junction to microwaves which will induce quasiparticle tunneling.

The spectrum of these fluctuations may be read by biasing the SIS junction and measuring the resulting quasiparticle current. As the total current is determined by the number of microwave photons above the tunneling energy gap $2\Delta_{SIS} - eV_{SIS}$ the spectrum is given by the change in the quasiparticle current as V_{SIS} is varied. By tuning V_{SIS} to be sensitive to photons just above the JJ gap, $\hbar\omega_g \approx 2\Delta_{SIS} - eV_{SIS}$, phase-dependent noise may be observed as a modulation of the current through the SIS detector at the Josephson frequency of the first. By taking advantage of the difference in the gaps, fluctuations above the JJ gap where quasiparticle currents and the

phase dependence are enhanced, (see fig. A.3) may be measured in the range of the SIS detector. Such an arrangement will require finite-frequency driving of the original junction, leading to side-bands in the noise measurement.

A measurement of the phase dependent noise would provide another avenue for accessing ε in addition to the direct measurements of the PDC. A more reliable determination of ε will allow for a better understanding of the effects which determine its size and magnitude, giving a lens into quasiparticle processes and disorder in these systems. Since obtaining long decoherence times is a primary obstacle in building useful superconducting qubits, we expect quasiparticle effects to be of increasing importance in technological applications. The presence of a phase dependent term in the dissipative current and noise may also have consequences for the lifetime of the superconducting state and motion of solitons in JJ used as detectors [VGS14, SGM$^+$17] and could lead to interesting dynamical effects. Lastly, in the interferometer mentioned previously [PDV14b] the supercurrent is confined to the superconducting loop, and thus is not accessible to determine the junction phase. The presence of phase-dependent thermal noise means that a noise measurement would be sufficient to determine the device phase, allowing for nondestructive readout of a superconducting memristor.

A.5 Authors contributions

All the authors were involved in the preparation of the manuscript. All the authors have read and approved the final manuscript.

A.6 Acknowledgements

This work has been supported by DOE under Grant No. DE-FG02-05ER46204 and the Center for Magnetic Recording Research at UCSD.

Appendix A, in full, is a reprint of the material as it appears in European Physical Journal Applied Physics Volume 81 Sheldon, Forrest; Peotta, Sebastiano; Di Ventra, Massimiliano, and is included with the kind permission of The European Physical Journal (EPJ). The dissertation author was the primary investigator and author of this paper.

Bibliography

[AAČ13] Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013.

[Abr18] Isaac Abraham. The case for rejecting the memristor as a fundamental circuit element. *Scientific reports*, 8(1):10972, 2018.

[AH15] Steven H Adachi and Maxwell P Henderson. Application of quantum annealing to training of deep neural networks. *arXiv preprint arXiv:1510.06356*, 2015.

[Arg68] F Argall. Switching phenomena in titanium oxide thin films. *Solid-State Electronics*, 11(5):535–541, 1968.

[ASMO$^+$12] Audrius V Avizienis, Henry O Sillin, Cristina Martin-Olmos, Hsien Hang Shieh, Masakazu Aono, Adam Z Stieg, and James K Gimzewski. Neuromorphic atomic switch networks. *PLoS One*, 7(8):e42772, jan 2012.

[AVS10] Augello, G., Valenti, D., and Spagnolo, B. Non-gaussian noise effects in the dynamics of a short overdamped josephson junction. *Eur. Phys. J. B*, 78(2):225–234, 2010.

[Bar82] Francisco Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.

[BCS57] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Theory of superconductivity. *Phys. Rev.*, 108:1175–1204, Dec 1957.

[BD07] Alan J Bray and David S Dean. Statistics of critical points of gaussian fields on large-dimensional spaces. *Phys. Rev. Lett.*, 98(15):150201, 2007.

[Bet35] H A Bethe. Statistical Theory of Superlattices. *Proc. R. Soc. London A*, 150(871):552–575, 1935.

