Statistical Science
2016, Vol. 31, No. 3, 362–398
DOI: 10.1214/16-STS560

Quantum Annealing with Markov Chain Monte Carlo Simulations and D-Wave Quantum Computers

Yazhen Wang, Shang Wu and Jian Zou

Abstract. Quantum computation performs calculations by using quantum devices instead of electronic devices following classical physics and used by classical computers. Although general purpose quantum computers of practical scale may be many years away, special purpose quantum computers are being built with capabilities exceeding classical computers. One prominent case is the so-called D-Wave quantum computer, which is a computing hardware device built to implement quantum annealing for solving combinatorial optimization problems. Whether D-Wave computing hardware devices display a quantum behavior or can be described by a classical model has attracted tremendous attention, and it remains controversial to determine whether quantum or classical effects play a crucial role in exhibiting the computational input–output behaviors of the D-Wave devices. This paper consists of two parts where the first part provides a review of quantum annealing and its implementations, and the second part proposes statistical methodologies to analyze data generated from annealing experiments. Specifically, we introduce quantum annealing to solve optimization problems and describe D-Wave computing devices to implement quantum annealing. We illustrate implementations of quantum annealing using Markov chain Monte Carlo (MCMC) simulations carried out by classical computers. Computing experiments have been conducted to generate data and compare quantum annealing with classical annealing. We propose statistical methodologies to analyze computing experimental data from a D-Wave device and simulated data from the MCMC based annealing methods, and establish asymptotic theory and check finite sample performances for the proposed statistical methodologies. Our findings confirm bimodal histogram patterns displayed in input–output data from the D-Wave device and both U-shape and unimodal histogram patterns exhibited in input–output data from the MCMC based annealing methods. Further statistical explorations reveal possible sources for the U-shape patterns. On the other hand, our statistical analysis produces statistical evidence to indicate that input–output data from the D-Wave device are not consistent with the stochastic behaviors of any MCMC based annealing models under the study. We present a list of statistical research topics for the future study on quantum annealing and MCMC simulations.

Key words and phrases: Quantum annealing, quantum computing, Markov chain Monte Carlo, Ising model, ground state success probability, Hamiltonian, quantum bit (qubit).

Yazhen Wang is Chair and Professor, Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA (e-mail: yzwang@stat.wisc.edu). Shang Wu is Graduate Student, Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA (e-mail: shangwu@stat.wisc.edu). Jian Zou is Assistant Professor,

1. INTRODUCTION

Quantum computation is based on the idea of using quantum devices to process information and perform computation, instead of electronic devices following the laws of classical physics and used by classical computers (Nielsen and Chuang, 2000 and Wang, 2012). Here, classical computers mean today's electronic based computers. Two quantum computing approaches are logic-gate based quantum computing and adiabatic quantum computing (Aharonov et al., 2007, Deutsch, 1985, DiVincenzo, 1995, Browne, 2014, Farhi et al., 2000, 2001, 2002 and Johnson et al., 2011). Logic-gate based quantum computing has as its purpose the development of a quantum version of classic logic gate operations and the construction of general purpose (or universal) quantum computers. Adiabatic quantum computing is based on quantum annealing to build special purpose quantum computers (i.e., quantum annealers) for solving tough combinatorial optimization problems, where quantum annealing is the quantum analog of classical annealing such as simulated annealing with thermodynamics in classical annealing replaced by quantum dynamics, and quantum annealers refer to physical hardware implementations of quantum annealing (see more details about classical annealing and quantum annealing later in this section and Section 3). Since its introduction by Feynman (1981/82), quantum computation has been proclaimed to be the future of computing, and intensive research efforts are under way around the globe to investigate a number of technologies that could lead to more powerful and more prevalent quantum devices for better computation, communication and cryptography. On the one hand, in spite of tremendous progresses made in the past two decades, general purpose quantum computers of practical scale still have a long way to go. On the other hand, the development of quantum technologies is at the critical point where quantum communication devices and special purpose quantum computers, such as quantum annealers, quantum simulators and quantum crypto devices, can be built with capabilities exceeding classical computer based devices (Aspuru-Guzik et al., 2005, Britton et al., 2012, Browne, 2014, Brumfiel, 2012, Nielsen and Chuang, 2000, Neumann et al., 2008, and Wang, 2012). For these quantum devices, it becomes increasingly important to test and/or certify their behaviors according to claims and/or specifications. One prominent example is the commercialization of a special purpose quantum computer manufactured by D-Wave Systems Incorporation for solving combinatorial optimization problems (Boixo et al., 2014a, 2014b, Boixo et al., 2015a, 2015b, Johnson et al., 2011, Jones, 2013, McGeoch, 2014, Rønnow et al., 2014 and Vinci et al., 2014, Hen et al., 2015, Venturelli et al., 2014, and Martin-Mayor and Hen, 2015).

The D-Wave quantum computer is an analog computing hardware device that is designed and built to physically implement quantum annealing. It consists of a quantum processor based on superconducting flux quantum bits (where quantum bits, called qubits for short, are the quantum analog of classic bits ones and zeros, see more details about qubits in Section 2) and surrounding system such as a cooling apparatus and a magnetic shielded box (Boixo et al., 2014a, 2014b, Clarke and Wilhelm, 2008, Devoret, Wallra and Martinis, 2004, DiCarlo et al., 2009, Johnson et al., 2011, Lanting et al., 2014 and Pudenz, Albash and Lidar, 2014). The hardware device is a quantum annealer, and the qubits are an array of chilled superconducting niobium loops that can be engineered to very quickly find the lowest point in an energy "landscape" of hills and valleys associated with a quantum system. The annealing idea is to define an optimization problem through the energy landscape and represent the lowest energy point as the solution of the problem posed. Such optimization problems turn up in everything from travel salesman problem to integer factoring, from software verification and validation to object recognition and classification, and from genome sequence analysis and protein folding to portfolio optimization and risk analysis (Brooke et al., 1999, Farhi et al., 2000, 2001, 2002, McGeoch, 2014, and Shor, 1994). For example, D-Wave devices have been tested on simple application problems in graphs and networks, machine learning, artificial intelligence and computational biology (Bian et al., 2012, O'Gorman et al., 2014, Perdomo-Ortiz et al., 2012, Perdomo-Ortiz et al., 2014, and Rieffel et al., 2014). As a case in point, the lowest-energy arrangement of a protein is thought to be its preferred state, and protein folding is to find the lowest-energy point in its energy landscape; a D-Wave device has been arranged to manipulate the qubits to reach their lowest-energy state and solve the problem of folding a simple protein (McGeoch, 2014 and Perdomo-Ortiz et al., 2012). For a physical system, its lowest-energy point is referred to as a ground state in physics, with

Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609, USA (e-mail: jzou@wpi.edu).

its higher energy points as excited states. Analogous to classical annealing such as simulated annealing, quantum annealing finds a ground state by allowing the qubits to exploit a quantum effect called quantum tunneling in the sense that they go through the energy hills, rather than climbing over the energy hills in classical thermal annealing (see more details about quantum tunneling in Section 3). It may be more efficient to find a ground state using quantum annealing via quantum tunneling than classical annealing via thermal jump (Boixo et al., 2015a, 2015b, Brooke et al., 1999, Farhi et al., 2001, Jones, 2013, McGeoch, 2014, Perdomo-Ortiz et al., 2014, Pudenz, Albash and Lidar, 2014, and Santoro et al., 2002).

In contrast to the fact that research laboratories can usually manage quantum computers with up to about a dozen of qubits, D-Wave devices utilize a solid state architecture with over a thousand of interlaced superconducting flux qubits. The manufacturing methods and computing technologies of the D-Wave devices are well documented, yet it is challenging to understand their computational power. Whether the D-Wave devices display a large-scale quantum behavior or can be described by a classical model has attracted tremendous attention, and it remains controversial to determine whether quantum or classical effects play a crucial role in exhibiting the computational input–output behaviors of the D-Wave devices. Boixo et al. (2014a, 2014b) employed random Ising model instances to test on a 128-qubit D-Wave device and compared the outputs to the data generated from MCMC based annealing models. They found that the computing experimental data exhibit similar bimodal histogram patterns for the D-Wave device and simulate quantum annealing but unimodal histogram patterns for simulated annealing and a classical spin model. The results are used to argue that the D-Wave performance is consistent with simulated quantum annealing but inconsistent with behaviors of the classical simulated annealing and classical spin models. Shin et al. (2014) proposed a MCMC model, which is simply referred to as the SSSV model, to yield at least as strong correlation with the input–output data of the D-Wave device as simulated quantum annealing and disputed the implication in Boixo et al. (2014a, 2014b) of quantum effects in the D-Wave device. Albash et al. (2014) adopted a measure based on the total variation distance between the probability distributions over excited states and demonstrated that the D-Wave data are not well correlated with SQA or SSSV data by the measure. Vinci et al. (2014) further tested on a 512-qubit D-Wave device to distinguish (or associate) its input–output behavior from (or with) those obtained from the classical and quantum annealing models in terms of the population ratio of the isolated and clustered ground states. Rønnow et al. (2014) reported to find no evidence of quantum speedup for a 512-qubit D-Wave device on the problem of random spin glass instances, despite of some other speedup claims (Browne, 2014, Hen et al., 2015, Katzgraber, Hamze and Andrist, 2014, McGeoch, 2014 and Venturelli et al., 2014).

Attempts to quantify the quantum nature of D-Wave devices have been not only met with excitement but also confronted with suspicion. These studies pose fundamental questions regarding the distinguishability between quantum annealers and classical thermal annealers. We may boil the distinguishability questions down to statistical analysis of annealing methods and the associated statistical inference problems. The consistence or inconsistence claims in the literature regarding the studies of the D-Wave devices along with MCMC based annealing methods are largely based on causal informal inspection of correlations and histogram patterns. The first part of the paper reviews quantum annealing and its implementations by D-Wave devices and MCMC based annealing approaches. We illustrate computing experiments for quantum annealing and carry out the MCMC simulations by classical computers to generate data and compare quantum annealing with classical annealing. The second part of the paper proposes statistical methodologies to analyze the input–output behaviors of the D-Wave devices and MCMC based annealing models. We have established the asymptotic theory to justify the proposed statistical methodologies, and conducted numerical simulations to check their finite sample performances. The statistical shape tests support bimodal but not U-shape histogram patterns exhibited in the data from a D-Wave device, and confirm U-shape histogram patterns in the data from the simulated quantum annealing and SSSV models and unimodal (including monotone) histogram patterns in the data from the classical simulated annealing model. We further explore the energy gap between the ground states and the excited states, and uncover some possible sources for the U-shape patterns. Our analysis results provide statistical evidence to suggest that the input–output data from the D-Wave device are not consistent with the statistical behaviors of these MCMC based annealing models.

The rest of the paper proceeds as follows. Section 2 provides a brief introduction to quantum mechanics

and quantum computation. The main review and research parts of the paper are given by Sections 3 and 4, respectively. Section 3 introduces quantum annealing and its implementations by D-Wave devices and by the MCMC based annealing methods in the context of the Ising model. We describe D-Wave computing experiments and carry out MCMC simulations by classical computers to generate data and compare quantum annealing with classical annealing. Section 4 proposes statistical methodologies to analyze input–output data from a D-Wave device and from various MCMC based annealing models, and presents our findings on the statistical behaviors of the D-Wave device and the MCMC based annealing models. Section 5 features concluding remarks with a list of future statistical research topics. All proofs are relegated in the Appendix.

2. A BRIEF QUANTUM BACKGROUND

Quantum physics is counter intuitive and hard to grasp. It is fundamentally different from what we might expect on the basis of everyday experiences. As a matter of fact, Neils Bohr, a founder of quantum theory, once said, "Anyone who thinks he understands quantum theory doesn't really understand it." As this paper involves only quantum computing based on quantum annealing, we will bypass high-level models of quantum computation and keep our review on quantum physics and quantum computation at the minimal level.

2.1 Notation

For the purpose of this paper, we consider only the finite dimension case. Denote by $\mathbb{C}^d$ the d-dimensional complex space. Given a vector ψ in $\mathbb{C}^d$, we follow the convention in quantum mechanics and quantum computation to use Dirac notations ket $|\cdot\rangle$ and bra $\langle\cdot|$ to indicate that $|\psi\rangle$ and $\langle\psi|$ are column and row vectors, respectively. Denote by superscripts $*$, $'$ and $\dagger$ the conjugate of a complex number, the transpose of a vector or matrix, and the conjugate transpose operation, respectively. A natural inner product in $\mathbb{C}^d$ is given by $\langle u|v\rangle = \sum_{j=1}^{d} u_j^* v_j = (u_1^*, \ldots, u_d^*)(v_1, \ldots, v_d)'$, where $\langle u| = (u_1, \ldots, u_d)$ and $|v\rangle = (v_1, \ldots, v_d)'$, and the modulus $|u| = \sqrt{\langle u|u\rangle}$. We say a matrix $\mathbf{A}$ is Hermitian if $\mathbf{A} = \mathbf{A}^\dagger$, and a matrix $\mathbf{U}$ is said to be unitary if $\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix.

2.2 Quantum Physics

A quantum system is completely characterized by its state and the time evolution of the state. A d-dimensional quantum system at a given time can be described by its quantum state, often given in terms of a unit vector $|\psi\rangle$ in $\mathbb{C}^d$. To study the quantum system, we perform measurements on the system to obtain data, which may be carried out in terms of the so-called observables. An observable $\mathbf{M}$ is defined to be a Hermitian matrix on $\mathbb{C}^d$. Assume that the eigen-ecomposition of $\mathbf{M}$ is as follows:

$$\mathbf{M} = \sum_{a=1}^{r} \lambda_a \mathbf{Q}_a,$$

where $\lambda_1, \ldots, \lambda_r$ are the real eigenvalues of $\mathbf{M}$, and $\mathbf{Q}_a$ is the projection onto the eigen-space corresponding to the eigenvalue λ_a. According to quantum theory, when we measure the quantum system in terms of $\mathbf{M}$ under the state $|\psi\rangle$, the measurement outcome Λ is a random variable that takes values in $\{\lambda_1, \lambda_2, \ldots, \lambda_r\}$, with probability distribution $P(\Lambda = \lambda_a) = \mathrm{tr}(\mathbf{Q}_a|\psi\rangle\langle\psi|) = \langle\psi|\mathbf{Q}_a|\psi\rangle$, $a = 1, 2, \ldots, r$. Statistically, we may perform measurements on $\mathbf{M}$ for the quantum system multiple times to obtain measurement data and infer the quantum state from the data. For a quantum system at a given time point, its state vector comprises all information about the system in the sense that it can yield the probability distributions of measurement outcomes and may encapsulate all things of importance about the quantum system.

To describe the time evolution of a quantum system, we denote by $|\psi(t)\rangle$ the state of the quantum system at time t, which is also referred to as a wave function. The states $|\psi(t_1)\rangle$ and $|\psi(t_2)\rangle$ at times t_1 and t_2 are connected through $|\psi(t_2)\rangle = \mathbf{U}(t_1, t_2)|\psi(t_1)\rangle$, where $\mathbf{U}(t_1, t_2) = \exp[-i\mathbf{H}(t_2 - t_1)]$ is a unitary matrix, and $\mathbf{H}$ is a Hermitian matrix on $\mathbb{C}^d$. In fact, the continuous time evolution of $|\psi(t)\rangle$ is governed by the Schrödinger equation

$$\sqrt{-1}\frac{\partial|\psi(t)\rangle}{\partial t} = \mathbf{H}|\psi(t)\rangle \quad \text{or equivalently} \quad |\psi(t)\rangle = e^{-\sqrt{-1}\mathbf{H}t}|\psi(0)\rangle, \tag{2.1}$$

where $\mathbf{H}$ is a possibly time-dependent Hermitian matrix on $\mathbb{C}^d$, which is known as the Hamiltonian of the quantum system. The Schrödinger equation shows that for the quantum system, its Hamiltonian can completely describe the dynamic evolution of its quantum states. It should be stressed that besides the approach by the Schrödinger equation to describe the quantum evolution, there are other formulations of quantum mechanics such as the so-called matrix mechanics created by Werner Heisenberg, Max Born and Pascual Jordan and the path integral formulation due to Richard

Feynman. See Holevo (1982), Sakurai and Napolitano (2010), Shankar (1994), Wang (2012, 2013), Wang and Xu (2015), Xu (2015) and Cai et al. (2016).

2.3 Qubits and Their Quantum Properties

Quantum computation grapples with understanding how to take advantage of the enormous information hidden in the quantum systems and to harness the immense potential of atoms and photons for the purpose of computing. It utilizes strange quantum phenomena such as quantum superposition, quantum entanglement and quantum tunneling to do the trick of performing computation and processing information. It intends to develop quantum computing devices for solving certain tough computational problems faster and/or more efficient than classical computers.

Any computers must utilize some states of physical systems to store digits. Classical computers use voltage levels to encode bits 0 and 1. Analog to bits 0 and 1 in classic computation, quantum computation uses qubits $|0\rangle$ and $|1\rangle$. However, unlike classical computation where bits are either 0's and 1's and transistors can crunch the ones and zeroes individually, quantum computation allows qubits to encode ones and zeroes simultaneously through what is known as quantum superposition. A qubit can be realized by various physical systems, such as the quantum spin of a particle where $|0\rangle$ and $|1\rangle$ correspond to the spin up state $|\uparrow\rangle$ and spin down state $|\downarrow\rangle$ of the particle, respectively. According to quantum physics, besides states $|0\rangle$ and $|1\rangle$ the particle may also exist in a superposition state, which is a blend of spin up and down states simultaneously. Thus, while at any instant a classical bit can be either 0 or 1, a qubit can be a superposition of both $|0\rangle$ and $|1\rangle$, that is, a superposition qubit is in a quantum state that may be viewed as both one and zero at the same time. Mathematically, a superposition qubit may take the form of $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, where α_0 and α_1 are two complex numbers satisfying $|\alpha_0|^2 + |\alpha_1|^2 = 1$. In other words, a qubit represents unit vectors in a two-dimensional complex vector space $\mathbb{C}^2$, and $|0\rangle$ and $|1\rangle$ consist of an orthonormal basis for the space and are often referred to as the computational basis. A classical bit can be examined to determine whether it is in the states 0 or 1, but for a qubit we can not determine its state and find the values of α_0 and α_1 by examining it. By quantum physics, we can perform a measurement on qubit $|\psi\rangle$ and obtain either the result 0, with probability $|\alpha_0|^2$, or the result 1, with probability $|\alpha_1|^2$.

Like classic bits, we can define multiple qubits. The states of b qubits are unit vectors in $\mathbb{C}^d$, where $d = 2^b$, with computational basis states of the form $|x_1 x_2 \cdots x_b\rangle$, $x_j = 0$ or 1, $j = 1, \ldots, b$. For example, the states of two qubits are unit vectors in $\mathbb{C}^4$, with the superposition states $|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$, where amplitudes α's are complex numbers satisfying $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$. Two qubits can be made to interact and stay in an entangled state so that they maintain persistent, instant influence on each other no matter how far apart they become and regardless of the nature of the intervening medium. A consequence of entanglement is that measuring one of the two entangled qubits will make the state of the other entangled qubit to be definite, and thus completely predictable, even though the two qubits may be very far away.