[BGP17] Olivier Bournez, Daniel S Graça, and Amaury Pouly. Polynomial time corresponds to solutions of polynomial ordinary differential equations of polynomial length. *Journal of the ACM (JACM)*, 64(6):38, 2017.

[BHL+02] Wolfgang Barthel, Alexander K Hartmann, Michele Leone, Federico Ricci-Tersenghi, Martin Weigt, and Riccardo Zecchina. Hiding solutions in random satisfiability problems: A statistical mechanics approach. *Physical review letters*, 88(18):188701, 2002.

[BMTD18] S. R. B. Bearden, H. Manukian, F. L. Traversa, and M. Di Ventra. Instantons in self-organizing logic gates. *Physical Review Applied*, 9:034029, 2018.

[BP82] A. Barone and G. Paterno. *Physics and applications of the Josephson effect*. Wiley, 1982.

[CB18] F Caravelli and P Barucca. A mean-field model of memristive circuit interaction. *EPL (Europhysics Letters)*, 122(4):40008, 2018.

[CC18] Francesco Caravelli and Juan Carbajal. Memristors for the curious outsiders. *Technologies*, 6(4):118, 2018.

[Chi10a] Dante R. Chialvo. Emergent complex neural dynamics. *Nat. Phys.*, 6(10):744–750, oct 2010.

[Chi10b] Dante R Chialvo. Emergent complex neural dynamics. *Nat. Phys.*, 6(10):744, 2010.

[CHP98] Kenneth Steiglitz Christos H. Papadimitriou. *Combinatorial Optimization*. Dover Publications Inc., 1998.

[Chu71] Leon Chua. Memristor-the missing circuit element. *IEEE Transactions on circuit theory*, 18(5):507–519, 1971.

[CK76] Leon O Chua and Sung Mo Kang. Memristive devices and systems. *Proceedings of the IEEE*, 64(2):209–223, 1976.

[CKF+11] G. Catelani, J. Koch, L. Frunzio, R. J. Schoelkopf, M. H. Devoret, and L. I. Glazman. Quasiparticle relaxation of superconducting qubits in the presence of flux. *Phys. Rev. Lett.*, 106:077002, Feb 2011.

[CLK+08] Seung Chul Chae, Jae Sung Lee, Sejin Kim, Shin Buhm Lee, Seo Hyoung Chang, Chunli Liu, Byungnam Kahng, Hyunjung Shin, Dong-Wook Kim, Chang Uk Jung, et al. Random circuit breaker network model for unipolar resistance switching. *Advanced Materials*, 20(6):1154–1159, 2008.

[CLZ17] Shaowei Cai, Chuan Luo, and Haochen Zhang. From decimation to local search and back: A new approach to maxsat. In *to appear in proceedings of International Joint Conference on Artificial Intelligence*, 2017.

[CMMS06] Simona Cocco, Remi Monasson, Andrea Montanari, and Guilhem Semerjian. Analyzing search algorithms with physical methods. In Allon Percus, Gabriel Istrate, and Cristopher Moore, editors, *Computational complexity and statistical physics*. OUP USA, 2006.

[cpl] Ibm ilog cplex optimizer.

[CSDG11] G. Catelani, R. J. Schoelkopf, M. H. Devoret, and L. I. Glazman. Relaxation and frequency shifts induced by quasiparticles in superconducting qubits. *Phys. Rev. B*, 84:064517, Aug 2011.

[CTDV17] Francesco Caravelli, Fabio Lorenzo Traversa, and Massimiliano Di Ventra. Complex dynamics of memristive circuits: Analytical results and universal slow relaxation. *Physical Review E*, 95(2):022140, 2017.

[CW08] J. Clarke and F. K. Wilhelm. Superconducting quantum bits. *Nature*, 453(7198):1031–1042, jun 2008.

[DBI+16] Vasil S. Denchev, Sergio Boixo, Sergei V. Isakov, Nan Ding, Ryan Babbush, Vadim Smelyanskiy, John Martinis, and Hartmut Neven. What is the computational value of finite-range tunneling? *Phys. Rev. X*, 6:031015, Aug 2016.

[Di 08] Massimiliano Di Ventra. *Electrical transport in nanoscale systems*, volume 14. Cambridge University Press Cambridge, 2008.

[DNG78] R. C. Dynes, V. Narayanamurti, and J. P. Garno. Direct measurement of quasiparticle-lifetime broadening in a strong-coupled superconductor. *Phys. Rev. Lett.*, 41:1509–1512, Nov 1978.

[DOGK03] R. Deblock, E. Onac, L. P. Gurevich, and L. Kouwenhoven. Detection of quantum noise from an electrically driven two-level system. *Science*, 301(5630):203–206, 2003.

[DS13] M. H. Devoret and R. J. Schoelkopf. Superconducting circuits for quantum information: An outlook. *Science*, 339(6124):1169–1174, 2013.

[DT17] M. Di Ventra and F. L. Traversa. Absence of periodic orbits in digital memcomputing machines with solutions. *Chaos*, 27:101101, 2017.

[DT18] Massimiliano Di Ventra and Fabio L. Traversa. Memcomputing: Leveraging memory and physics to compute efficiently. *J. Appl. Phys.*, 123:180901, 2018.

[DVP13a] Massimiliano Di Ventra and Yuriy V Pershin. On the physical properties of memristive, memcapacitive and meminductive systems. *Nanotechnology*, 24(25):255201, 2013.

[DVP13b] Massimiliano Di Ventra and Yuriy V. Pershin. The parallel approach. *Nature Physics*, 9:200–202, 2013.

[DVT17] M. Di Ventra and Fabio L. Traversa. Absence of chaos in digital memcomputing machines with solutions. *Phys. Lett. A*, 381:3255, 2017.

[DVTO17] M. Di Ventra, Fabio L. Traversa, and Igor V. Ovchinnikov. Topological field theory and computing with instantons. *Ann. Phys. (Berlin)*, 529:1700123, 2017.

[ED05] David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.

[EKPC13] Stanislaw H. Zak Edwin K. P. Chong. *An Introduction to Optimization*. John Wiley & Sons Inc, 2013.

[ERT11] Mária Ercsey-Ravasz and Zoltán Toroczkai. Optimization hardness as transient chaos in an analog approach to constraint satisfaction. *Nat. Phys.*, 7(12):966, 2011.

[FC05] Eric Falcon and Bernard Castaing. Electrical conductivity in granular media and branly?s coherer: A simple experiment. *American journal of physics*, 73(4):302–307, 2005.

[Fei98] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45(4):634–652, jul 1998.

[FH93] Konrad H Fischer and John A Hertz. *Spin glasses*, volume 1. Cambridge University Press, 1993.

[FPT73] C. M. Falco, W. H. Parker, and S. E. Trullinger. Observation of a phase-modulated quasiparticle current in superconducting weak links. *Phys. Rev. Lett.*, 31:933–936, Oct 1973.

[GBB11] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

[GJ90] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990.

[GKSS08] Carla P Gomes, Henry Kautz, Ashish Sabharwal, and Bart Selman. Satisfiability solvers. In Frank Van Harmelen, Vladimir Lifschitz, and Bruce Porter, editors, *Handbook of knowledge representation*, volume 1, chapter 2, pages 89–134. Elsevier, 2008.

[GMP12] F. Giazotto and M. J. Martínez-Pérez. The josephson heat interferometer. *Nature*, 492:401–405, 2012.

[GVS15] C. Guarcello, D. Valenti, and B. Spagnolo. Phase dynamics in graphene-based josephson junctions in the presence of thermal and correlated fluctuations. *Phys. Rev. B*, 92:174519, Nov 2015.

[Hal10] J.K. Hale. *Asymptotic Behavior of Dissipative Systems*, volume 25 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, Rhode Island, 2nd edition, 2010.

[Har74] R. E. Harris. Cosine and other terms in the josephson tunneling current. *Phys. Rev. B*, 10:84–94, Jul 1974.