The quantum system of b qubits is described by $\mathbb{C}^d$ with each superposition state specified by 2^b amplitudes. As 2^b increases exponentially in b, it is very easy for such a system to grow with an enormously big space. As a result, we encounter a key signature in handling a quantum system that its complexity usually grows exponentially with its size. It takes an exponential number of bits of memory on a classical computer to store the state of a quantum system, and simulations of quantum systems via classical computers face great computational challenge. For example, besides the memory storage difficulty, we need to evaluate $e^{-\sqrt{-1}\mathbf{H}t}$ in (2.1) for the study of the quantum evolution; as the matrix size of Hamiltonian $\mathbf{H}$ grows exponentially with the size of the system, it is extremely difficult to exponentiate the quantum Hamiltonian using mathematical analysis or classical computers. On the other hand, since quantum systems are able to store and keep track an exponential number of complex numbers and perform data manipulations and calculations as the systems evolve, quantum computation and quantum information search for ways to access the exponential information reservoir hidden in the quantum systems and utilize the tremendous potential for computing. See Nielsen and Chuang (2000) and Wang (2011, 2012).

3. QUANTUM ANNEALING AND MCMC SIMULATIONS

Annealing a material by heating and slowly cooling it to enhance its quality is an ancient technique that has been used for materials like glasses and metals over seven thousand years. Mimicking this process by computer simulations creates simulated annealing

as an optimization method. It is based on the analogy between the behavior of a complex physical system with many degrees of freedom and an optimization problem of finding the global minimum of a given objective (or cost) function depending on many parameters. By viewing the objective function of the optimization problem as the energy of the physical system, we naturally formulate the optimization problem as the (possible NP-hard) problem of finding minimum energy configurations (or ground states) of the many-body physical system, and develop computer simulation algorithms to mimic the system behavior and search for the minimum energy configurations of the corresponding physical model.

A classical approach is simulated annealing (SA), which takes into account the relative configuration energies and a fictitious time-dependent temperature when probabilistically exploring the immense search space. For a given optimization problem, its objective function to be minimized is identified with the energy of a physical system, and we then give the physical system a temperature as an artificially-introduced control parameter. By reducing the temperature gradually from a high value to zero, we wish to drive the system to the state with the lowest value of the energy (objective function), and thus reach the solution of the optimization problem. The initial temperature is usually set high relative to the system energy scale in order to induce thermal fluctuations and sample its configurations using MCMC simulations like the Metropolis algorithm, and the annealing process escapes from local minima by thermal fluctuations to climb over energy barriers and searches for lower energy configurations. As the system evolves with the change of temperature sufficiently slow, the system is expected to stay close to thermal equilibrium during time evolution, and thus in the end we lead the SA system to the zero-temperature equilibrium state, the lowest-energy state. See Bertsimas and Tsitsiklis (1992), Kirkpatrick, Gelatt and Vecchi (1983) and Winker (2001).

Quantum annealing is based on the physical process of a quantum system whose lowest energy, or equivalently, a ground state of the system, represents the solution to an optimization problem posed. It starts with building a simple quantum system initialized in its ground state, and then moves the simple system gradually toward the target complex system. According to the quantum adiabatic theorem (Farhi et al., 2000, 2001, 2002), as the system slowly evolves, it tends to remain in a ground state, and hence measuring the state of the final system will yield an answer to the original optimization problem with some probability. The key idea behind quantum annealing is to replace thermal fluctuations in SA by quantum fluctuations via quantum tunneling so that the system is kept close to some instantaneous ground state of the quantum system during the quantum annealing evolution, analog to some quasi-equilibrium state to be kept during the time evolution of SA.

Both classical and quantum annealing techniques are powerful tools for solving hard optimization problems, whether they are utilized as physical devices or simulation methods. The physical scheme is to employ a natural system or build a device to engineer a physical system so that ground states of the system represent the sought-after solution of an optimization problem (McGeoch, 2014). The simulation approach is to apply "escape" rules in computer simulations to prevent the system from getting trapped in local minima of an energy or cost function, and eventually reach the global minimum with some probability (Rieger and Kawashima, 1999 and Martoňák, Santoro and Tosatti, 2002). In both situations, the system is allowed to probabilistically explore its immense configuration space and ultimately "freeze" in the global minimum with certain probability, and by enough repeated tries we can find the global minimum and solve the optimization problem.

3.1 Classical Ising Model and Simulated Annealing

The Ising model is often used to describe natural systems in physics, and many optimization problems can be mapped into physical systems described by the Ising model whose ground states provide the solutions to the optimization problems. Examples include traveling salesman problem, portfolio optimization, integer factoring, social economics network, protein folding, protein modeling and statistical genetics. See Irback, Peterson and Potthast (1996), Majewski, Li and Ott (2001), McGeoch (2014) and Stauffer (2008).

Consider the Ising model described by a graph $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$, where $\mathcal{V}(\mathcal{G})$ and $\mathcal{E}(\mathcal{G})$ stand for the vertex and edge sets of $\mathcal{G}$, respectively. Each vertex is occupied by a random variable taking values in $\{+1, -1\}$, and each edge represents the coupling (or interaction) between the two vertex variables connected by the edge. A configuration $\mathbf{s} = \{s_j, j \in \mathcal{V}(\mathcal{G})\}$ is defined to be an assignment of a set of values to all vertex variables s_j, $j \in \mathcal{V}(\mathcal{G})$. Vertices are also referred to as sites, and vertex variables as spins in physics, with $+1$ for spin up and -1 for spin down. For example, consider a graph corresponding to a two-dimensional lattice with

a magnet placed at each lattice site pointing either up or down. Suppose that there are b lattice sites labelled by $j = 1, \ldots, b$. Site variable s_j stands for a binary random variable indicating the position of the magnet at site j, where $s_j = \pm 1$ may be interpreted as the jth magnet pointing up or down, respectively.

The Hamiltonian of the classical Ising model is given by

$$\mathbf{H}_I^c(\mathbf{s}) = -\sum_{(i,j)\in\mathcal{E}(\mathcal{G})} J_{ij} s_i s_j - \sum_{j\in\mathcal{V}(\mathcal{G})} h_j s_j, \tag{3.1}$$

where (i, j) stands for the edge between sites i and j, with the first sum over all pairs of vertices with edge $(i, j) \in \mathcal{E}(\mathcal{G})$, J_{ij} stands for the interaction (or coupling) between sites i and j associated with edge $(i, j) \in \mathcal{E}(\mathcal{G})$, and h_j describes an external magnetic field on vertex $j \in \mathcal{V}(\mathcal{G})$. A set of fixed values $\{J_{ij}, h_j\}$ is referred to as one instance of the Ising model. For a given configuration $\mathbf{s}$, the energy of the Ising model is equal to $\mathbf{H}_I^c(\mathbf{s})$.

According to Boltzmann's law, the probability of a given configuration $\mathbf{s}$ is described by the Boltzmann (or Gibbs) distribution

$$P_\beta(\mathbf{s}) = \frac{e^{-\beta \mathbf{H}_I^c(\mathbf{s})}}{Z_\beta}, \quad Z_\beta = \sum_{\mathbf{s}} e^{-\beta \mathbf{H}_I^c(\mathbf{s})},$$

where $\beta = (k_B T)^{-1}$ is an inverse temperature, with k_B a generic physical constant called the Boltzmann constant and T the absolute temperature, and the normalization constant Z_β is called the partition function. If we take $k_B = 1$, T is interpreted as the fundamental temperature of the system with units of energy, and β is treated as the reciprocal to the fundamental temperature. The configuration probability $P_\beta(\mathbf{s})$ represents the probability that the physical system is in a state with configuration $\mathbf{s}$ in equilibrium.

When a combinatorial optimization is represented by the Ising model, the objective is to find a ground state of the Ising model, that is, we need to find a state whose configuration minimizes the energy function $\mathbf{H}_I^c(\mathbf{s})$. If the Ising model has b sites, the configuration space is $\{-1, +1\}^b$ with the total number of the configurations equal to 2^b. For a system with many sites, because of the exponential complexity, it is an extremely difficult task to numerically find ground states and solve the minimization problem. In fact, it is prohibitive for deterministic exhaustive search algorithms to solve the minimization problem with such an exponential growth search space. Annealing approaches such as SA are often employed to probabilistically explore the search space and find the global minimum of $\mathbf{H}_I^c(\mathbf{s})$. SA finds the minimum with repeated tries by using MCMC methods such as the Metropolis–Hastings algorithm to generate configuration samples from the Boltzmann distribution with temperature decreasing slowly. It starts with a random initial spin configuration and randomly flips spins at each time step; we always accept a new spin configuration if it lowers the energy and accept it probabilistically using the Metropolis rule otherwise; and we gradually lower the temperature to reduce the escape probability of trapping in local minima. Specifically, the SA algorithm initializes spins with -1 and $+1$ randomly and independently from each other to obtain initial spins $\mathbf{s}^{(0)} = \{s_j^{(0)}\}$. We update spins one by one, and each complete updating over all spins constitutes one sweep. At the kth sweep, for spin i, we attempt to flip its state $s_i^{(k-1)}$ to new state $s_i^{(k)} = -s_i^{(k-1)}$ while keeping all other spins unchanged, and calculate energy change between its original state $s_i^{(k-1)}$ and the newly flipped state $s_i^{(k)}$,

$$\begin{aligned}\Delta E_i^{(k)} = {} & -h_i\big(s_i^{(k)} - s_i^{(k-1)}\big) \\ & - \sum_{j=1}^{i-1} J_{ij} s_j^{(k)} \big(s_i^{(k)} - s_i^{(k-1)}\big) \\ & - \sum_{j=i+1}^{b} J_{ij} s_j^{(k-1)} \big(s_i^{(k)} - s_i^{(k-1)}\big).\end{aligned}$$

The new state $s_i^{(k)}$ is accepted with probability $\min\{1, \exp(-\Delta E_i^{(k)}/T_k)\}$, that is, we change spin i's state from $s_i^{(k-1)}$ to new state $s_i^{(k)}$ if $\Delta E_i^{(k)} \le 0$ and otherwise update its state with probability $\exp(-\Delta E_i^{(k)}/T_k)$, where T_k is the annealing schedule to lower the temperature. Typical annealing schedules for lowering temperature include T_k proportional to $\frac{1}{k}$ or $\frac{1}{\log k}$. See Bertsimas and Tsitsiklis (1992), Geman and Geman (1984) and Hajek (1988).

3.2 Quantum Annealing

We describe a quantum Ising system associated with quantum annealing through the same graph $\mathcal{G}$ used for the classical Ising model, where similar to the classical case, the vertex set $\mathcal{V}(\mathcal{G})$ now represents quantum spins, with edge set $\mathcal{E}(\mathcal{G})$ for couplings (or interactions) between two quantum spins. As illustrated in Section 2.3, each quantum spin may realize a qubit, and thus each vertex represents a qubit, with the total number of qubits in the system equal to the total number, b, of the vertices in $\mathcal{G}$. The quantum system

is characterized by complex space $\mathbb{C}^d$, where $d = 2^b$, with its quantum state described by a unit vector in $\mathbb{C}^d$ and its dynamic evolution governed by the Schrödinger equation defined in (2.1) via a quantum Hamiltonian, which is a Hermitian matrix of size d. The energies of the quantum system correspond to the eigenvalues of the quantum Hamiltonian, with a ground state being an eigenvector corresponding to the smallest eigenvalue. To specify the quantum Hamiltonian we need to define

$$\mathbf{I}_j = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\sigma}_j^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$\boldsymbol{\sigma}_j^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad j = 1, \ldots, b,$$

where $\boldsymbol{\sigma}_j^x$ and $\boldsymbol{\sigma}_j^z$ are called Pauli matrices in x and z axes, respectively (here we do not need Pauli matrix in y axis, see Nielsen and Chuang, 2000 and Wang, 2012). To describe the quantum system, we replace each classical vertex variable $s_j = \pm 1$ in (3.1) by $\boldsymbol{\sigma}_j^z$ for quantum spin j. Pauli matrix $\boldsymbol{\sigma}_j^z$ has two eigenvalues ± 1 whose eigen-states $|+1\rangle$ and $|-1\rangle$ correspond to spin up state $|\uparrow\rangle$ and spin down state $|\downarrow\rangle$, respectively, so that a qubit at vertex j is realized by the quantum spin. Quantum configurations consist of the 2^b possible combinations of $2b$ eigen-states $|\pm 1\rangle$ of the Pauli matrices $\{\boldsymbol{\sigma}_j^z\}_{j=1}^b$. Replacing s_j in classical Ising Hamiltonian $\mathbf{H}_I^c(\mathbf{s})$ by $\boldsymbol{\sigma}_j^z$, we obtain the following quantum Hamiltonian of the quantum Ising model:

$$\mathbf{H}_I^q = -\sum_{(i,j)\in\mathcal{E}(\mathcal{G})} J_{ij}\boldsymbol{\sigma}_i^z\boldsymbol{\sigma}_j^z - \sum_{j\in\mathcal{V}(\mathcal{G})} h_j\boldsymbol{\sigma}_j^z, \tag{3.2}$$

where J_{ij} is the Ising coupling along the edge $(i, j) \in \mathcal{E}(\mathcal{G})$, and h_j is the local field on the vertex $j \in \mathcal{V}(\mathcal{G})$. Here, we use the convention in the quantum literature that $\boldsymbol{\sigma}_j^z$ and $\boldsymbol{\sigma}_i^z\boldsymbol{\sigma}_j^z$ in (3.2) denote their tensor products along with identical matrices in such a way that

$$\begin{aligned}\boldsymbol{\sigma}_i^z\boldsymbol{\sigma}_j^z \equiv{}& \mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_{i-1} \\ &\otimes \underbrace{\boldsymbol{\sigma}_i^z \otimes \mathbf{I}_{i+1} \otimes \cdots \otimes \mathbf{I}_{j-1} \otimes \boldsymbol{\sigma}_j^z}_{\text{vertices } i \text{ and } j} \\ &\otimes \mathbf{I}_{j+1} \otimes \cdots \otimes \mathbf{I}_b,\end{aligned}$$

$$\boldsymbol{\sigma}_j^z \equiv \mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_{j-1} \underbrace{\otimes \boldsymbol{\sigma}_j^z \otimes}_{\text{vertex } j} \mathbf{I}_{j+1} \otimes \cdots \otimes \mathbf{I}_b.$$

That is, we have one matrix, either a Pauli matrix $\boldsymbol{\sigma}_j^z$ or an identity matrix $\mathbf{I}_j$, to act on the jth qubit, each term in (3.2) is a tensor product of b matrices of size two, and the quantum convention simply identifies the qubits with Pauli matrices for real actions but omits the identical matrices and tensor product signs. Because of the tensor products, all terms in $\mathbf{H}_I^q$ are diagonal matrices of size 2^b, and so does $\mathbf{H}_I^q$.

The goal of finding a ground state of quantum Hamiltonian $\mathbf{H}_I^q$ is to search for an eigenvector of $\mathbf{H}_I^q$ corresponding to its smallest eigenvalue, or equivalently, a quantum spin configuration with the minimal energy. Because $\mathbf{H}_I^q$ involves only tensor products of commuting diagonal matrices $\mathbf{I}_j$ and $\boldsymbol{\sigma}_j$, $\mathbf{H}_I^q$ is a diagonal matrix with eigenvalues equal to its diagonal entries. Moreover, it is easy to show from the same structure of (3.1) and (3.2) that the eigenvalues of $\mathbf{H}_I^q$ are actually the 2^b values of classical Hamiltonian $\mathbf{H}_I^c(\mathbf{s})$, $\mathbf{s} \in \{-1, +1\}^b$. Thus, the system governed by $\mathbf{H}_I^q$ behaves essentially like a classical system, and finding the minimal energy of the quantum Ising Hamiltonian $\mathbf{H}_I^q$ is equivalent to finding the minimal energy of the classical Ising Hamiltonian $\mathbf{H}_I^c$.

Since the original optimization problem described in Section 3.1 can be formulated in the quantum framework, the computational task for solving the optimization problem remains the same as in the classical case so far. The key in quantum annealing is to engineer a magnetic field orthogonal to the Ising axis and obtain the Hamiltonian of the Ising model in the transverse field. With $\mathbf{H}_I^q$ representing a potential energy, the transverse field stands for a kinetic energy that does not commute with $\mathbf{H}_I^q$, thus it induces transitions between the up and down states of each single spin, and turns the model behavior from classical to quantum. Specifically, assume that the transverse magnetic field is governed by a quantum Hamiltonian

$$\mathbf{H}_X = -\sum_{j\in\mathcal{V}(\mathcal{G})} \boldsymbol{\sigma}_j^x, \tag{3.3}$$

where we again follow the quantum convention that $\boldsymbol{\sigma}_j^x$ stands for tensor products of b matrices of size 2, that is,

$$\boldsymbol{\sigma}_j^x \equiv \mathbf{I}_1 \otimes \cdots \otimes \mathbf{I}_{j-1} \underbrace{\otimes \boldsymbol{\sigma}_j^x \otimes}_{\text{vertex } j} \mathbf{I}_{j+1} \otimes \cdots \otimes \mathbf{I}_b,$$

which does not commute with $\boldsymbol{\sigma}_j^z$ in $\mathbf{H}_I^q$. Since $\boldsymbol{\sigma}_j^x$ has two eigenvalues ± 1 with corresponding eigenvectors $|\mathbf{v}_{j,\pm 1}\rangle = (1, \pm 1)^\dagger$, $\mathbf{H}_X$ has simple eigenvector $|\mathbf{v}_+\rangle = |\mathbf{v}_{1,+1}\rangle \otimes |\mathbf{v}_{2,+1}\rangle \otimes \cdots \otimes |\mathbf{v}_{b,+1}\rangle$ corresponding to its smallest eigenvalue, namely, $|\mathbf{v}_+\rangle$ is the ground state of $\mathbf{H}_X$.

The quantum annealing procedure starts with a quantum system driven by the transverse magnetic field $\mathbf{H}_X$ and initialized in its ground state. During the process

of quantum annealing, the system is evolved gradually from initial Hamiltonian $\mathbf{H}_X$ to the final target Hamiltonian $\mathbf{H}_I^q$. For such Hamiltonian changes, the quantum adiabatic theorem indicates that the system tends to remain in ground states of the instantaneous Hamiltonian through quantum tunneling (Farhi et al., 2000, 2001, 2002, and McGeoch, 2014). At the end of quantum annealing, if the system is in a ground state of the final Hamiltonian $\mathbf{H}_I^q$, an optimal solution is obtained by measuring the system. Specifically, quantum annealing is realized by an instantaneous Hamiltonian for the Ising model in the transverse field as follows:

$$\mathbf{H}_D(t) = A(t)\mathbf{H}_X + B(t)\mathbf{H}_I^q, \quad t \in [0, t_f], \tag{3.4}$$

where $A(t)$ and $B(t)$ are time-dependent smooth functions controlling the annealing schedules, and t_f is the total annealing time. Typically, $A(t_f) = B(0) = 0$, $A(t)$ is decreasing and $B(t)$ is increasing. The gradual move of the Hamiltonian from $\mathbf{H}_D(0)$ to $\mathbf{H}_D(t_f)$ is achieved through controlling the annealing schedules $A(t)$ and $B(t)$. At initial time $t = 0$, $\mathbf{H}_D(0) = A(0)\mathbf{H}_X$, at the final time $t = t_f$, $\mathbf{H}_D(t_f) = B(t_f)\mathbf{H}_I^q$, and $A(0)$ and $B(t_f)$ are known scalars, then $\mathbf{H}_D(t)$ has the same eigenvectors as $\mathbf{H}_X$ at the initial time and as $\mathbf{H}_I^q$ at the final time, where the corresponding eigenvalues differ by factors of $A(0)$ and $B(t_f)$, respectively. Therefore, the quantum annealing evolution driven by $\mathbf{H}_D(t)$ essentially moves the system $\mathbf{H}_X$ initialized at its ground state to the final system $\mathbf{H}_I^q$. According to the quantum adiabatic theorem (Aharonov et al., 2007, Born and Fock, 1928, Farhi et al., 2000, 2001, 2002, McGeoch, 2014, and Morita and Nishimori, 2008), for appropriately chosen $A(t)$ and $B(t)$ we can measure the quantum system at the final annealing time t_f to find a ground state of $\mathbf{H}_I^q$ with sufficiently high probability. Thus, with certain probability the quantum annealing procedure driven by (3.4) can obtain the global minimum of the objective function $\mathbf{H}_I^c(\mathbf{s})$ and solve the minimization problem.