[Har75] R. E. Harris. Josephson tunneling current in the presence of a time-dependent voltage. *Phys. Rev. B*, 11:3329–3333, May 1975.

[Har76] R. E. Harris. Intrinsic response time of a josephson tunnel junction. *Phys. Rev. B*, 13:3818–3821, May 1976.

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, jul 2001.

[HBB92] F. Hlawatsch and G. F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2):21–67, April 1992.

[HJA+15] Itay Hen, Joshua Job, Tameem Albash, Troels F Rønnow, Matthias Troyer, and Daniel A Lidar. Probing for quantum speedup in spin-glass problems with planted solutions. *Phys. Rev. A*, 92(4):042325, 2015.

[HOT+10] Tsuyoshi Hasegawa, Takeo Ohno, Kazuya Terabe, Tohru Tsuruoka, Tomonobu Nakayama, James K Gimzewski, and Masakazu Aono. Learning abilities achieved by a single solid-state atomic switch. *Adv. Mater.*, 22(16):1831–4, apr 2010.

[Hro10] Juraj Hromkovic. *Algorithmics for Hard Problems: introduction to combinatorial optimization, randomization, approximation, and heuristics*. Springer, 2010.

[HTTA12] Tsuyoshi Hasegawa, Kazuya Terabe, Tohru Tsuruoka, and Masakazu Aono. Atomic Switch: Atom/Ion Movement Controlled Devices for Beyond Von-Neumann Computers. *Adv. Mater.*, 24(2):252–267, 2012.

[IS15] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

[IYAT08] I. H. Inoue, S. Yasuda, H. Akinaga, and H. Takagi. Nonpolar resistance switching of metal/binary-transition-metal oxides/metal sandwiches: Homogeneous/inhomogeneous transition of current distribution. *Phys. Rev. B - Condens. Matter Mater. Phys.*, 77(3):1–7, 2008.

[Iye85] TSKV Iyer. *Circuit Theory*. Tata McGraw-Hill, New Delhi, 1985.

[JMS04] Haixia Jia, Cris Moore, and Bart Selman. From spin glasses to hard satisfiable formulas. In *International Conference on Theory and Applications of Satisfiability Testing*, pages 199–210. Springer, 2004.

[JMS07] Haixia Jia, Cristopher Moore, and Doug Strain. Generating hard satisfiable formulas by hiding solutions deceptively. *Journal of Artificial Intelligence Research*, 28:107–118, 2007.

[JOP$^+$] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed ¡today¿].

[Jos62] B. D. Josephson. Possible new effects in superconductive tunnelling. *Physics letters*, 1(7):251–253, 1962.

[Jos65] B.D. Josephson. Supercurrents through barriers. *Advances in Physics*, 14(56):419–451, 1965.

[KBR$^+$88] B Kahng, G G Batrouni, S Redner, L de Arcangelis, and H J Herrmann. Electrical breakdown in a fuse network with random, continuously distributed breaking strengths. *Phys. Rev. B*, 37(13):7625, 1988.

[KGV83] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[KHZ$^+$15] Helmut G. Katzgraber, Firas Hamze, Zheng Zhu, Andrew J. Ochoa, and H. Munoz-Bauza. Seeking quantum speedup through spin glasses: The good, the bad, and the ugly. *Phys. Rev. X*, 5:031026, Sep 2015.

[Kik51] Ryoichi Kikuchi. A Theory of Cooperative Phenomena. *Phys. Rev.*, 81(6):988–1003, 1951.

[KK06] Sigismund Kobe and Jarek Krawczyk. Ground states, energy landscape, and low-temperature dynamics of $\pm j$ spin glasses. In Allon Percus, Gabriel Istrate, and Cristopher Moore, editors, *Computational complexity and statistical physics*. OUP USA, 2006.

[KLH15] Andrew D King, Trevor Lanting, and Richard Harris. Performance of a quantum annealer on range-limited constraint satisfaction problems. *arXiv preprint arXiv:1502.02098*, 2015.