In quantum annealing, the move from the initial simple system $\mathbf{H}_X$ to the final complex system $\mathbf{H}_I^q$ is often realized through engineering magnetic fields. For example, we may manufacture the move from the initial system to the final system by turning magnetic fields on and then off adiabatically, like the physical implementation by D-Wave devices that initializes its system in the ground state of $\mathbf{H}_X$ by chilling the system to a near absolute zero temperature and turning on the transverse magnetic field. Quantum annealing induces quantum fluctuations by introducing artificial degrees of freedom of quantum nature, and then ingeniously controls the strength of these quantum fluctuations via annealing schedules $A(t)$ and $B(t)$ in order to make the system finally reach a ground state of $\mathbf{H}_I^q$, just like the slow reduction of temperature in classical annealing. Initially, the strength of quantum fluctuations is set to be relatively high for the system to search for the global structure of the configuration space, corresponding to the initial high-temperature situation in SA, and then the strength is slowly decreased to finally vanish to recover the original target system hopefully in the lowest-energy state. Quantum annealing differs from the classical annealing in terms of quantum-mechanical fluctuations in quantum annealing instead of thermal fluctuations in SA. While classical annealing relies on thermal fluctuations to make the system hop from state to state over intermediate energy barriers and search for the desired lowest-energy state, quantum annealing replaces thermal hopping by quantum-mechanical fluctuations for state transitions. The quantum fluctuations in quantum annealing are realized via quantum tunneling that allows the annealing process to explore the different states by traveling directly through energy barriers, rather than climbing over them thermally. Quantum tunneling refers to the quantum phenomenon where particles tunnel through a barrier in the circumstance that could not be possible by classical physics. The tunneling process cannot be directly observed nor adequately explained by classical physics. Quantum tunneling is often explained using the Heisenberg uncertainty principle and the wave-particle duality of matter in quantum physics (Sakurai and Napolitano, 2010 and Shankar, 1994). Here, we try to provide some heuristic explanation. To facilitate our understanding of the phenomenon, we may compare particles attempting to travel between barriers to balls trying to roll over a hill or balls trying to penetrate a wall. Classical physics indicates that balls without sufficient energy to surmount the hill would roll back down, and balls which do not have enough energy to penetrate the wall would either bounce back or bury themselves inside the wall. In both cases, balls will not be able to reach the other sides of the hill and wall. While these particles without enough energy to classically surmount a barrier cannot get to the other side of the barrier by classical physics, quantum physics predicts that the particles can, with some probability, cross the barrier and reach the other side, thus tunneling through the barrier. Figure 1 provides some cartoon illustrations of thermal hopping versus quantum tunneling. The illustrations may indicate that if energy barriers are high and thin, quantum tunneling may be

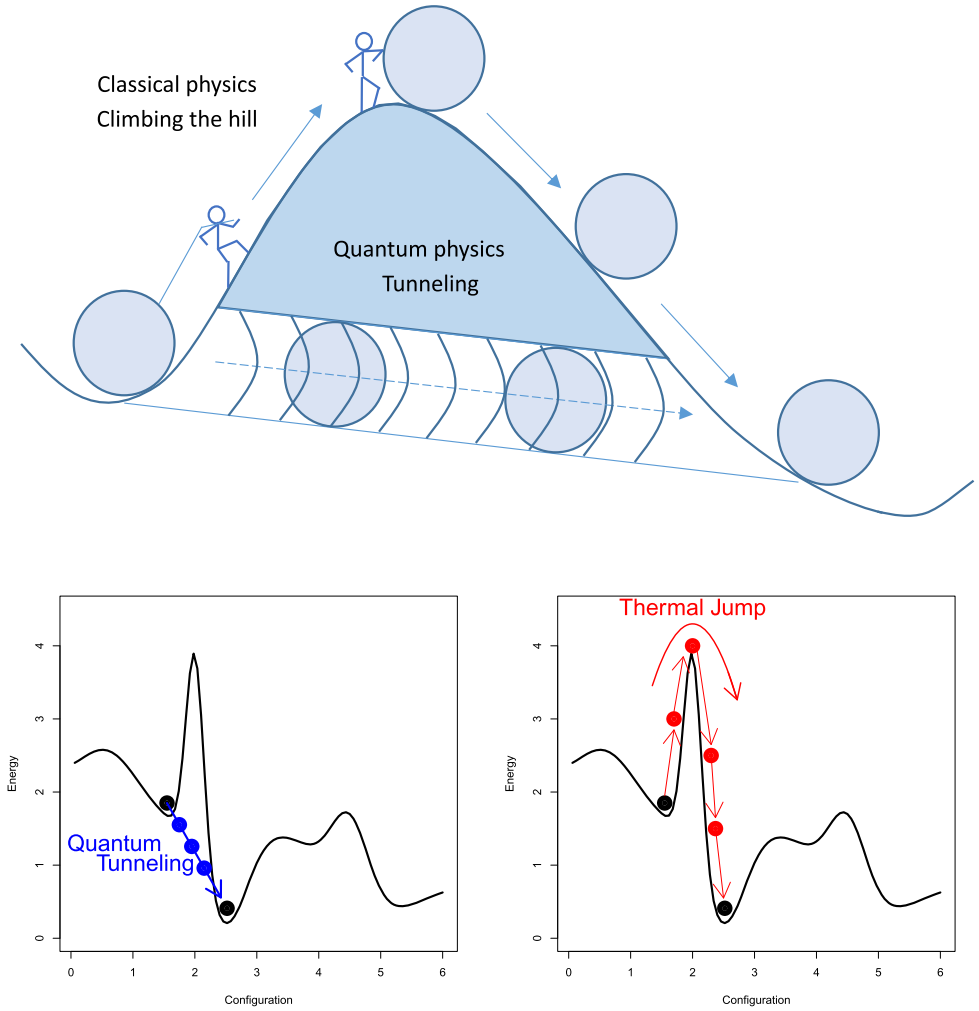

FIG. 1. *A cartoon illustration of quantum tunneling vs. thermal climbing on the top panel with annealing elucidations of quantum tunneling on the left bottom panel and thermal climbing on the right bottom panel.*

more efficient than thermal hopping, and quantum annealing may be faster than classical thermal annealing to reach the equilibrium distribution. In fact, research has shown that it can be more efficient to explore the state space quantum mechanically, and quantum annealing may have advantages over thermal annealing (Brooke et al., 1999, Denchev et al., 2016, Farhi et al., 2000, 2001, 2002, Kechedzhi and Smelyanskiy, 2015 and McGeoch, 2014).

Quantum annealing may be implemented by physical devices or MCMC simulations. We will illustrate in next two sections D-Wave physical devices to implement the unitary dynamic evolution of quantum annealing by natural Schrödingier dynamics and MCMC methods to approximate the dynamic evolution of quantum annealing by artificial time evolutions of Monte Carlo dynamics.

3.3 Physical Implementation of Quantum Annealing by the D-Wave Quantum Computer

The D-Wave quantum computer is a commercially available computing hardware device designed to implement quantum annealing. As illustrated in Figure 2, it includes a quantum processor, a magnetically shielded box and a cooling equipment housed inside a ten square meter shielded room. The quantum processor consists of superconducting flux qubits based on niobium superconducting quantum interference devices, and the shielded box and the cooling equipment protect the system from coupling to the environment and maintain the system at near absolute zero temperature to keep superconducting flux qubits in quantum states. D-Wave One "Rainier" device of 128 qubits and D-Wave Two "Vesuvius" device of 512 qubits were released in May 2011 and May 2013, respectively, while

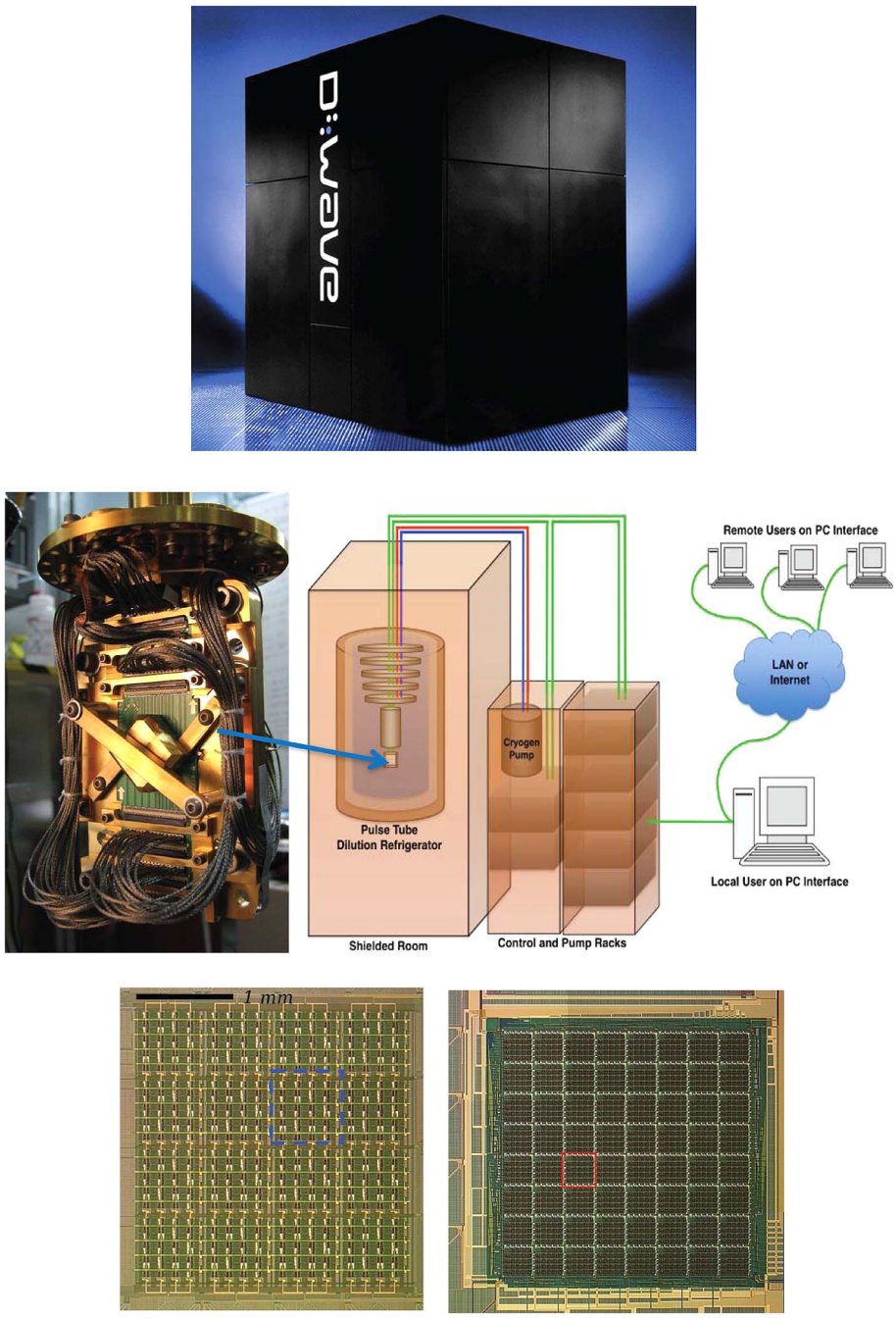

FIG. 2. *The D-Wave quantum computer in the top panel, its inside structural illustration in the middle panel, and* 128 *and* 512 *qubit quantum chips of D-Wave One and Two devices, respectively, in the bottom panel* (*Courtesy of D-Wave Systems Inc.*).

D-Wave's 2X processor with over 1000 qubits was made available in 2015.

The quantum annealing implemented by D-Wave devices is to evolve the quantum Ising system of superconducting flux qubits subject to a transverse field, and the quantum evolution is described by the Schrödinger equation (2.1) via the Ising Hamiltonian $\mathbf{H}_I^q$ in (3.2), the transverse field $\mathbf{H}_X$ in (3.3), and the instantaneous Hamiltonian $\mathbf{H}_D(t)$ in (3.4) for the quantum Ising model in the transverse field, where Figure 3 displays the graph $\mathcal{G}$ employed by the D-Wave One device to describe qubits and their couplings, and plots annealing schedules $A(t)$ and $B(t)$ used by the D-Wave device for carrying out quantum annealing. The graph in Figure 3 is a subgraph of the so-called Chimera graph, where each vertex represents a qubit in the D-

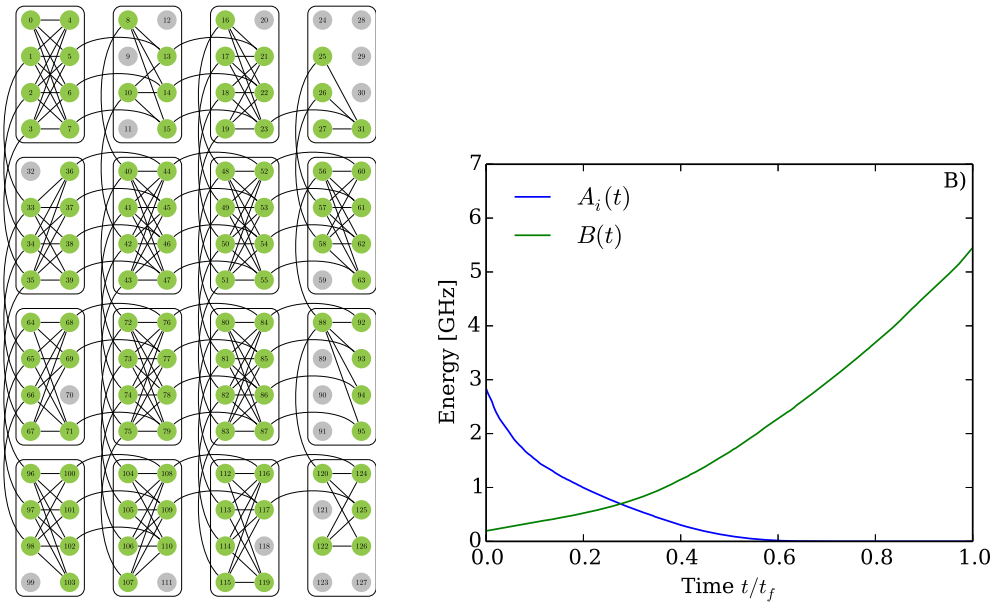

FIG. 3. *The Chimera graph in the left panel and the annealing schedules in the right panel used by D-Wave One quantum computer, which are adopted from Figures* 1 *and* 3 *in Supplementary Information of Boixo et al.* (2014a, 2014b).

Wave One device, and each edge corresponds to the coupler (or interaction) between the two incident vertices with coupling J_{ij} for the edge weight between vertices i and j. The Chimera graph illustrates 128 qubits in the system of the D-Wave One device, and of the 128 qubits on the device, 108 were fully functioning and used in our experiments. In the graph, the working qubits are shown in green, and the couplers between them are marked as black lines. This graph is built from unit cells containing eight vertices each, and within each unit cell the qubits and couplers realize a complete bipartite graph, where within a unit cell each of the four qubits on the left is coupled to all of the four on the right and vice versa; each right qubit in a unit cell is horizontally coupled with the corresponding right qubits in the cells immediately to the right and left of the unit cell, with each left qubit vertically coupled to the corresponding left qubits in the cells immediately above and below the unit cell. There are 16 cells in D-Wave One device, with 64 and 144 cells in the D-Wave Two device and D-Wave 2X device, respectively.

Data used were collected from a computing experiment performed on the D-Wave One device owned by the Lockheed-Martin Corporation and housed at Center for Quantum Information Science & Technology at the University of Southern California (USC) and USC-Lockheed-Martin Quantum Computing Center (Boixo et al., 2014a, 2014b, Albash et al., 2014, and Wang et al., 2013). As the Ising model with local field h_j tends to make corresponding optimization problems easier, the experiment took external magnetic field $h_j = 0$, and we consider the Ising model without external magnetic field h_j in the rest of the paper. In the experiment, we randomly assigned each coupler J_{ij} in the device a value of either $+1$ or -1 to obtain an instance $\{J_{ij}, (i, j) \in \mathcal{E}(\mathcal{G})\}$. The same procedure was repeated to select 1000 sets of different random couplings $J_{ij} = \pm 1$, and thus obtain 1000 different instances. The random selection may result in rough energy landscapes for the quantum system, and this and next two sections apply annealing methods to solve the optimization problem of finding the minimum of $\mathbf{H}_I^c(\mathbf{s})$, and adopt the following procedure to determine if a run of an annealing algorithm finds a ground state: First, some brute force methods can be used to obtain the global minimum of $\mathbf{H}_I^c(\mathbf{s})$ for each selected instance (Dechter, 1999 and Boixo et al., 2014a, 2014b); second, with the known global minimum value of $\mathbf{H}_I^c(\mathbf{s})$ for each selected instance, we declare that a particular run of an annealing algorithm finds a ground state if the annealing run yields a minimum value of $\mathbf{H}_I^c(\mathbf{s})$ equal to its global minimum value. For each of the 1000 selected instances, after its couplings were programmed, 1000 annealing runs were performed on the device to determine whether the system reached a ground state during each run and compute the frequency of finding the ground states among the 1000 runs to estimate the ground state success probability for the instance. Each annealing operation was carried out

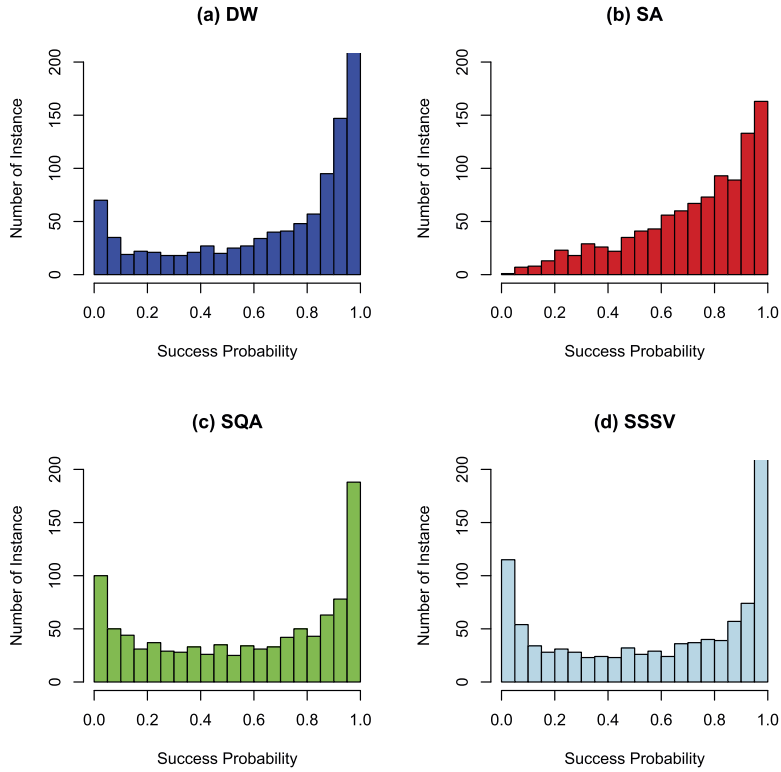

FIG. 4. *Histogram plots of ground state success probability data for DW, SA, SQA and SSSV in* (a)–(d), *respectively.*

at a temperature of 20 milli-Kelvin, with an annealing time $t_f = 5$ microseconds. While the transverse field starts initially at $A(0) \approx 3$ GHz and decreases to zero during the annealing process, the couplings field increases from near zero to about $B(t_f) \approx 5$ GHz at the end of the annealing process.

For each of the 1000 selected instances, we computed the frequency of finding the ground states among the 1000 runs, and the obtained 1000 frequencies consist of the ground state success probability data from the D-Wave (DW) experiment. Figure 4(a) displays the histogram of the DW ground state success probability data. We plotted in Figure 4(b) the histogram of ground state success probability data generated from SA described in Section 3.1, where the same Chimera graph and 1000 instances as in the DW case were employed by the classical Ising model in SA with annealing schedule T_k proportional to $\frac{1}{k}$, 1000 runs were used to obtain the ground state success probability data for each instance, and each annealing run was carried out with 10,000 sweeps. One striking difference between the two histograms in Figure 4(a), (b) is the bimodal vs. unimodal phenomenon that the DW histogram exhibits a U-shape pattern, while the SA histogram shows a unimodal (including monotone) pattern. In fact, we ran SA under the same set-up except for different annealing times with various sweeps and plotted the histograms of ground state success probability data in Figure 6. The plots show that the SA data exhibit unimodal histogram patterns for all sweeps, and as the number of sweeps increases, the mode shifts toward the larger success probability side and gets stronger. We illustrated in Figure 5(a) the scatter plot of DW against SA data displayed in Figure 4(a), (b), and found their correlation to be 0.84.

3.4 Classical MCMC Simulations to Approximate Quantum Annealing

It is extremely difficult to simulate quantum annealing by classical computers. The matrix size of Hamiltonians in quantum annealing exponentially increases with its system size, and the simulations require to exponentiate such exponentially large noncommutable matrices, which is not feasible by classical computers. In this section, we introduce two classical MCMC simulation methods, simulated quantum annealing (SQA) and the SSSV annealing method in Shin et al. (2014), to approximate quantum annealing and provide possible candidate models for the D-Wave experimental data. Like SA they are classical MCMC methods. However, there are some intrinsic differences.

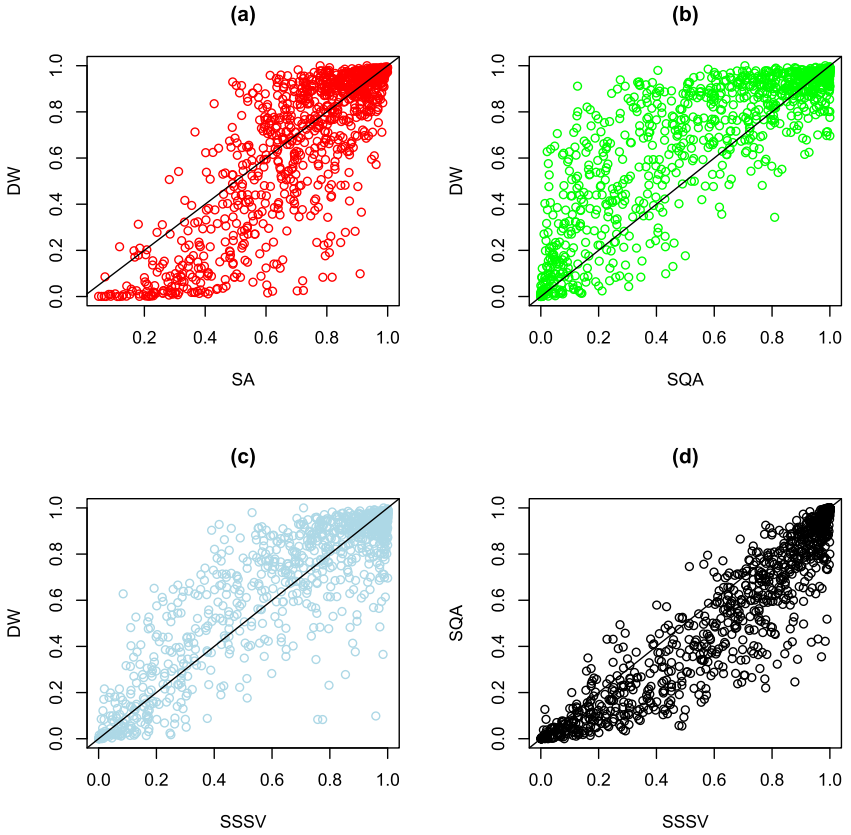

FIG. 5. *Ground state success probability data scatter plots of DW against SA, SQA, SSSV in* (a)–(c), *respectively, along with scatter plot of SQA vs SSSV in* (d).

SA solves the minimization problem of an objective function by starting with a random initial spin configuration and flipping spins at random at each time step, we always accept a new spin configuration if it lowers the energy and accept it probabilistically using the Metropolis rule otherwise, and we gradually lower the temperature to reduce the escape probability of trapping in local minima. It is a pure thermal annealing. Quantum annealing replaces spin flips and the thermal escape rule by introducing the noncommuting transverse field and reducing the escape probability by turning on and then off this noncommuting field adiabatically to allow quantum tunneling out of local minima. SQA and SSSV are two classical MCMC simulations to approximate quantum annealing by path-integral formulation and mean field approximation, respectively. These two annealing methods employ the same schedules in quantum annealing to mimic some effect of generating the Ising system with quantum fluctuations based on magnetic fields, and to make the system to ultimately end up in its lowest energy configuration, which represents the solution to the optimization problem. The MCMC methods bear important practical significance as they can treat practical large-size problems on classical computers.

3.4.1 *Simulated quantum annealing*. For quantum annealing Hamiltonian $\mathbf{H}_D(t)$ in (3.4), to find the Boltzmann state of the quantum system we need to evaluate canonical partition function $\mathrm{tr}[e^{-\mathbf{H}_D(t)/T}]$ of the transverse field quantum Ising model. However, we can not analytically exponentiate $e^{-\mathbf{H}_D(t)/T}$ to evaluate the partition function, since the two matrices in $\mathbf{H}_D(t)$, Ising Hamiltonian $\mathbf{H}_I^q$ and transverse field Hamiltonian $\mathbf{H}_X$, do not commute. As described in Martoňák, Santoro and Tosatti (2002), the basic idea of SQA is to apply the path-integral technique to the canonical partition by using the Trotter formula (Kato, 1978, Suzuki, 1976 and Trotter, 1959) to obtain an approximation of the partition function for the quantum annealing Hamiltonian $\mathbf{H}_D(t)$. Specifically substituting $T/B(t)$ for temperature parameter and $A(t)/B(t)$ for transverse field parameter in Martoňák et al. (2002), equations (2)–(6), we obtain an approximation of the partition function for the quantum annealing Hamiltonian $\mathbf{H}_D(t)$ with temperature T that maps the trans-

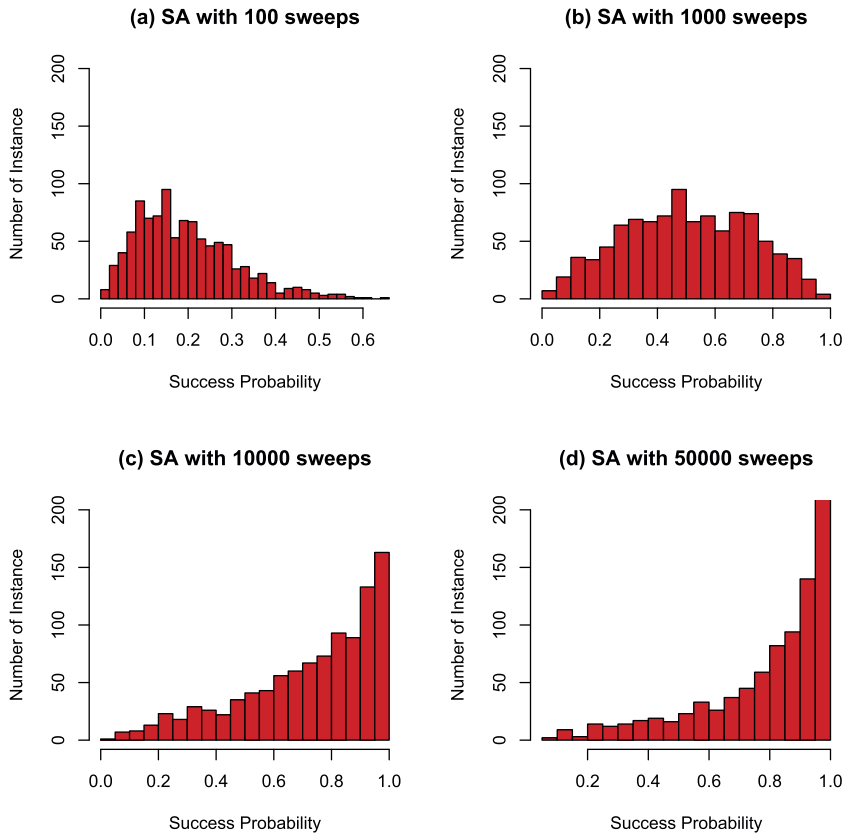

FIG. 6. *Histogram plots of ground state success probability data for SA with different annealing sweeps.*

verse quantum Ising model to a classical $(2+1)$-dimensional anisotropic Ising model with temperature τT and Hamiltonian

$$\mathbf{H}_{aI}^{c}(\mathbf{s}) = -\sum_{\ell=1}^{\tau}\Bigg[B(t)\sum_{(i,j)\in\mathcal{E}(\mathcal{G})} J_{ij}s_{i\ell}s_{j\ell} + J(t)\sum_{j\in\mathcal{V}(\mathcal{G})} s_{j\ell}s_{j,\ell+1}\Bigg], \tag{3.5}$$

where $s_{j\ell}$ are random variables taking values in $\{+1, -1\}$, τ is an integer, ℓ is the index for an extra dimension that is often referred to as the imaginary-time dimension, J_{ij} are the couplings along the original 2-dimensional direction of the Ising model described by (3.1) on graph $\mathcal{G}$,

$$J(t) = -\frac{\tau T}{2}\ln\left[\tanh\left(\frac{A(t)}{\tau T}\right)\right]$$

is the coupling along the imaginary-time dimension, and $A(t)$ and $B(t)$ are the same annealing schedules as in quantum annealing.

Let $\mathbf{s}_\ell = \{s_{i\ell}, i = 1, \ldots, b\}$ denote a configuration of b spins that is called the ℓth Trotter slice, here $1 \le \ell \le \tau$, and b is the total number of the vertices in graph $\mathcal{G}$. The classical anisotropic Ising model has two directions, one along the original 2-dimensions described by graph $\mathcal{G}$ and another along the imaginary-time dimension. In the original 2-dimensional graph direction the model has coupling J_{ij} and annealing schedule $B(t)$, where the couplings are the same for all τ Trotter slices, and in the imaginary-time direction the model has a finite length τ and uniform coupling, with annealing schedule $J(t)$.

Since $\mathbf{H}_{aI}^{c}$ is a classical anisotropic Ising Hamiltonian, the path-integral representation of the transverse field quantum Ising model allows approximate simulations of the quantum system by classical MCMC sampling from the classical anisotropic Ising model. Because of the extra dimension in the classical anisotropic Ising model, the MCMC simulation method needs to employ a standard Metropolis algorithm with both local and global moves to sample the system. For the local moves, we perform usual independent spin flips at all sites in all Trotter slices, while each global move attempts to flip simultaneously all the replicas

of the same spin in all Trotter slices. It is clear that in the Metropolis procedure, the acceptance ratio of the global moves is free of the transverse field $J(t)$, because we always flip at the same time both replicas of the same spin in neighboring Trotter slices, which interact via $J(t)$, and thus depend on $A(t)$. To choose the initial configuration, we may assume that there is little interactions between the neighboring Trotter slices, and the $(2+1)$-dimensional anisotropic Ising system behaves like a collection of noninteracting two-dimensional systems, and an appropriate choice of the initial configuration may be to set all Trotter slices equal to spin configurations corresponding to the equilibrium Boltzmann distribution, which can be achieved with some burn-in simulations.

The SQA algorithm starts with random initialization: initiate spins in all Trotter slices with -1 and $+1$ at random and independent among spins and Trotter slices to obtain initial spin configuration $\mathbf{s}^{(0)} = \{s_{j\ell}^{(0)}, j \in \mathcal{G}(\mathcal{V}), \ell = 1, \ldots, \tau\}$. We update spins one by one for local move and spin replicas site by site for global move. Each complete updating all spins locally and globally constitutes one sweep. Denote by R the total number of sweeps in SQA, and let $t_k = k/R$, $k = 1, \ldots, R$. At the kth sweep, in the case of local move, for spin i in the ℓth Trotter slice, we attempt to flip state $s_{i\ell}^{(k-1)}$ to new state $s_{i\ell}^{(k)} = -s_{i\ell}^{(k-1)}$ while keeping all others unchanged, and calculate energy change between its original state $s_{i\ell}^{(k-1)}$ and newly flipped state $s_{i\ell}^{(k)}$,

$$\begin{aligned}\Delta E_{1i\ell}^{(k)} = &-B(t_k)\Bigg[\sum_{j=1}^{i-1} J_{ij} s_{j\ell}^{(k)}\big(s_{i\ell}^{(k)} - s_{i\ell}^{(k-1)}\big) \\ &+ \sum_{j=i+1}^{b} J_{ij} s_{j\ell}^{(k-1)}\big(s_{i\ell}^{(k)} - s_{i\ell}^{(k-1)}\big)\Bigg] \\ &- J(t_k)\big[s_{i\ell}^{(k)} s_{i,\ell+1}^{(k)} + s_{i,\ell-1}^{(k)} s_{i\ell}^{(k)} \\ &- s_{i\ell}^{(k-1)} s_{i,\ell+1}^{(k-1)} - s_{i,\ell-1}^{(k-1)} s_{i\ell}^{(k-1)}\big].\end{aligned}$$

The local move accepts the new state $s_{i\ell}^{(k)}$ with probability $\min\{1, \exp[-\Delta E_{1i\ell}^{(k)}/(\tau T)]\}$. At the kth sweep, in the case of global move, for spins at site i of all Trotter slices, we attempt to flip states $\{s_{i\ell}^{(k-1)}, \ell = 1, \ldots, \tau\}$ to new states $\{s_{i\ell}^{(k)} = -s_{i\ell}^{(k-1)}, \ell = 1, \ldots, \tau\}$ while keeping all others unchanged, and calculate energy change between its original states $\{s_{i\ell}^{(k-1)}, \ell = 1, \ldots, \tau\}$ and newly flipped states $\{s_{i\ell}^{(k)}, \ell = 1, \ldots, \tau\}$,

$$\begin{aligned}\Delta E_{2i}^{(k)} = -\sum_{\ell=1}^{\tau} B(t_k)\Bigg[&\sum_{j=1}^{i-1} J_{ij} s_{j\ell}^{(k)}\big(s_{i\ell}^{(k)} - s_{i\ell}^{(k-1)}\big) \\ &+ \sum_{j=i+1}^{b} J_{ij} s_{j\ell}^{(k-1)}\big(s_{i\ell}^{(k)} - s_{i\ell}^{(k-1)}\big)\Bigg].\end{aligned}$$

The global move accepts the new states $\{s_{i\ell}^{(k)}, \ell = 1, \ldots, \tau\}$ with probability $\min\{1, \exp[-\Delta E_{2i}^{(k)}/(\tau T)]\}$. Finally, to evaluate the original classical Hamiltonian $\mathbf{H}_I^c(\mathbf{s})$, we simply use the first Trotter slice at the $k = R$th sweep by taking $s_i^{(k)} = s_{i1}^{(k)}$, and then obtain $\mathbf{s}^{(k)} = \{s_i^{(k)}, i \in \mathcal{V}(\mathcal{G})\}$ and evaluate $\mathbf{H}_I^c(\mathbf{s}^{(k)})$ given by (3.1) with $h_j = 0$.

The path-integral formulation maps quantum annealing to a classical $(2+1)$-dimensional anisotropic Ising model, where the model in the imaginary-time direction is made completely decoupled and identical, and SQA is a MCMC based method to perform a path-integral simulation of a transverse field quantum Ising model, where the system is brought into the equilibrium state at each MCMC sweep and then a small change is made in the Hamiltonian, after which the system is again brought into the equilibrium state and so on so forth (Martoňák, Santoro and Tosatti, 2002, and Morita and Nishimori, 2008).

Using the same Chimera graph and 1000 instances along with their corresponding global minima of $\mathbf{H}_I^c(\mathbf{s})$ as in DW, and taking the following annealing schedules close to the DW annealing schedule curves displayed in Figure 3,

$$\begin{aligned}A(t) &= \begin{cases} 8t^2 - 9.6t + 2.88, & \text{if } 0 \le t \le 0.6, \\ 0, & \text{if } 0.6 < t \le 1, \end{cases} \\ B(t) &= 5.2t^2 + 0.2t, \quad t \in [0, 1],\end{aligned} \tag{3.6}$$

we ran the SQA algorithm with up to 10,000 sweeps, temperature $0.1 \le T \le 1$ and various Trotter slices τ. For each run of the SQA algorithm on a selected instance, as in the DW case we determined whether the run found a ground state of $\mathbf{H}_I^c(\mathbf{s})$ by comparing the minimum value yielded from the run of the SQA algorithm with the known global minimum value of $\mathbf{H}_I^c(\mathbf{s})$ corresponding to the instance, and for each instance we repeatedly ran the SQA algorithm 1000 times to compute the frequency of finding the ground states over the 1000 times. We displayed in Figure 4(c) the histogram of ground state success probability data generated by the SQA algorithm with $\tau = 30$, $T = 0.1$ and

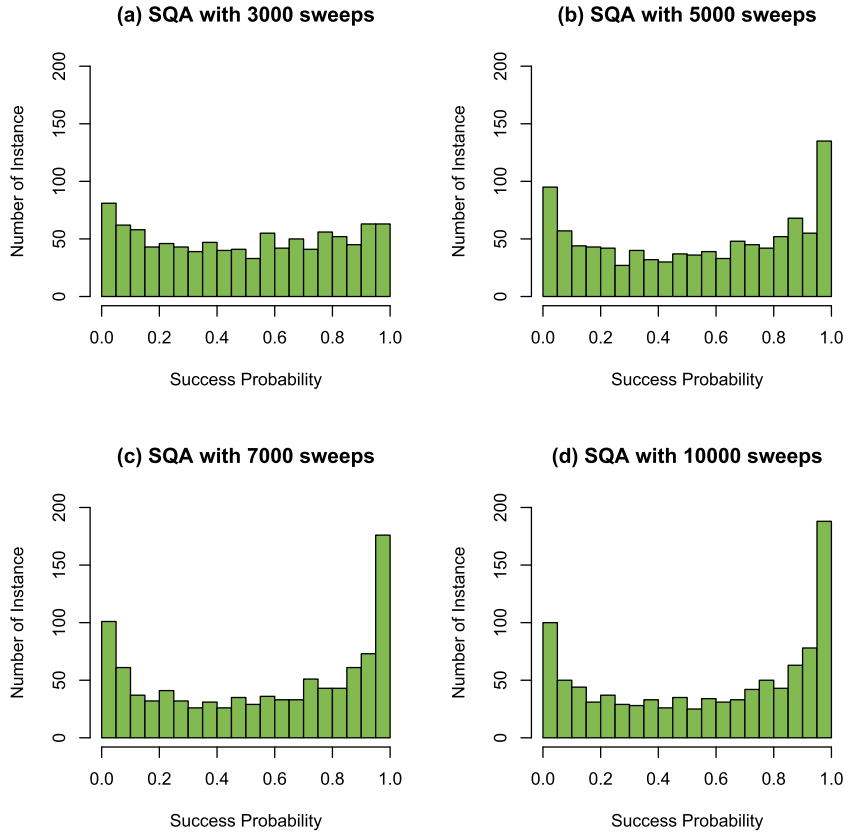

FIG. 7. *Histogram plots of ground state success probability data for SQA with temperature $T = 0.1$ and different annealing sweeps.*

10,000 sweeps. The histogram exhibits a U-shape pattern similar to the DW case. We also ran the SQA algorithm with different temperatures and various sweeps and plotted the histograms of ground success probability data in Figures 7–8. The plots show that at $T = 0.1$, the SQA data exhibit U-shape histogram patterns for all sweeps, and as the number of sweeps increases, the patterns grow stronger; as the temperature T increases when the number of sweeps is fixed to be 10,000, the U-shape patterns get weaker, eventually disappear and become unimodal patterns. This indicates that as temperature increases, thermal effects tend to dominate quantum effects, and the SQA histogram patterns approach to the SA histogram patterns. We illustrated in Figure 5(b) the scatter plot of DW against SQA data displayed in Figure 4(c), and found their correlation to be 0.85.