[KSS09] Henry A Kautz, Ashish Sabharwal, and Bart Selman. Incomplete algorithms. In Armin Biere, Marijn Heule, and Hans van Maaren, editors, *Handbook of satisfiability*, volume 185, pages 185–203. IOS press, 2009.

[KTHA06] Manisha Kundu, Kazuya Terabe, Tsuyoshi Hasegawa, and Masakazu Aono. Effect of sulfurization conditions and post-deposition annealing treatment on structural and electrical properties of silver sulfide films. *Journal of applied physics*, 99(10):103501, 2006.

[Lan90a] CG Langton. Computation at the edge of chaos: phase transitions and emergent computation. *Phys. D Nonlinear Phenom.*, 42:12–37, 1990.

[Lan90b] Chris G Langton. Computation at the edge of chaos: phase transitions and emergent computation. *Physica D*, 42(1-3):12–37, 1990.

[LG07] R. M. Lutchyn and L. I. Glazman. Kinetics of quasiparticle trapping in a cooper-pair box. *Phys. Rev. B*, 75:184520, May 2007.

[Lik79] K. K. Likharev. Superconducting weak links. *Rev. Mod. Phys.*, 51:101–159, Jan 1979.

[Lik86] K. K. Likharev. *Dynamics of Josephson junctions and circuits*. Gordon and Breach, 1986.

[LMS11] J. Leppäkangas, M. Marthaler, and G. Schön. Phase-dependent quasiparticle tunneling in josephson junctions: Measuring the cosφ term with a superconducting charge qubit. *Phys. Rev. B*, 84:060505, Aug 2011.

[LWB+11] M. Lenander, H. Wang, R. C. Bialczak, E. Lucero, M. Mariantoni, M. Neeley, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, T. Yamamoto, Y. Yin, J. Zhao, A. N. Cleland, and J. M. Martinis. Measurement of energy decay in superconducting qubits from nonequilibrium quasiparticles. *Phys. Rev. B*, 84:024501, Jul 2011.

[MAA09] J. M. Martinis, M. Ansmann, and J. Aumentado. Energy decay in superconducting josephson-junction qubits from nonequilibrium quasiparticle excitations. *Phys. Rev. Lett.*, 103:097002, Aug 2009.

[Mat17] The Mathworks, Inc., Natick, Massachusetts. *MATLAB version 9.3.0.713579 (R2017b)*, 2017.

[MAX] Max-sat 2016. Accessed: 2017-06-30.

[MKT17] Salvatore Mandra, Helmut G Katzgraber, and Creighton Thomas. The pitfalls of planar spin-glass benchmarks: raising the bar for quantum annealers (again). *Quantum Science and Technology*, 2(3):038501, 2017.

[MM09] M Mezard and A Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

[MM11] Cristopher Moore and Stephan Mertens. *The Nature of Computation*. Oxford University Press, Inc., New York, NY, USA, 2011.

[MOL+99] J. E. Mooij, T. P. Orlando, L. Levitov, L. Tian, C. H. van der Wal, and S. Lloyd. Josephson persistent-current qubit. *Science*, 285(5430):1036–1039, 1999.

[Mor93] Paul Morris. The breakout method for escaping from local minima. In *AAAI 93*, 1993.

[MPZ02] Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002.

[MRLYGV96] A. Martín-Rodero, A. Levy Yeyati, and F. J. García-Vidal. Thermal noise in superconducting quantum point contacts. *Phys. Rev. B*, 53:R8891–R8894, Apr 1996.

[MSS01] Y. Makhlin, G. Schön, and A. Shnirman. Quantum-state engineering with josephson-junction devices. *Rev. Mod. Phys.*, 73:357–400, May 2001.

[MTD18] H. Manukian, F. L. Traversa, and M. Di Ventra. Accelerating deep learning with memcomputing. *arXiv:1801.00512*, 2018.

[NS11] Ehsan Nedaaee Oskoee and Muhammad Sahimi. Electric currents in networks of interconnected memristors. *Phys. Rev. E*, 83(3):031105, mar 2011.