3.4.2 *The SSSV annealing*. The SSSV model is a MCMC simulation model proposed by Shin et al. (2014) and Smolin and Smith (2014). It replaces qubits at the Chimera graph by planar rotors with angles $\theta_i \in [0, 2\pi]$ and Pauli matrices $\boldsymbol{\sigma}_i^x$ and $\boldsymbol{\sigma}_i^z$ in quantum Hamiltonian $\mathbf{H}_D(t)$ in (3.4) by $\sin\theta_i$ and $\cos\theta_i$, respectively. The resulted rotor system is governed by Hamiltonian

$$\begin{aligned}\mathbf{H}_S(t) = &-A(t) \sum_{j\in\mathcal{V}(\mathcal{G})} \sin\theta_j \\ &- B(t) \sum_{(i,j)\in\mathcal{E}(\mathcal{G})} J_{ij}\cos\theta_i\cos\theta_j,\end{aligned} \tag{3.7}$$

where graph $\mathcal{G}$, couplings J_{ij}, and annealing schedules $A(t)$ and $B(t)$ are the same as in quantum annealing such as those used by DW and SQA. The time evolution of the system can be generated by standard MCMC algorithms with a Metropolis scheme to update rotor angles as follows. Starting with initial values θ_i, $i = 1, \ldots, b$, each selected at random from $[0, 2\pi]$, one variable at a time we generate angle θ_i from the uniform distribution on $[0, 2\pi]$. The generated angle is accepted if it does not increase the energy given by $\mathbf{H}_S(t)$. If the generated angle increases the energy, it is accepted with probability $\exp(-\Delta E/T)$, where $\Delta E > 0$ is the change in energy. A sweep is referred to as each complete update of all b rotors during the entire annealing schedule. After each sweep, the time t in Hamiltonian $\mathbf{H}_S(t)$ is increased by one time unit until the final annealing time.

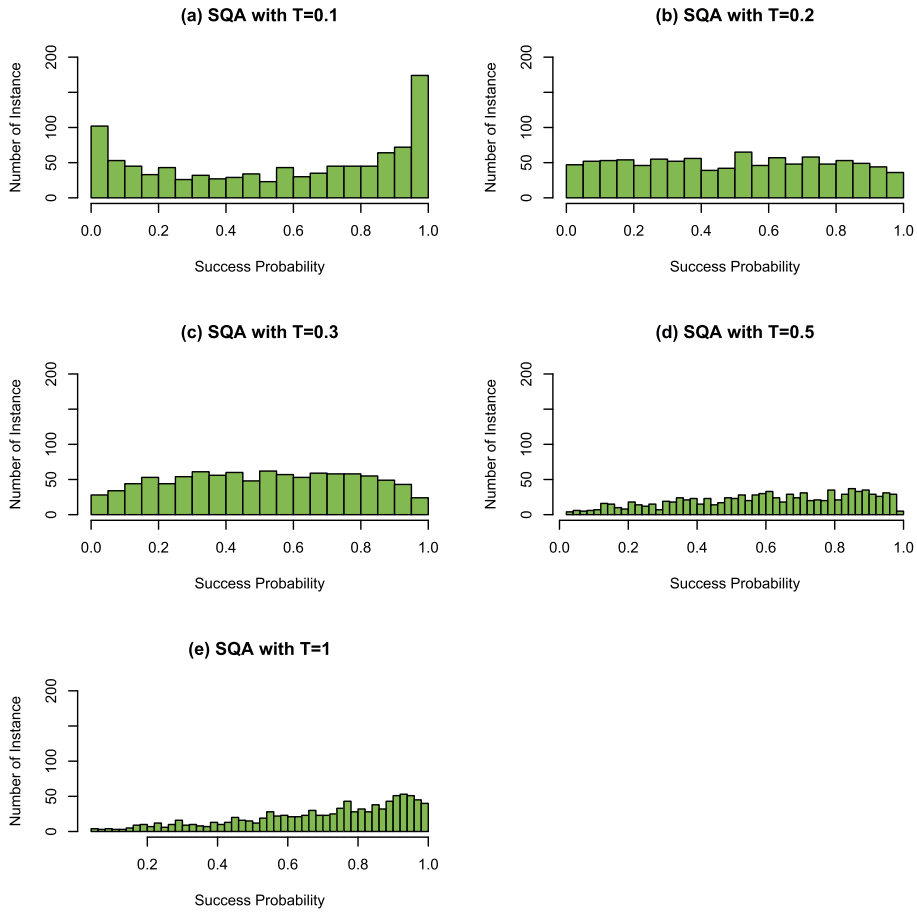

FIG. 8. *Histogram plots of ground state success probability data for SQA with* 10,000 *sweeps and different temperatures.*

Specifically, as in SQA, denote by R the total number of sweeps and $t_k = k/R$ the time step at the kth sweep, $k = 1, \ldots, R$. We perform the Metropolis updating scheme one variable at a time in each time step as follows. Starting with random initial values $\theta_i^{(0)}$ selected independently from the uniform distribution on $[0, 2\pi]$, we generate an angle $\theta_i^{(k)}$ from the uniform distribution on $[0, 2\pi]$ for vertex i at the kth sweep.

1. For rotor i, compute the change in energy between its original state $\theta_i^{(k-1)}$ and the newly generated state $\theta_i^{(k)}$,

$$\begin{aligned}\Delta E_i^{(k)} = &-A(t_k)\big(\sin\theta_i^{(k)} - \sin\theta_i^{(k-1)}\big)\\ &- B(t_k)\Bigg[\sum_{j=1}^{i-1} J_{ij}\cos\theta_j^{(k)}\big(\cos\theta_i^{(k)}\\ &- \cos\theta_i^{(k-1)}\big)\\ &\cdot \sum_{j=i+1}^{b} J_{ij}\cos\theta_j^{(k-1)}\big(\cos\theta_i^{(k)} - \cos\theta_i^{(k-1)}\big)\Bigg],\end{aligned}$$

where $\theta_1^{(k)}, \ldots, \theta_{i-1}^{(k)}, \theta_{i+1}^{(k-1)}, \ldots, \theta_b^{(k-1)}$ denote the states of the other rotors when updating rotor i.

2. Change rotor i's state from $\theta_i^{(k-1)}$ to new state $\theta_i^{(k)}$ if $\Delta E_i^{(k)} \le 0$ and otherwise update its state with probability $\exp(-\Delta E_i^{(k)}/T)$, where T is the temperature of the system which is kept constant throughout the simulation.

As each complete time evolution following the entire annealing schedule constitutes one sweep, the kth sweep corresponds to a set of angles $\{\theta_j^{(k)}, j = 1, \ldots, b\}$. We interpret angles in terms of states in the computational basis according to the signs of $\cos(\theta_j^{(k)})$, that is, define vertex variables s_j, $j \in \mathcal{V}(\mathcal{G})$, at the kth sweep according their corresponding angles $\theta_j^{(k)}$ as follows:

$$s_j^{(k)} = \operatorname{sign}\big(\cos\theta_j^{(k)}\big), \quad \mathbf{s}^{(k)} = \big\{s_j^{(k)}, j \in \mathcal{V}(\mathcal{G})\big\}.$$

We plug the obtained configuration $\mathbf{s}^{(k)}$ into the original classical Ising Hamiltonian $\mathbf{H}_I^c$ given by (3.1) with

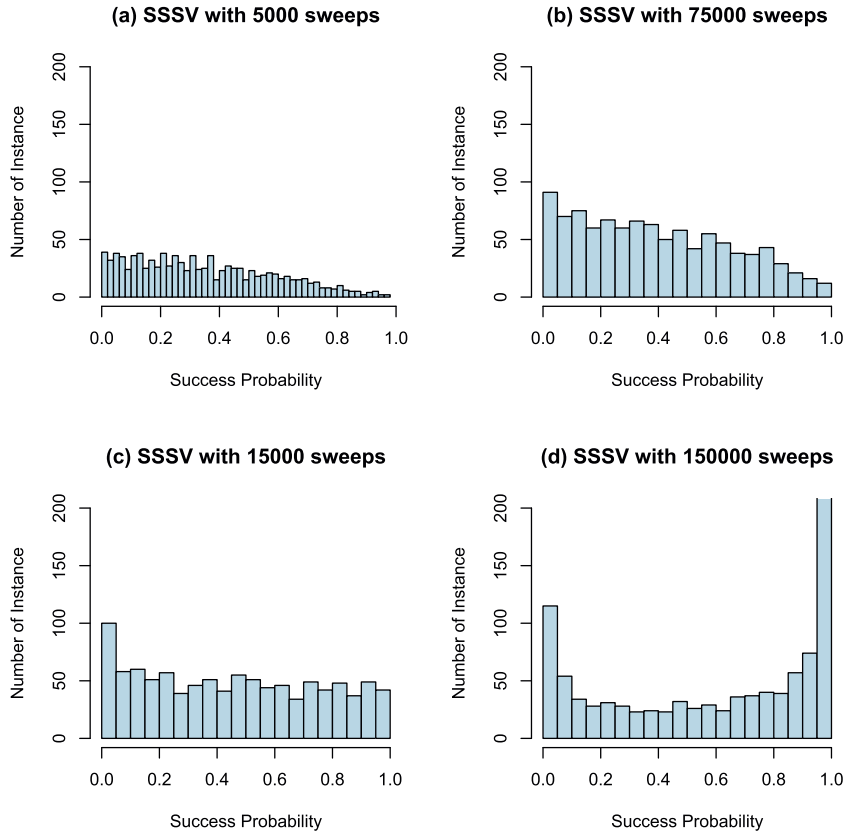

FIG. 9. *Histogram plots of ground state success probability data for SSSV with temperature* $T = 0.1$ *and different annealing sweeps.*

$h_j = 0$ to evaluate

$$\mathbf{H}_I^c(\mathbf{s}^{(k)}) = -\sum_{(i,j)\in\mathcal{E}(\mathcal{G})} J_{ij} s_i^{(k)} s_j^{(k)}.$$

Similar to the SA procedure we can find the global minimum of $\mathbf{H}_I^c(\mathbf{s})$ with enough repetitions.

Both SSSV and SQA are classical MCMC simulation methods to approximate quantum annealing by following the same annealing schedules as in quantum annealing. While SQA uses the path-integral formulation to obtain an approximation of the quantum annealing by a classical $(2+1)$-dimensional anisotropic Ising model, the SSSV model may be understood as approximating quantum annealing by a mean-field approximation of the path-integral formulation.

As in Shin et al. (2014), similar to the SQA case, we ran the SSSV algorithm by using the same Chimera graph, 1000 instances and annealing schedules $A(t)$ and $B(t)$ as in DW and SQA, where we set temperature $0.1 \le T \le 1$ and sweeps up to 150,000. For each combination of instance, temperature and sweep set-up, we repeatedly ran the SSSV algorithm 1000 times to compute the frequency of successfully finding the ground states over the 1000 times. We displayed in Figure 4(d) the histogram of ground state success probability data generated by the SSSV algorithm with 150,000 sweeps and $T = 0.1$. The histogram exhibits a U-shape pattern similar to the DW and SQA cases. We ran the SSSV algorithm with various sweeps and temperatures and plotted the histograms of ground state success probability data in Figures 9–10. At temperature $T = 0.1$, the SSSV data initially exhibit decreasing patterns, and as the number of sweeps increases, the decreasing patterns become weaker, the histograms at two ends get larger, and eventually the histograms display strong U-shape patterns similar to the DW and SQA cases. When the number of sweeps is fixed to be 150,000, as the temperature increases, the U-shape patterns get weaker, finally fade away and turn into monotone patterns. We illustrated in Figure 5(c)–(d) the scatter plots of DW against SSSV data and of SQA against SSSV data displayed in Figure 4(c)–(d), and computed their correlations to be 0.86 and 0.94, respectively. The similar behaviors of SQA and SSSV may be explained by their extremely high correlation.

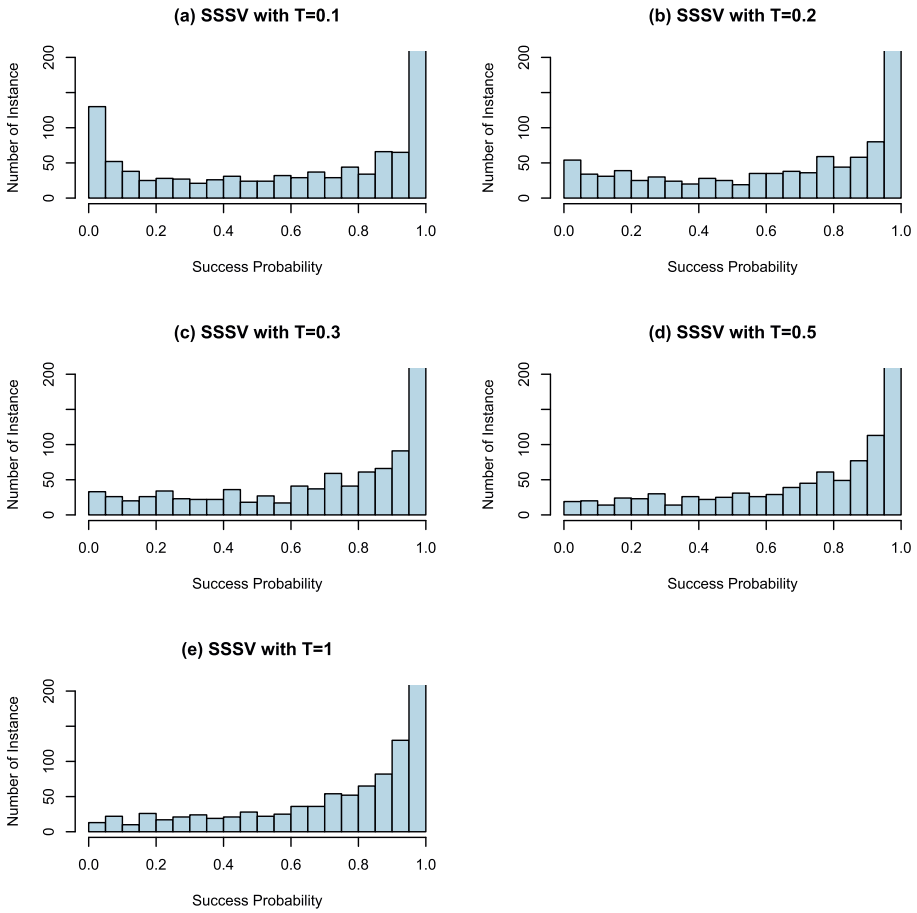

FIG. 10. *Histogram plots of ground state success probability data for SSSV with* 150,000 *sweeps and different temperatures.*

4. STATISTICAL INFERENCE FOR GROUND STATE SUCCESS PROBABILITY DATA

Our motivation and goal are to develop statistical methodologies to analyze annealing data described in Section 3. First, we introduce notation and state statistical problems. Suppose that we have selected n instances for the Ising model, and for a given instance, we run each of the DW, SQA, SSSV and SA algorithms to find the ground states with the same parameter set-up repeatedly for m times, and their frequencies of successfully finding the ground states are obtained for the instance. For the rth instance, $r = 1, \ldots, n$, denote by $\widehat{p}_{0,m,r}$, $\widehat{p}_{1,m,r}$, $\widehat{p}_{2,m,r}$ and $\widehat{p}_{3,m,r}$ the ground state success frequencies among m repeated runs for DW, SA, SQA and SSSV, respectively, and also denote by $p_{0,\infty,r}$, $p_{1,\infty,r}$, $p_{2,\infty,r}$ and $p_{3,\infty,r}$ the true probability of successfully finding the ground states by the DW, SA, SQA and SSSV algorithms, respectively. The true ground state success probabilities may be viewed as the frequency limits of running the algorithms an infinite number of times. With ground state success probability data, we may develop statistical tests to detect if there is any difference between two methods with respect to probability distributions and their patterns over the instances.

4.1 Statistical Tests of Ground State Success Probability

4.1.1 *Methodologies and theory.* For any two annealing methods, we may test whether they yield the same ground state success probabilities for each instance or all instances. In particular, we are interested in testing the equality of ground state success probabilities between DW and one of the MCMC based annealing methods. First, consider instance based multiple tests. For the rth instance define null hypothesis $H_0 : p_{k,\infty,r} = p_{\ell,\infty,r}$ against alternative $H_a : p_{k,\infty,r} \neq p_{\ell,\infty,r}$, where $k \neq \ell$ and $k, \ell = 0, 1, 2, 3$. It is easy to see that $m\widehat{p}_{k,m,r}$ follows a binomial distribution $\text{Bin}(m, p_{k,\infty,r})$. Applying the two sample binomial

proportion test, we define a test statistic

$$T_{rm} = \frac{m(\widehat{p}_{k,m,r} - \widehat{p}_{\ell,m,r})^2}{\widehat{p}_{k,m,r}(1-\widehat{p}_{k,m,r}) + \widehat{p}_{\ell,m,r}(1-\widehat{p}_{\ell,m,r})}. \tag{4.1}$$

For each r, when m is large, the null distribution of T_{rm} can be approximated by a χ^2 distribution with one degree of freedom. We may use the variance stabilization transformation arcsin to transform the frequency data and define

$$\begin{aligned} T_{0rm} = 2m\big[&\arcsin\big(\sqrt{\widehat{p}_{k,m,r}}\big) \\ &- \arcsin\big(\sqrt{\widehat{p}_{\ell,m,r}}\big)\big]^2. \end{aligned} \tag{4.2}$$

Again for each r, for large m the null distribution of T_{0rm} can be approximated by a χ^2 distribution with one degree of freedom. Since there are n hypotheses, for large n we need the following uniform convergence over $r = 1, \ldots, n$ to provide asymptotic justifications for carrying out the multiple tests.

THEOREM 1. *Assume that $p_{k,\infty,r}$, $k = 0, 1, 2, 3$, $r = 1, \ldots, n$, are bounded away from 0 and 1. Then given $k \neq \ell$, as $m, n \to \infty$,*

$$\sup_{1\le r\le n} \sup_{-\infty<x<\infty} \big|\Psi_{rm}(x) - \Psi(x)\big| \to 0,$$

$$\sup_{1\le r\le n} \sup_{-\infty<x<\infty} \big|\Psi_{0rm}(x) - \Psi(x)\big| \to 0,$$

where Ψ is the distribution function of χ_1^2, Ψ_{rm} and Ψ_{0rm} are the distribution functions of T_{rm} and T_{0rm} under the null hypotheses $H_0 : p_{k,\infty,r} = p_{\ell,\infty,r}$, respectively.

Theorem 1 provides asymptotic justifications for uniform approximations of the p-values for all n hypotheses by their corresponding large sample p-values. For testing $H_0 : p_{k,\infty,r} = p_{\ell,\infty,r}$ against $H_a : p_{k,\infty,r} \neq p_{\ell,\infty,r}$, the p-values for the tests based on T_{rm} and T_{0rm} are equal to $1 - \Psi_{rm}(|T_{rm}|)$ and $1 - \Psi_{0rm}(|T_{0rm}|)$, respectively. By Theorem 1, we can use χ_1^2 distribution to approximate the p-values by $1 - \Psi(|T_{rm}|)$ and $1 - \Psi(|T_{0rm}|)$ uniformly over $r = 1, \ldots, n$. The obtained p-values can be used to carry out the FDR procedure in the multiple hypothesis testing.

Now we consider the goodness-of-fit test for all instances. For testing null hypothesis $H_0 : p_{k,\infty,r} = p_{\ell,\infty,r}$ for all instances $r = 1, \ldots, n$ against alternative $H_a : p_{k,\infty,r} \neq p_{\ell,\infty,r}$ for some r, where $k \neq \ell$ and $k, \ell = 0, 1, 2, 3$, we define test statistics based on T_{rm} and T_{0rm} as follows:

$$\begin{aligned} U_{mn} &= \sum_{r=1}^{n} \frac{T_{rm} - 1}{\sqrt{2n}} \\ &= (2n)^{-1/2} \\ &\quad \cdot \Bigg\{ \sum_{r=1}^{n} \big(m(\widehat{p}_{k,m,r} - \widehat{p}_{\ell,m,r})^2\big) / \big(\widehat{p}_{k,m,r}(1 \\ &\qquad - \widehat{p}_{k,m,r}) + \widehat{p}_{\ell,m,r}(1 - \widehat{p}_{\ell,m,r})\big) - n \Bigg\}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} U_{0mn} &= \sum_{r=1}^{n} \frac{T_{0rm} - 1}{\sqrt{2n}} \\ &= (2n)^{-1/2} \Bigg\{ 2m \sum_{r=1}^{n} \big[\arcsin\big(\sqrt{\widehat{p}_{k,m,r}}\big) \\ &\qquad - \arcsin\big(\sqrt{\widehat{p}_{\ell,m,r}}\big)\big]^2 - n \Bigg\}, \end{aligned} \tag{4.4}$$

where k and ℓ denote any two of the four methods. Although the null distributions of T_{rm} and T_{0rm} can be approximated by χ_1^2 with an error of order $m^{-1/2}$, the naive calculations immediately show that the null distribution of U_{mn} may be approximated by a standard normal distribution with an error of order $(n/m)^{1/2}$, which renders a difficulty in justifying applications of the large sample tests to the D-Wave data with $m = n = 1000$. Delicate asymptotic analysis reveals that the normal approximation actually has an error order of $n^{1/2}/m$, which provides asymptotic justifications for the D-Wave testing problem. The result is given in the following theorem.

THEOREM 2. *Assume that $p_{k,\infty,r}$, $k = 0, 1, 2, 3$, $r = 1, \ldots, n$, are bounded away from 0 and 1, and as $n, m \to \infty$, $\sqrt{n}/m \to 0$. Then given $k \neq \ell$, under the null hypothesis that $p_{k,\infty,r} = p_{\ell,\infty,r}$ for $1 \le r \le n$, U_{mn} and U_{0mn} converge in distribution to a standard normal distribution as $n, m \to \infty$.*

For large sample by Theorem 2, we may approximate the p-values of the goodness-of-fit tests based on U_{mn} and U_{0mn} by $1 - \Phi(U_{mn})$ and $1 - \Phi(U_{0mn})$, respectively, where Φ is the standard normal distribution function.

4.1.2 *Simulations.* Our statistical tests are based on large sample theory where the null limiting distributions require the numbers of runs and/or instances to go to infinity. We need to assess how good are the approximations of the finite sample null distributions

by the null limiting distributions for the experimental data with $m = 1000$ runs and $n = 1000$ instances described in Section 3. A statistical simulation study was conducted to carry out the assessment of the finite sample performances of the approximations. First, we need to set some plausible true success probability scenarios for the simulation study. We explored the success probability scenarios by taking the common true success probability for the two methods under consideration to be the success frequency obtained from each of DW, SA, SQA and SSSV and displayed in Figure 4. Specifically, we took the true common success probability as the ground state success frequencies of DW, SA with 10,000 sweeps, SQA with $T = 0.1$ and 10,000 sweeps and SSSV with $T = 0.1$ and 150,000 sweeps. For each of the four scenarios, denote by p_r, $r = 1, \ldots, 1000$, the true common success probabilities for the 1000 instances. We simulated $1000\widehat{p}_{1r}$ and $1000\widehat{p}_{2r}$ from the binomial distribution $\text{Bin}(1000, p_r)$, $r = 1, \ldots, 1000$, and used $(\widehat{p}_{1r}, \widehat{p}_{2r})$ to substitute $(\widehat{p}_{k,m,r}, \widehat{p}_{\ell,m,r})$ in (4.1)–(4.4) and evaluate a value set for $(T_{rm}, T_{0rm}, U_{mn}, U_{0mn})$. We repeated the procedure 500 times to obtain 500 value sets for $(T_{rm}, T_{0rm}, U_{mn}, U_{0mn})$, and used them to compute the empirical distribution functions of T_{rm}, T_{0rm}, U_{mn} and U_{0mn}. We compared the empirical distribution functions of T_{rm} and T_{0rm} with the χ^2_1 distribution function and the empirical distribution functions of U_{mn} and U_{0mn} with the standard normal distribution function.

Figures 11–12 plot the null empirical distribution functions of T_{rm}, T_{0rm}, U_{mn} and U_{0mn} and their corresponding null limiting distribution functions for the four true success probability scenarios. The simulations show that for $m = 1000$ runs and $n = 1000$ instances, the empirical distribution functions of T_{rm}, T_{0rm}, U_{mn} and U_{0mn} under the null hypotheses are well approximated by their corresponding null limiting distributions under the four true success probability scenarios considered, except for the cases of U_{mn} under the SQA and SSSV success probability scenarios. The approximations for T_{0rm} and U_{0mn} are much better for SQA and SSSV and slightly worse for DW and SA than those for T_{rm} and U_{mn}. The better distribution approximations for T_{0rm} and U_{0mn} in the case SQA and SSSV may be partly due to that the variance stabilization transformation is more stable than normalization especially for hard and easy instances with small or large success probabilities. The good news is that for the four success probability scenarios, there is at least one of test statistics whose null distributions can be well approximated by their limiting null distributions.

4.1.3 *Applications*. As described in Section 3, we have explored the MCMC annealing methods and generated data with various sweeps and different temperatures. This section applies the statistical methodologies developed in Section 4.1.1 to these data. As in Boixo et al. (2014a, 2014b) and Shin et al. (2014), we have selected the MCMC annealing data displayed in Figures 4 and 5 to match the overall patterns of the DW data. Below we present statistical analysis regarding fitting MCMC based annealing models to the DW data.

We applied the tests based on T_{rm}, T_{0rm}, U_{mn} and U_{0mn} in Section 4.1.1 to ground state success probability data displayed in Figures 4–5 for testing whether the DW device yields the same success probability as any of SA, SQA and SSSV. For the multiple tests based on T_{rm} and T_{0rm}, we computed the p-values corresponding to the 1000 instances for each pair of the annealing methods. Figure 13(a)–(c) plots the 1000 p-values of the T_{0rm} test against the DW ground state success frequencies for testing DW against one of SA, SQA and SSSV, with Figure 13(d) for the case of SQA against SSSV. The plots show many p-values are very close to zero. Using the FDR procedure (Benjamini and Hochberg, 1995 and Benjamini, 2010), we found that the q-values are essentially zero for all six pairs of the four annealing methods. The results indicate that there is statistical evidence to suggest that the DW data are not compatible with data generated from any of the MCMC based annealing procedures. In fact, the ground state success probability data provide very strong statistical evidence to indicate significant pairwise differences in the instance distributions among the four annealing data displayed in Figure 4.

For the goodness-of-fit tests based on U_{mn} and U_{0mn}, again we found that the p-values are essentially zero for the six pairs of the four annealing methods, and the testing results provide statistical evidence to indicate significant pairwise differences in the instance distributions among the four annealing data displayed in Figure 4. In particular, they suggest that neither SQA nor SSSV data in Figure 4 have the same distribution as the DW data. The conclusions are not in agreement with the implications of the findings in Boixo et al. (2014a, 2014b) and Shin et al. (2014) based on the correlation and histogram pattern results. As the high correlations of DW and SQA or SSSV suggest only statistical association among the data generated from the annealing models, they do not imply they yield the same success probability. In fact, subsequent investigations were carried out to distinguish SQA and SSSV from the DW device through much more complicated physical experimental studies on isolated and clustered ground states

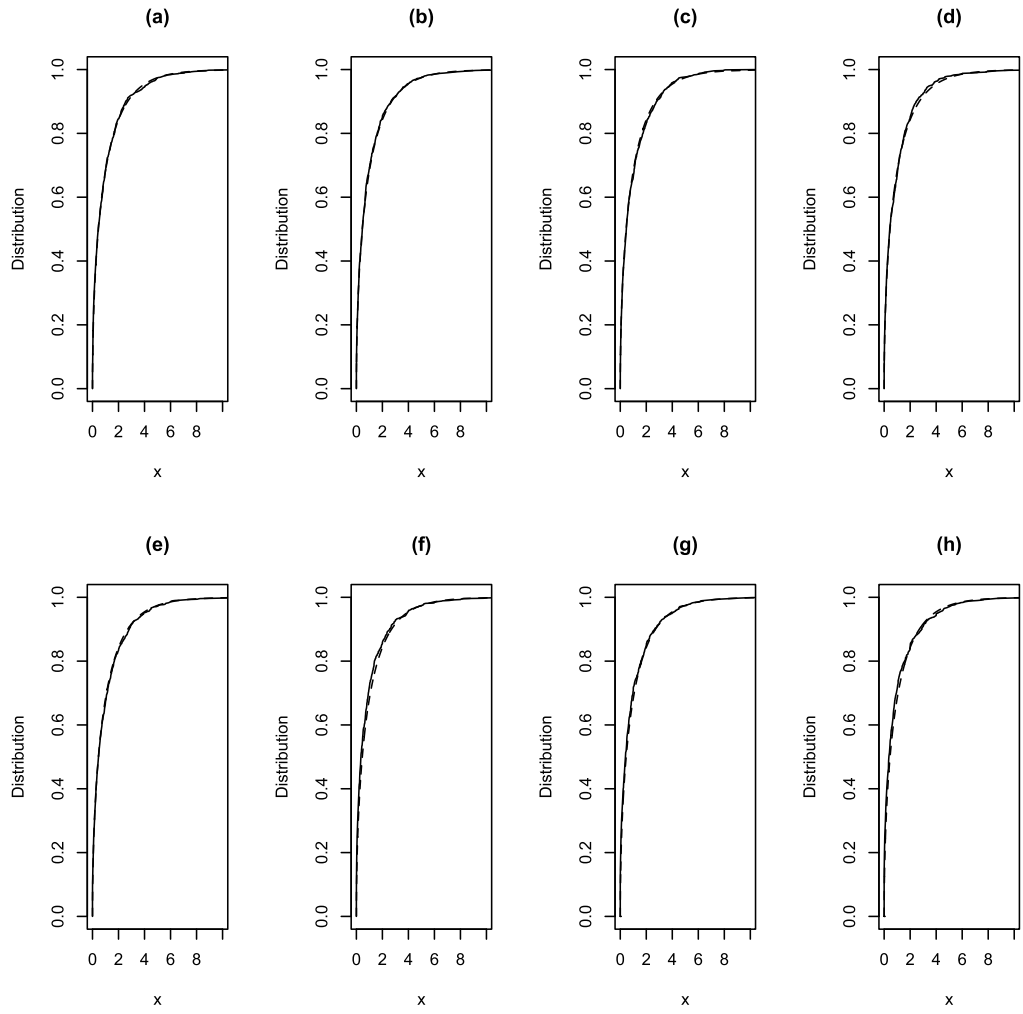

FIG. 11. *Plots of typical null empirical distribution functions* (*solid line*) *and the null limiting chi-square distribution function* (*dash line*) *for the multiple test statistics* T_{rm} *and* T_{0rm} *under four true success probability scenarios. The plot pairs* (a, b), (c, d), (e, f) *and* (g, h) *correspond to* (T_{rm}, T_{0rm}) *with true success probabilities equal to the success frequencies of DW, SA, SQA and SSSV, respectively.*

(Vinci et al., 2014) and on ground state degeneracy and excited states (Albash et al., 2014). However, our analysis shows that through proper statistical analysis of data on the probability of successfully finding ground states we can distinguish SQA and SSSV from the D-Wave device, rather than supporting the agreement between DW and SQA or SSSV. In fact, our test results provide statistical evidence to indicate that SSSV and SQA do not yield the same ground state success probability over all instances, in spite of extremely high correlation between the SQA and SSSV data.

4.2 Shape Pattern Analysis for Ground State Success Probability

4.2.1 *Statistical tests.* From the histogram plots in Figure 4 for the ground state success probability data, we have observed some shape patterns. For example, there are U-shape histogram patterns in the DW, SQA and SSSV data, and unimodal (including monotone) histogram patterns in the SA data. Since these histograms may be viewed as some "density estimators" of distributions of the underlying ground state success probabilities over instances, a unimodal distribution implies that the corresponding distribution function has a convex and then concave shape, while the U-shape pattern is a upside down flip over of the unimodal pattern with a concave and then convex distribution function. We applied the dip test (Hartigan and Hartigan, 1985) to the ground state success probability data for checking the distributional shape patterns. Consider the null hypothesis that the underlying distribution of the success probability data is unimodal

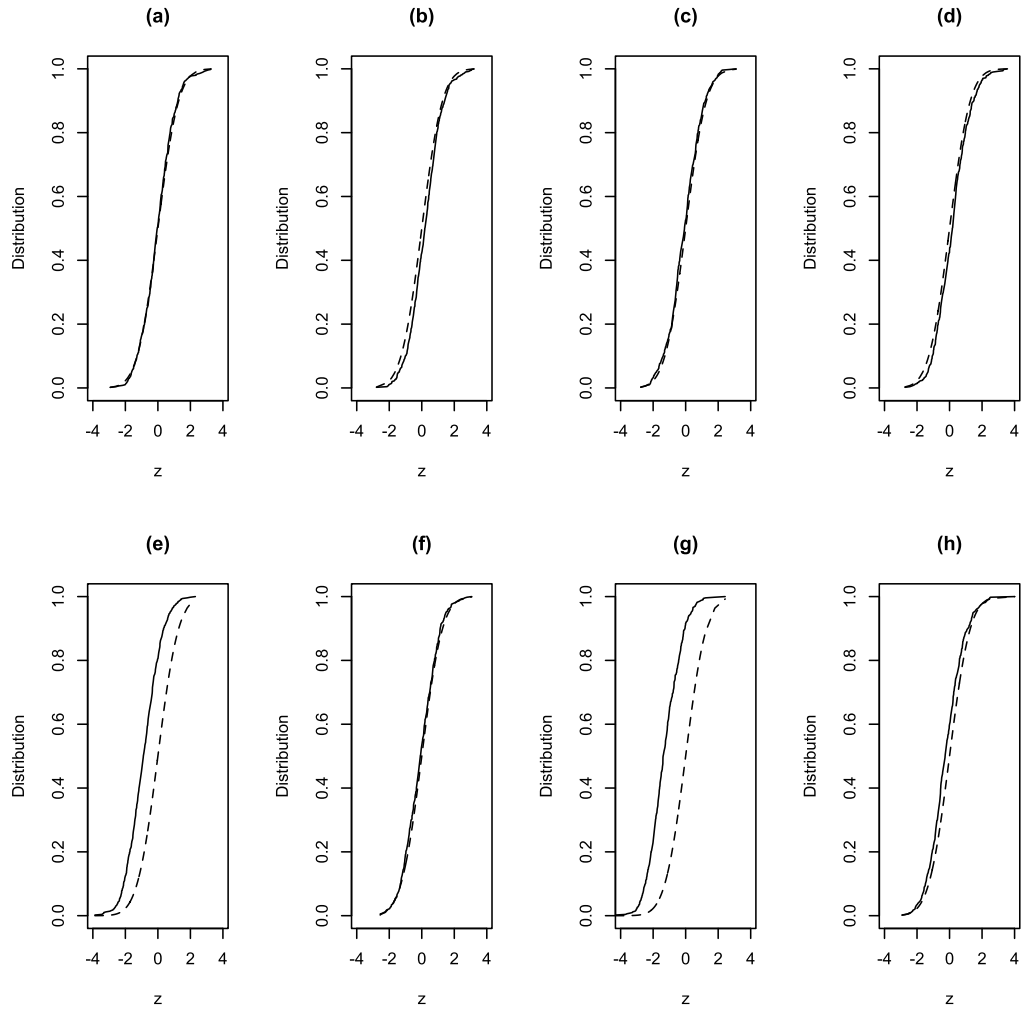

FIG. 12. *Plots of the null empirical distribution functions* (*solid line*) *and the null limiting standard normal distribution function* (*dash line*) *for the goodness-of-fit test statistics* U_{mn} *and* U_{0mn} *under four true success probability scenarios. The plot pairs* (a, b), (c, d), (e, f) *and* (g, h) *correspond to* (U_{mn}, U_{0mn}) *with true success probabilities equal to the success frequencies of DW, SA, SQA and SSSV, respectively.*

against alternative that the underlying distribution is not unimodal. For a given success probability data set of sample size n, denote by F_n its empirical distribution function. The dip test statistic $D(F_n)$ for testing unimodality is defined to be the maximum difference between the empirical distribution function of the data and the unimodal distribution function that minimizes the maximum difference over all unimodal distribution functions. Order restricted statistical inferences show that $D(F_n)$ is equal to the maximum distance between $F_n(p)$ and the unimodal distribution function estimator $\widehat{F}_n(p)$, which turns out to be the largest convex minorant of $F_n(p)$ on $[0, \widehat{p}]$ and then least concave majorant of $F_n(p)$ on $[\widehat{p}, 1]$, where the turning point $\widehat{p}$ is selected to minimize the maximum distance $\sup_{0\le p\le 1} |F_n(p) - \widehat{F}_n(p)|$, the largest convex minorant of F_n on an interval is the largest convex function bounded from above by F_n on the interval, and the least concave majorant of F_n on an interval is the smallest concave function bounded from below by F_n on the interval (Robertson, Wright and Dykstra, 1988 and Wang, 1995).