[NW75] M. Nisenoff and S. Wolf. Observation of a $cos\varphi$ term in the current-phase relation for "dayem"-type weak link contained in an rf-biased superconducting quantum interference device. *Phys. Rev. B*, 12:1712–1714, Sep 1975.

[NY96] Masahiro Nagamatu and Torao Yanaru. On the stability of lagrange programming neural networks for satisfiability problems of prepositional calculus. *Neurocomputing*, 13(2-4):119–133, 1996.

[OHT+11] Takeo Ohno, Tsuyoshi Hasegawa, Tohru Tsuruoka, Kazuya Terabe, James K Gimzewski, and Masakazu Aono. Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.*, 10(8):591–5, aug 2011.

[PD11] Yuriy V. Pershin and Massimiliano Di Ventra. Memory effects in complex materials and nanoscale systems. *Adv. Phys.*, 60(2):145–227, apr 2011.

[PD13] Yuriy V. Pershin and Massimiliano Di Ventra. Self-organization and solution of shortest-path optimization problems with memristive networks. *Phys. Rev. E*, 88(1):013305, jul 2013.

[PDV14a] S. Peotta and M. Di Ventra. Superconducting memristors. *Physical Review Applied*, 2:034011, 2014.

[PDV14b] S. Peotta and M. Di Ventra. Superconducting memristors. *Phys. Rev. Applied*, 2:034011, Sep 2014.

[PFL72] N. F. Pedersen, T. F. Finnegan, and D. N. Langenberg. Magnetic field dependence and q of the josephson plasma resonance. *Phys. Rev. B*, 6:4151–4159, Dec 1972.

[PGC+14] I. M Pop, K. Geerlings, G. Catelani, R. J Schoelkopf, L. I Glazman, and Michel H Devoret. Coherent suppression of electromagnetic dissipation due to superconducting quasiparticles. *Nature*, 508(7496):369–372, apr 2014.

[PS95] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[PTC12] Themistoklis Prodromakis, Christofer Toumazou, and Leon Chua. Two centuries of memristors. *Nature Materials*, 11(6):478, 2012.

[PTDV17] Yan Ru Pei, Fabio L. Traversa, and Massimiliano Di Ventra. On the universality of memcomputing machines. *arXiv:1712.08702*, 2017.

[RD76] R. Rifkin and B. S. Deaver. Current-phase relation and phase-dependent conductance of superconducting point contacts from rf impedance measurements. *Phys. Rev. B*, 13:3894–3901, May 1976.

[RS74] D. Rogovin and D. J. Scalapino. Fluctuation phenomena in tunnel junctions. *Annals of Physics*, 86(1):1 – 90, 1974.

[RTWZ01] Federico Ricci-Tersenghi, Martin Weigt, and Riccardo Zecchina. Simplest random k-satisfiability problem. *Physical Review E*, 63(2):026702, 2001.

[SA12] AZ Stieg and AV Avizienis. Emergent Criticality in Complex Turing BType Atomic Switch Networks. *Adv. Mater.*, 24:286–293, 2012.

[SA14] AZ Stieg and AV Avizienis. Self-organized atomic switch networks. *Jpn. J. Appl. Phys.*, 02(53), 2014.

[SAS+13] Henry O Sillin, Renato Aguilera, Hsien-Hang Shieh, Audrius V Avizienis, Masakazu Aono, Adam Z Stieg, and James K Gimzewski. A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing. *Nanotechnology*, 24(38):384004, sep 2013.

[Sch98] Schrijver. *Theory of Linear Integer Programming*. John Wiley & Sons, 1998.

[Sch99] T. Schoning. A probabilistic algorithm for k-sat and constraint satisfaction problems. In *40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039)*, pages 410–414, Oct 1999.

[SCTD] Forrest Sheldon, Pietro Cicotti, Fabio L. Traversa, and Massimiliano Di Ventra. Stress-testing memcomputing on hard combinatorial optimization problems. *arXiv:1807.00107*.