When the underlying null distribution is the uniform distribution on [0, 1], which is the asymptotic least favorable unimodal distribution, $\sqrt{n}D(F_n)$ converges in distribution to $D(W)$ as $n \to \infty$, where W is a standard Brownian bridge on [0, 1], and $D(W)$ is the maximum difference between W and the convex and then concave function on [0, 1] that minimizes the maximum difference over all such convex-concave functions on [0, 1] (Hartigan and Hartigan, 1985). For testing unimodality for each of the DW, SA, SQA and SSSV data displayed in Figures 4 and 5, we computed its dip test statistic

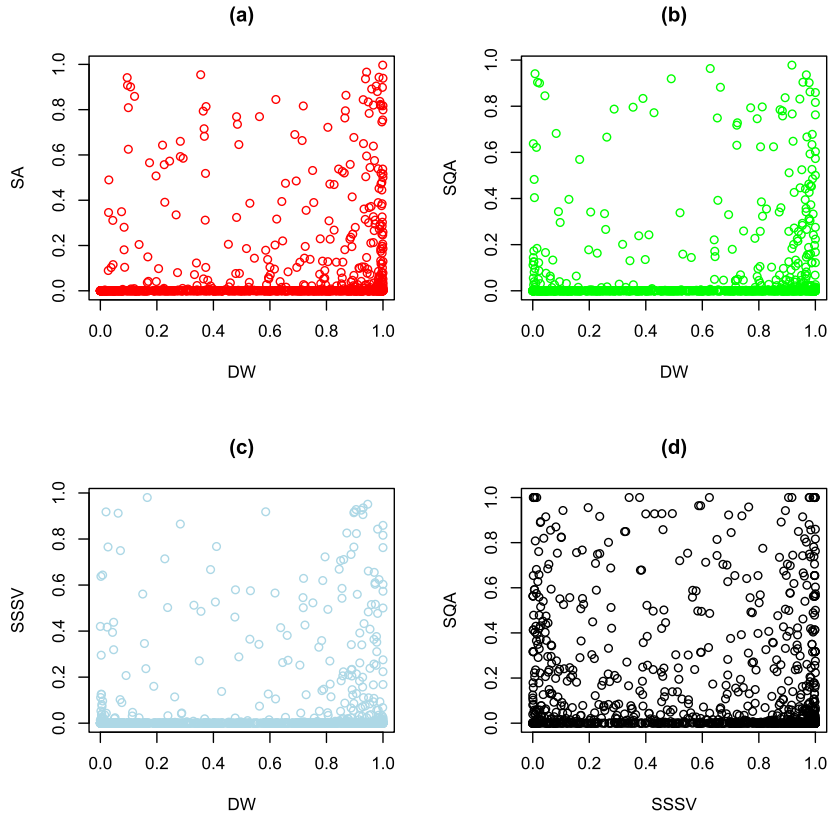

FIG. 13. *The plots of p-values against DW success probability for testing DW vs SA, SQA and SSSV in* (a)–(c), *respectively, along with the plot of p-values against SSSV success probability for testing SSSV vs. SQA in* (d).

$D(F_n)$. We numerically evaluated the limiting distribution of $D(W)$ and found the p-values to be 0.006 for DW, 0.998 for SA and 0.000 for both SQA and SSSV. The test results provide statistical evidence to confirm the unimodal histogram pattern exhibited only in the SA data.

Now we consider testing for the U-shape pattern, that is, the null hypothesis is that the underlying distribution of the success probability data has a U-shape pattern against the alternative that the underlying distribution does not exhibit a U-shape pattern. Although the dip test was designed for testing unimodality, since the U-shape pattern is a upside-down flip-over of the unimodal pattern, we define the dip test statistic D_0 to be the maximum difference between the empirical distribution function and the U-shape distribution function that minimizes the maximum difference over all U-shape distribution functions. The results for the U-shape case are essentially the same as in the unimodal case except for switching between convexity and concavity and between minorant and majorant.

THEOREM 3. *Given the empirical distribution function F_n of a success probability data set, the dip test statistic $D_0(F_n)$ is equal to the maximum distance between the empirical distribution function F_n and the U-shape distribution function estimator $\tilde{F}_n$, which is equal to the least concave majorant and then the largest convex minorant of F_n with the turning point chosen to minimize the maximum distance between F_n and $\tilde{F}_n$; when the underlying distribution is the uniform distribution on* $[0, 1]$, *then as* $n \to \infty$, $\sqrt{n}D_0(F_n)$ *converges in distribution to the same limiting distribution as in the unimodal case.*

For each of the DW, SQA and SSSV data displayed in Figures 4 and 5, we evaluated the dip test statistics $D_0(F_n)$ and computed the p-values to be 0.000 for DW, 0.996 for SQA and 0.322 for SSSV. The test results indicate strong statistical evidence to support the U-shape pattern for SQA and SSSV but not for DW. We further explored the same ground state success probability data displayed in Figure 4 and checked their histograms by varying the numbers of bins. Figure 14 features histogram plots for the data with a larger number of bins than that used in Figure 4. The DW histogram plotted in Figure 14 seems to reveal a possible mode near 0.92 and indicate that the DW distribution may

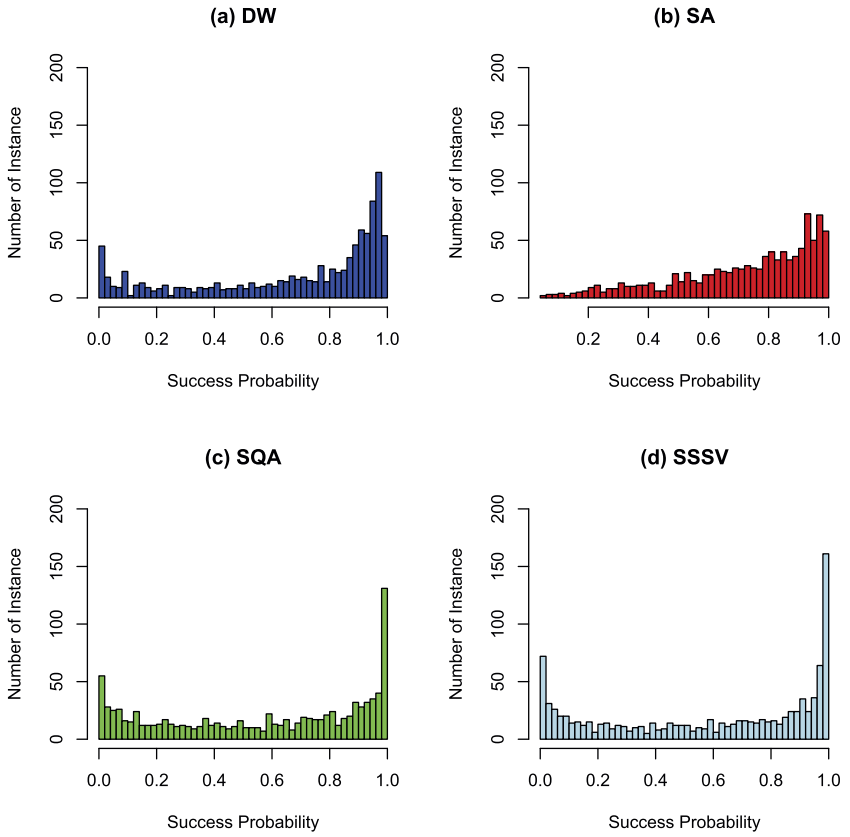

FIG. 14. *Histogram plots of the same ground state success probability data as in Figure 4 for a larger number of bins than that used in Figure 4. Again plots* (a)–(d) *are for DW, SA, SQA and SSSV, respectively.*

be bimodal instead of a U-shape pattern. On the other hand, histogram plots for the SQA and SSSV data with different numbers of bins all point to the U-shape pattern.

4.2.2 *Exploratory shape pattern analysis by regression*. Quantum mechanics indicates that for the Ising model, the ground state success probability may relate to the ground states, the first excited states and the energy gap between the two types of states, where with the ground states corresponding to the smallest energy value, the first excited states correspond to the second smallest energy value. For the MCMC annealing methods, for each instance we repeatedly ran $m = 1000$ times to compute ground state success frequency. Of the m runs, some reached the ground states, and others did not. For those runs that did not reach any ground states, we used their minimal energy and the associated configurations as the proxies for the second smallest energy and the first excited states, respectively. Using the proxies, we defined the energy gap as the difference between the proxy energy of the first excited states and the energy of the ground states and the Hamming distance between the two types of states as the minimum of the Hamming distances between all pairs of configurations with one being a ground state and one being a proxy of the first excited states, where the Hamming distance between two configurations is equal to the number of different spins between their configurations.

Figure 15 displays scatter plots of ground state success frequency versus energy gap or Hamming distance for SA, SQA and SSSV. The scatter plots indicate that while SA, SQA and SSSV all have similar positive associations between ground state success probability and energy gap, they exhibit different patterns in the scatter plots of success probability versus Hamming distance. For SQA and SSSV, as Hamming distance increases, the ground state success probability data tend to cluster around zero and one, but the SA scatter plot does not exhibit such a pattern. To further explore the bi-cluster tendency of ground state success probability data associated with Hamming distance, we divided the 1000 ground state success probability data into two subsets according to whether the Hamming distances

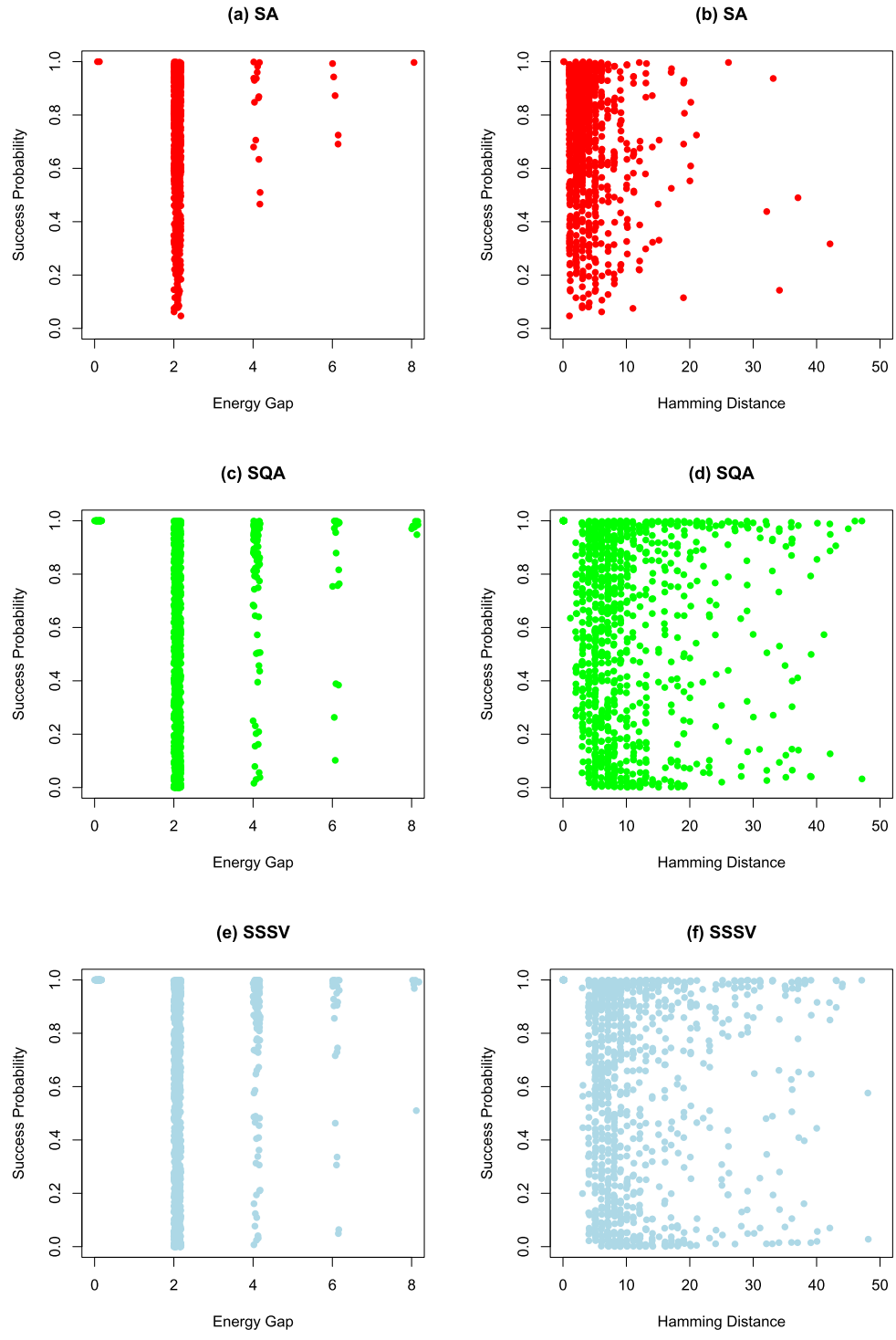

FIG. 15. *Plots of ground state success probability data against energy gap and Hamming distance for SA, SQA and SSSV.*

for the corresponding instances exceed 5 or not, and illustrated in Figure 16 the histogram plots for the subset data. The figure shows that while SA histograms display a consistent monotone tendency regardless of the Hamming distance, both SQA and SSSV exhibit a strong U-shape pattern for the subset data with Hamming distance at least 5 and a clear monotone pattern for the subset data with Hamming distance less than 5. This reveals that the U-shape pattern in the ground state success probability data of SQA and SSSV may be

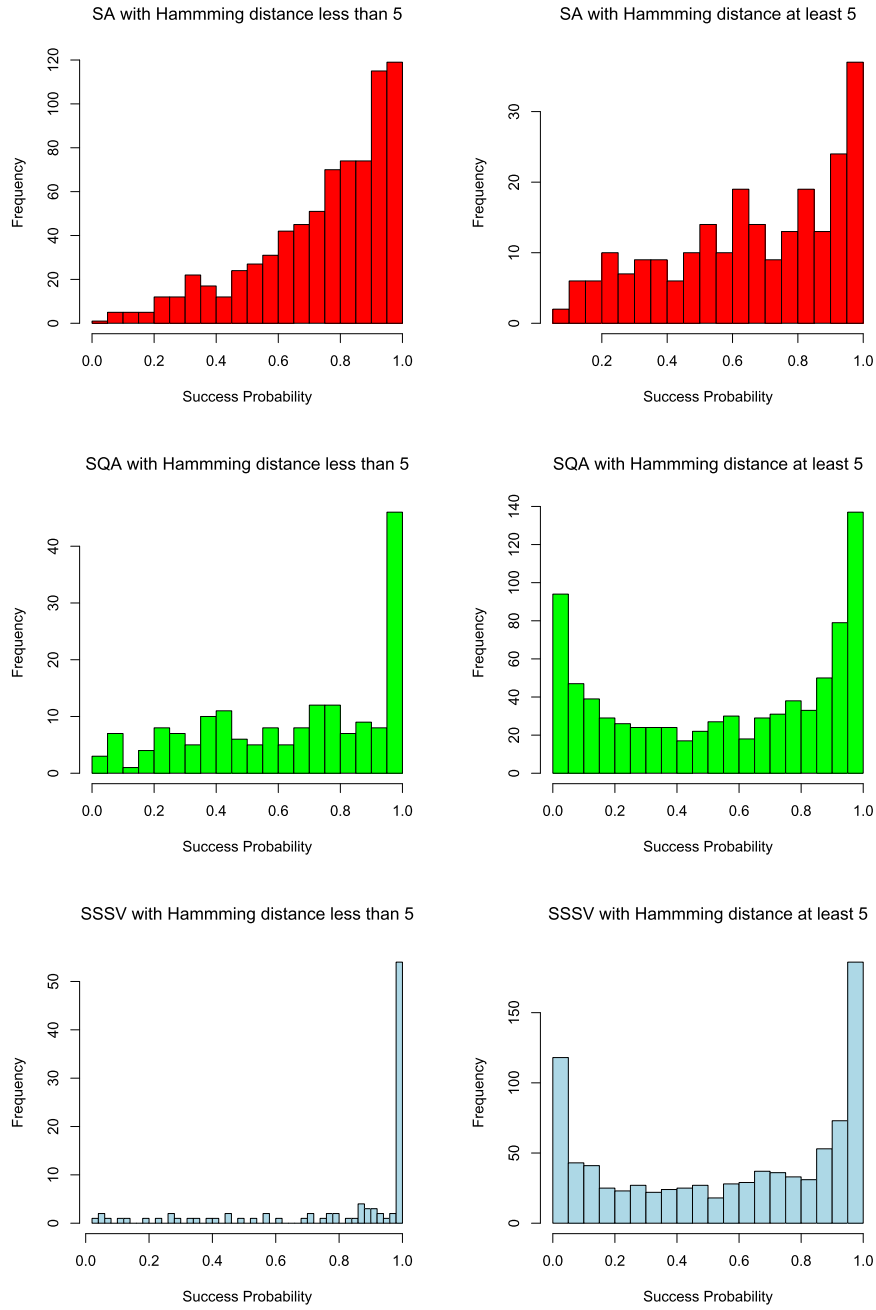

FIG. 16. *Histograms of ground state success probability data of SA, SQA and SSSV classified according to whether Hamming distance exceeds* 5 *or not.*

largely due to the instances with large Hamming distances. The findings point to the direction for further exploring the bimodal phenomenon displayed in the ground state success probability data from D-Wave devices as well as MCMC based annealing methods.

We may fit the data to a logistic regression model. For method k and instance r, we obtained ground state success frequency $\widehat{p}_{k,m,r}$, energy gap $\Delta_{k,m,r}$ and the Hamming distance $D_{k,m,r}$. Since $m\widehat{p}_{k,m,r}$ are independent and follow $\mathrm{Bin}(m, p_{k,\infty,r})$, treating $(\Delta_{k,m,r}, D_{k,m,r})$ as covariates and $\widehat{p}_{k,m,r}$ as a dependent variable, we fit the data to a logistic regression model with link function

$$\log\left(\frac{p_{k,\infty,r}}{1-p_{k,\infty,r}}\right) = \beta_0 + \beta_1\Delta_{k,m,r} + \beta_2 D_{k,m,r},$$

where $(\beta_0, \beta_1, \beta_2)$ are model parameters. We estimated the model parameters $(\beta_0, \beta_1, \beta_2)$ by the standard max-

TABLE 1
Parameter estimates with standard errors for the logistic regression model

	$\widehat{\beta}_0$	$se_{\widehat{\beta}_0}$	$\widehat{\beta}_1$	$se_{\widehat{\beta}_1}$	$\widehat{\beta}_2$	$se_{\widehat{\beta}_2}$
SA	0.340	0.0135	0.410	0.0069	−0.056	0.0006
SQA	−0.070	0.0053	0.209	0.0024	−0.006	0.0003
SSSV	0.161	0.0048	0.169	0.0021	−0.012	0.0002

imum likelihood estimation for the generalized linear model and reported the results in Table 1.

For each annealing method, we conducted statistical tests to check whether covariates $\Delta_{k,m,r}$ and $D_{k,m,r}$ have any significant effects on the ground state success probability $p_{k,\infty,r}$ by considering hypotheses that either one of β_1 and β_2 or both are equal to zero. For SA, SQA and SSSV, all the tests resulted in essentially zero p-values, and both energy gap and Hamming distance are highly significant. Again the obtained distinct models for SA, SQA and SSSV point out that they behave differently.