[SDK+93] James P Sethna, Karin Dahmen, Sivan Kartha, James A Krumhansl, Bruce W Roberts, and Joel D Shore. Hysteresis and hierarchies: Dynamics of disorder-driven first-order phase transformations. *Physical Review Letters*, 70(21):3347, 1993.

[SDV17] Forrest C. Sheldon and Massimiliano Di Ventra. Conducting-insulating transition in adiabatic memristive networks. *Phys. Rev. E*, 95:012305, Jan 2017.

[SGM+17] Bernardo Spagnolo, Claudio Guarcello, Luca Magazz, Angelo Carollo, Dominique Persano Adorno, and Davide Valenti. Nonlinear relaxation phenomena in metastable condensed matter systems. *Entropy*, 19(1):20, 2017.

[SK93] Bart Selman and Henry Kautz. Domain-independent extension to gsat: Solving large structured satisfiability problems. In *Proc. IJCAI-93*, pages 290–295, 1993.

[SLM+92] Bart Selman, Hector J Levesque, David G Mitchell, et al. A new method for solving hard satisfiability problems. In *AAAI*, volume 92, pages 440–446. Citeseer, 1992.

[SMP77] O. H. Soerensen, J. Mygind, and N. F. Pedersen. Measured temperature dependence of the $cos\phi$ conductance in josephson tunnel junctions. *Phys. Rev. Lett.*, 39:1018–1021, Oct 1977.

[SMTC02] Giuseppe E. Santoro, Roman Martoňák, Erio Tosatti, and Roberto Car. Theory of quantum annealing of an ising spin glass. *Science*, 295(5564):2427–2430, 2002.

[SPZS11] Ashivni Shekhawat, Stefanos Papanikolaou, Stefano Zapperi, and James P Sethna. Dielectric Breakdown and Avalanches at Nonequilibrium Metal-Insulator Transitions. *Phys. Rev. Lett.*, 107:276401, 2011.

[SRS08] Amos Sharoni, Juan Gabriel Ramírez, and Ivan K. Schuller. Multiple avalanches across the metal-insulator transition of vanadium oxide nanoscaled junctions. *Phys. Rev. Lett.*, 101:026404, 2008.

[SSA+14] Henry O Sillin, Eric J Sandouk, Audrius V Avizienis, Masakazu Aono, Adam Z Stieg, and James K Gimzewski. Benchtop fabrication of memristive atomic switch networks. *J. Nanosci. Nanotechnol.*, 14(4):2792–8, apr 2014.

[SSFM15] P. K. Sarswat, Y. R. Smith, M. L. Free, and M. Misra. Duality in Resistance Switching Behavior of TiO2-Cu2ZnSnS4 Device. *ECS J. Solid State Sci. Technol.*, 4(8):Q83–Q91, 2015.

[SSSW08] Dmitri B Strukov, Gregory S Snider, Duncan R Stewart, and R Stanley Williams. The missing memristor found. *nature*, 453(7191):80, 2008.

[STD] Forrest Sheldon, Fabio L. Traversa, and Massimiliano Di Ventra. Taming a non-convex landscape with long-range order. *In preparation*.

[Ste69] M. J. Stephen. Lectures on josephson tunneling. In Phillip Russel Wallace, editor, *Superconductivity*, volume 1, pages 297–326, New York, 1969. Gordon and Breach.

[Szi64] Leo Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral science*, 9(4):301–310, 1964.

[TCSD18] F. L. Traversa, P. Cicotti, F. Sheldon, and M. Di Ventra. Evidence of exponential speed-up in the solution of hard optimization problems. *Complexity*, 2018:7982851, 2018.

[TD15] Fabio L. Traversa and Massimiliano Di Ventra. Universal Memcomputing Machines. *IEEE Trans. Neural Networks Learn. Syst.*, 26(11):1–11, may 2015.

[TD18] F. L. Traversa and M. Di Ventra. Memcomputing integer linear programming. *arXiv preprint arXiv:1808.09999*, 2018.

[TDV15] Fabio Lorenzo Traversa and Massimiliano Di Ventra. Universal memcomputing machines. *IEEE Trans. Neural Networks Learn. Syst.*, 26(11):2702, 2015.

[TDV17] Fabio Lorenzo Traversa and Massimiliano Di Ventra. Polynomial-time solution of prime factorization and np-complete problems with digital memcomputing machines. *Chaos*, 27:023107, 2017.

[THNA05] K Terabe, T Hasegawa, T Nakayama, and M Aono. Quantized conductance atomic switch. *Nature*, 433(7021):47, 2005.

[Tin96] M. Tinkham. *Introduction to superconductivity*. Courier Corporation, 1996.

[TRBD15] Fabio L Traversa, Chiara Ramella, Fabrizio Bonani, and M. Di Ventra. Memcomputing NP-complete problems in polynomial time using polynomial resources and collective states. *Sci. Adv.*, 1(6):e1500031–e1500031, 2015.

[TRBDV15] Fabio L. Traversa, Chiara Ramella, Fabrizio Bonani, and M. Di Ventra. Memcomputing NP-complete problems in polynomial time using polynomial resources and collective states. *Science Advances*, 1(6):e1500031, 2015.

[VD74] D. A. Vincent and B. S. Deaver. Observation of a phase-dependent conductivity in superconducting point contacts. *Phys. Rev. Lett.*, 32:212–215, Feb 1974.

[VGS14] D. Valenti, C. Guarcello, and B. Spagnolo. Switching times in long-overlap josephson junctions subject to thermal fluctuations and non-gaussian noise sources. *Phys. Rev. B*, 89:214510, Jun 2014.

[vK07] N. G. van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 2007.

[VM15] Sascha Vongehr and Xiangkang Meng. The missing memristor has not been found. *Scientific reports*, 5:11657, 2015.

[VP13] Massimiliano Di Ventra and YV Pershin. The parallel approach. *Nat. Phys.*, 9(April):200–202, 2013.

[Whi84] Steve R White. Concepts of scale in simulated annealing. In *AIP Conference Proceedings*, volume 122, pages 261–270. AIP, 1984.

[WMK15] Wenlong Wang, Jonathan Machta, and Helmut G. Katzgraber. Comparing monte carlo methods for finding ground states of ising spin glasses: population annealing, simulated annealing, and parallel tempering. *Phys. Rev. E*, 92:013303, Jul 2015.

[XBW+10] Zhi Xu, Yoshio Bando, Wenlong Wang, Xuedong Bai, and Dmitri Golberg. Real-time in situ HRTEM-resolved resistance switching of Ag2S nanoscale ionic conductor. *ACS Nano*, 4(5):2515–2522, 2010.

[YN05] J. Q. You and F. Nori. Superconducting circuits and quantum information. *Physics Today*, 58:42, 2005.

[ZC92] Shengwei Zhang and AG Constantinides. Lagrange programming neural networks. *IEEE Trans. Circuits Syst. II*, 39(7):441–452, 1992.

[ZKLS79] A. B. Zorin, I. O. Kulik, K. K. Likharev, and J. R. Schrieffer. On the sign of the quasiparticle-pair interference current in superconducting tunnel junctions. *Fiz. Nizk. Temp. (Sov. J. Low Temp. Phys.)*, page 1138 (537), 1979.

[ZLS03] Erhai Zhao, Tomas Löfwander, and J. A. Sauls. Phase modulated thermal conductance of josephson weak links. *Phys. Rev. Lett.*, 91:077003, Aug 2003.

[Zor81] A. B. Zorin. Voltage fluctuations in josephson tunnel junctions. *Physica B+C*, 108(1):1293 – 1294, 1981.

[ZRSV99] Stefano Zapperi, Purusattam Ray, H. Eugene Stanley, and Alessandro Vespignani. Avalanches in breakdown and fracture processes. *Phys. Rev. E*, 59(5):5049–5057, may 1999.