5. CONCLUDING REMARKS

This paper reviews quantum annealing and its implementations by D-Wave devices and by MCMC based annealing methods. We describe computing experiments by D-Wave devices and carry out simulation experiments by MCMC based annealing algorithms on classical computers to generate ground state success probability data and compare quantum annealing with classical annealing. Using the experiment data, we illustrate the consistence or inconsistence claims in the quantum literature regarding the studies of the D-Wave devices along with MCMC based annealing methods. We propose statistical methodologies to analyze computing experimental data from a D-Wave device and simulated data from SA, SQA, and SSSV annealing. We develop asymptotic theory for the proposed methodologies and conduct simulation studies to check their finite sample performances. Our hypothesis test results show bimodal histogram patterns displayed in input–output data from the D-Wave device, U-shape histogram patterns shown in the data from SQA and SSSV methods and unimodal (including monotone) histogram patterns exhibited in the data from SA method. Our statistical analysis discovers that the U-shape histogram patterns may be possibly due to the large Hamming distance between the ground states and the excited states for some instances and that the sampling tends to be around ground states and high energy states. Our test results supply statistical evidence to imply that none of SA, SQA and SSSV models fit to input–output data from the D-Wave device.

Quantum computing is the development of computer technology based on the principles of quantum physics, and this highly interdisciplinary field is of great current interest in frontier sciences. This is an exciting time, and we stand at the edge of quantum era where many quantum technologies are being developed, and quantum devices are being built with capabilities exceeding classical computer based devices. Statistics and quantum computing are a two-way street. On the one hand, quantum science may yield quantum computing devices to revolutionize statistical computing and inferences, and on the other hand, quantum theory is of stochastic nature, and quantum procedures and quantum devices are based on random phenomena, therefore, frontier research and technological developments of quantum computing encounter complex data and face many statistical challenges.

This paper may serve as a stepping stone to the interface between quantum computing and statistics. It opens new research directions by raising many statistical issues and lots of open problems for the future study. We may study statistical properties of quantum annealing and investigate the probability lower bound for successfully solving combinatorial optimization problems by quantum annealing. We may study MCMC based annealing methods such as the statistical properties of SQA and SSSV approaches. For example, in comparison with SA, SQA is very slow, and SSSV requires an extremely large number of sweeps to get bimodal histogram patterns for the ground state success probability data. We need to investigate their convergence properties and study the relationships between sweeps and patterns for these MCMC based annealing methods. We may develop statistical methods to analyze computing experimental data from D-Wave devices and simulated data from SA, SQA and SSSV annealing. We need to study bimodal histogram patterns displayed in input–output data from the D-Wave devices, U-shape histogram patterns shown in the data from SQA and SSSV methods, and unimodal (including monotone) histogram patterns exhibited in the data from SA and uncover the sources for these patterns. It is very interesting to find out what cause the bimodal (or U-shape) vs. unimodal (or monotone) patterns in ground state success probability data for these methods. Moreover, quantum

annealing and D-Wave devices solve combinatorial optimization problems by performing some sort of Monte Carlo sampling, but we know little about their sampling properties and their usefulness for certain computational tasks in particular for machine learning and other statistical computing problems. It is challenging to study random errors in the qubits of D-Wave devices and their computational impact on the ground state success probability. It is critical to investigate quantum states of D-Wave devices by quantum tomography (Cai et al., 2016). It is extremely important to find appropriate models for quantum devices such as D-Wave devices and use experimental data to test whether they work according to the claims and/or specifications.

APPENDIX: PROOFS

Denote by C's generic constants whose values are free of (m, n, M, d) and may change from appearance to appearance.

A.1 Proof of Theorem 1

Because of similarity, we give the proof argument only for T_{rm}. Define

$$\check{T}_{rm} = \frac{m(\widehat{p}_{k,m,r} - \widehat{p}_{\ell,m,r})^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})}.$$

For each r, as $m \to \infty$, $\check{T}_{rm} \to \chi_1^2$, where χ_1^2 is a chi-square distribution with one degree of freedom. Since for $x \geq 0$,

$$\begin{aligned} &P(\check{T}_{rm} \leq x) \\ &\quad = P\Bigg(-x^{1/2} \\ &\qquad \leq \frac{m^{1/2}(\widehat{p}_{k,m,r} - \widehat{p}_{\ell,m,r})}{[p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})]^{1/2}} \\ &\qquad \leq x^{1/2} \Bigg), \end{aligned}$$

applying the Berry–Esseen theorem we obtain

$$\sup_{-\infty<x<\infty} \left| P(\check{T}_{rm} \leq x) - P\left(\chi_1^2 \leq x\right) \right| \leq Cm^{-1/2}.$$

As the right-hand side of above inequality is free of r, we have

$$\begin{aligned} &\sup_{1\leq r\leq n} \sup_{-\infty<x<\infty} \left| P(\check{T}_{rm} \leq x) - P\left(\chi_1^2 \leq x\right) \right| \\ &\qquad \leq Cm^{-1/2}. \end{aligned} \tag{A.1}$$

The Bernstein inequality shows that there exists a constant C free of (r, n, m) such that for any $x \geq 0$ and $k = 0, 1, 2, 3$,

$$P\left(|\widehat{p}_{k,m,r} - p_{k,\infty,r}| \geq x\right) \leq e^{-Cmx^2}. \tag{A.2}$$

Let

$$w_{rm} = \frac{\widehat{p}_{k,m,r}(1-\widehat{p}_{k,m,r}) + \widehat{p}_{\ell,m,r}(1-\widehat{p}_{\ell,m,r})}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{l,\infty,r})}.$$

Then

$$\begin{aligned} &P\left(|w_{rm} - 1| \geq a\right) \\ &\quad \leq P\left(\left|(\widehat{p}_{k,m,r} - p_{k,m,r})(1 + \widehat{p}_{k,m,r} + p_{k,\infty,r})\right| \right. \\ &\qquad \left. \geq p_{k,\infty,r}(1-p_{k,\infty,r})a\right) \\ &\qquad + P\left(\left|(\widehat{p}_{\ell,m,r} - p_{\ell,m,r})(1 + \widehat{p}_{\ell,m,r} + p_{\ell,\infty,r})\right| \right. \\ &\quad \geq p_{k,\infty,r}(1-p_{k,\infty,r})a\Big) \\ &\quad \leq P\left(|\widehat{p}_{k,m,r} - p_{k,m,r}| \geq p_{k,\infty,r}(1-p_{k,\infty,r})a/3\right) \\ &\qquad + P\left(|\widehat{p}_{\ell,m,r} - p_{\ell,m,r}| \right. \\ &\qquad \left. \geq p_{k,\infty,r}(1-p_{k,\infty,r})a/3\right) \\ &\quad \leq e^{-C_0 m a^2}, \end{aligned} \tag{A.3}$$

where the last inequality is due to (A.2). Take $a = \sqrt{2/C_0}m^{-1/2}\log^{1/2} m$, where C_0 is taken to be the same constant as in (A.3) so that $e^{-C_0 m a^2} = m^{-2}$. Then for $1 \leq r \leq n$ and any $x \geq 0$,

$$\begin{aligned} &P(T_{rm} \leq x) \\ &\quad \leq P\left(T_{rm} \leq x, |w_{rm} - 1| < a\right) + P\left(|w_{rm} - 1| \geq a\right) \\ &\quad \leq P\left(\check{T}_{rm} \leq (1+a)x\right) + P\left(|w_{rm} - 1| \geq a\right), \\ &P(T_{rm} \leq x) \\ &\quad \geq P(T_{rm} \leq x, |w_{rm} - 1| < a) - P(|w_{rm} - 1| \geq a) \\ &\quad \geq P\left(\check{T}_{rm} \leq (1-a)x\right) - P(|w_{rm} - 1| \geq a), \end{aligned}$$

and

$$\begin{aligned} &\left|P(T_{rm} \leq x) - P(\check{T}_{rm} \leq x)\right| \\ &\quad \leq P\left(\check{T}_{rm} \leq (1+a)x\right) - P\left(\check{T}_{rm} \leq (1-a)x\right) \\ &\qquad + 2P\left(|w_{rm} - 1| \geq a\right) \\ &\quad \leq P\left(\chi_1^2 \leq (1+a)x\right) - P\left(\chi_1^2 \leq (1-a)x\right) \\ &\qquad + 2Cm^{-1/2} + 2P\left(|w_{rm} - 1| \geq a\right) \\ &\quad \leq C_1\sqrt{2/C_0}m^{-1/2}\log^{1/2} m + 2Cm^{-1/2} + 2m^{-2}, \end{aligned} \tag{A.4}$$

where the second to last inequality is from (A.1), and the last inequality is due to (A.3), $a = \sqrt{2/C_0}m^{-1/2}$.

$\log^{1/2} m$, and $0 \le P(\chi_1^2 \le (1+a)x) - P(\chi_1^2 \le (1-a)x) \le C_1 a$ for all $x \ge 0$. Finally, the theorem is a consequence of (A.1) and (A.4).

REMARK 1. If $m^{-1/2}\log n \to 0$ as $m, n \to \infty$, then we have a stronger uniform convergence result: $\sup_{1\le r\le n} |T_{rm} - \breve{T}_{rm}|$ converges in probability to zero as $m, n \to \infty$. Indeed, take $x = \sqrt{2/C} m^{-1/2} \log^{1/2} n$ in (A.2) where C is the same constant as in (A.2) so that $e^{-Cmx^2} = n^{-2}$, and

$$P\Big(\sup_{1\le r\le n} |\widehat{p}_{k,m,r} - p_{k,\infty,r}| \ge \sqrt{2/C} m^{-1/2}\log^{1/2} n\Big) \le 1 - \big(1-n^{-2}\big)^n \sim n^{-1},$$

which implies that the difference between the denominators of T_{rm} and $\breve{T}_{rm}$ converges in probability to zero uniformly over $1 \le r \le n$, and hence the uniform convergence result.

A.2 Proof of Theorem 2

Given a method and an instance r, the simulation experiments are independently conducted for m times, thus $\widehat{p}_{k,m,1}, \ldots, \widehat{p}_{k,m,n}$, $k = 0, 1, 2, 3$, are independent, and $m\widehat{p}_{k,m,r} \sim \mathrm{Bin}(m, p_{k,\infty,r})$. Let

$$T = m\sum_{r=1}^{n} \frac{(\widehat{p}_{k,m,r} - \widehat{p}_{\ell,m,r})^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,m,r}(1-p_{\ell,\infty,r})}, \qquad k \ne \ell,\ k, \ell = 0, 1, 2, 3.$$

First, we show that as $m, n \to \infty$, $(T-n)/\sqrt{2n}$ converges in distribution to the standard normal distribution. We establish it by applying central limit theorem to T and checking Liapounov' condition.

$$\begin{aligned} &mE\big\{\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^2\big\} \\ &\quad = mE\big\{(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big\} \\ &\qquad + mE\big\{(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^2\big\} \\ &\quad = p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r}), \end{aligned}$$

$$E\left[\frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})}\right] = 1, \tag{A.5}$$

$$E\left[\left|\frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})}\right|^2\right] = \frac{m^2E\{[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^4\}}{[p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})]^2} = 3 + O\big(m^{-1}\big), \tag{A.6}$$

where the last equality is due to the facts that

$$\begin{aligned} &m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^4\big] \\ &\quad = \frac{1}{m^2} mp_{k,\infty,r}(1-p_{k,\infty,r}) \\ &\qquad \cdot \big[3p_{k,\infty,r}(1-p_{k,\infty,r})(m-2) + 1\big] \\ &\quad = 3\big[p_{k,\infty,r}(1-p_{k,\infty,r})\big]^2 + O\big(m^{-1}\big), \\ &m^2E\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^4\big] \\ &\quad = 3\big[p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big]^2 + O\big(m^{-1}\big), \\ &mE\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big] = p_{k,\infty,r}(1-p_{k,\infty,r}), \\ &mE\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^2\big] = p_{\ell,\infty,r}(1-p_{\ell,\infty,r}), \end{aligned}$$

and

$$\begin{aligned} &m^2E\big\{\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) + (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^4\big\} \\ &\quad = m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^4\big] \\ &\qquad + m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^4\big] \\ &\qquad + 6m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big] \\ &\qquad \cdot E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big] \\ &\quad = 3\big[p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big]^2 \\ &\qquad + O\big(m^{-1}\big). \end{aligned}$$

(A.5) and (A.6) together imply

$$\begin{aligned} &\mathrm{Var}\left[\frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})}\right] \\ &\quad = 2 + O(1/m), \\ &\mathrm{Var}(T) = 2n + O(n/m). \end{aligned} \tag{A.7}$$

On the other hand,

$$\begin{aligned} &E\left[\left|\frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})}\right|^3\right] \\ &\quad \le \big(480\big[p_{k,\infty,r}^3(1-p_{k,\infty,r})^3 \\ &\qquad + p_{\ell,\infty,r}^3(1-p_{\ell,\infty,r})^3\big] + O(1/m)\big) \\ &\qquad /\big(\big[p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big]^3\big) \\ &\quad \le 480 + O(1/m), \end{aligned} \tag{A.8}$$

where the first inequality is due to the facts that

$$\begin{aligned} &m^3E\big\{\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^6\big\} \\ &\quad \le 32m^3E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^6 \\ &\qquad + (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^6\big], \end{aligned}$$

$$
\begin{aligned}
&m^3 E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^6\big] \\
&\quad = 15p_{k,\infty,r}^3(1-p_{k,\infty,r})^3 \\
&\qquad + 5m^{-1}p_{k,\infty,r}^2(1-p_{k,\infty,r})^2(5-26p_{k,\infty,r} \\
&\qquad + 26p_{k,\infty,r}^2) + m^{-2}p_{k,\infty,r}(1-31p_{k,\infty,r} \\
&\qquad + 180p_{k,\infty,r}^2 - 390p_{k,\infty,r}^3 + 360p_{k,\infty,r}^4 \\
&\qquad - 160p_{k,\infty,r}^5) \\
&\quad = 15p_{k,\infty,r}^3(1-p_{k,\infty,r})^3 + O(1/m), \\
&m^3 E\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^6\big] \\
&\quad = 15p_{\ell,\infty,r}^3(1-p_{\ell,\infty,r})^3 + O(1/m).
\end{aligned}
$$

From (A.7) and (A.8), we have

$$
\begin{aligned}
&\big[\mathrm{Var}(T)\big]^{-3/2} \\
&\qquad \times \sum_{r=1}^{n} E\big[\big|\big(m\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^2\big)/\big(p_{k,\infty,r}(1-p_{k,\infty,r}) \\
&\qquad + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big)\big|^3\big] \\
&\le \frac{480n + O(n/m)}{[2n + O(n/m)]^{3/2}} = 70\sqrt{2}n^{-1/2}\big[1 + O(1/m)\big],
\end{aligned}
$$

and Liapounov's condition is satisfied. Hence, $(T - n)/\sqrt{2n}$ converges in distribution to the standard normal distribution. Next, we show that $U_{mn} = (T - n)/\sqrt{2n} + O_P(\sqrt{n}/m)$.

Let $G(u, v) = 1/[u(1-u) + v(1-v)]$, and denote by G_u and G_v the partial derivatives of G with respect to u and v, respectively.

$$
\begin{aligned}
&\frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{\widehat{p}_{k,\infty,r}(1-\widehat{p}_{k,\infty,r}) + \widehat{p}_{\ell,\infty,r}(1-\widehat{p}_{\ell,\infty,r})} \\
&\qquad - \frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})} \\
&\quad = m\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^2 \\
&\qquad \cdot \big[G(\widehat{p}_{k,m,r}, \widehat{p}_{\ell,m,r}) - G(p_{k,\infty,r}, p_{\ell,\infty,r})\big], \\
&G(\widehat{p}_{k,m,r}, \widehat{p}_{\ell,m,r}) - G(p_{k,\infty,r}, p_{\ell,\infty,r}) \\
&\quad = G_u(p_{k,\infty,r}, p_{\ell,\infty,r})(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad + G_v(p_{k,\infty,r}, p_{\ell,\infty,r})(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r}) \\
&\qquad + O_P(m^{-1}).
\end{aligned}
$$

Then

$$
\begin{aligned}
&\sqrt{2n}U_{mn} + n - T \\
&\quad = \sum_{r=1}^{n} \frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{\widehat{p}_{k,\infty,r}(1-\widehat{p}_{k,\infty,r}) + \widehat{p}_{\ell,\infty,r}(1-\widehat{p}_{\ell,\infty,r})} \\
&\qquad - \sum_{r=1}^{n} \frac{m[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})]^2}{p_{k,\infty,r}(1-p_{k,\infty,r}) + p_{\ell,\infty,r}(1-p_{\ell,\infty,r})} \\
&\quad = m\sum_{r=1}^{n} G_u(p_{k,\infty,r}, p_{\ell,\infty,r})\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^2(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad + m\sum_{r=1}^{n} G_v(p_{k,\infty,r}, p_{\ell,\infty,r})\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^2(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big] \\
&\qquad + O_P(nm^{-1}) \\
&\quad = A_1 + A_2 + O_P(nm^{-1}).
\end{aligned} \tag{A.9}
$$

Direct calculations show

$$
\begin{aligned}
&E\big\{m\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^2 \\
&\qquad \cdot (\widehat{p}_{k,m,r} - p_{k,\infty,r})\big\} \\
&\quad = mE\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^3\big] \\
&\qquad + mE\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^2(\widehat{p}_{k,m,r} - p_{k,\infty,r})\big] \\
&\qquad - 2mE\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big] \\
&\quad = mE\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^3\big] \\
&\quad = m^{-1}p_{k,\infty,r}(1-p_{k,\infty,r})(1-2p_{k,\infty,r}),
\end{aligned}
$$

and $E(A_1) = O(n/m)$. Also

$$
\begin{aligned}
&E\big\{m^2\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^4 \\
&\qquad \cdot (\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big\} \\
&\quad = m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^6\big] \\
&\qquad - 4m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^5\big] \\
&\qquad \cdot E\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big] \\
&\qquad + 6m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^4\big] \\
&\qquad \cdot E\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^2\big] \\
&\qquad - 4m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^3\big] \\
&\qquad \cdot E\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^3\big] \\
&\qquad + m^2E\big[(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big]
\end{aligned}
$$

$$\cdot E\big[(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^4\big]$$
$$= O\big(m^{-1}\big) + 0 + O\big(m^{-1}\big) + O\big(m^{-2}\big) + O\big(m^{-1}\big)$$
$$= O\big(m^{-1}\big),$$

and

$$\begin{aligned}
&\operatorname{Var}(A_1/\sqrt{n}) \\
&\quad = \frac{1}{n}\operatorname{Var}(A_1) \\
&\quad = \frac{1}{n}\sum_{r=1}^{n} G_u^2(p_{k,\infty,r}, p_{\ell,\infty,r}) \operatorname{Var}\big\{m\big[(\widehat{p}_{k,m,r} \\
&\qquad - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^2 \\
&\qquad \cdot (\widehat{p}_{k,m,r} - p_{k,\infty,r})\big\} \\
&\quad \le \frac{1}{n}\sum_{r=1}^{n} G_u^2(p_{k,\infty,r}, p_{\ell,\infty,r}) E\big\{m^2\big[(\widehat{p}_{k,m,r} \\
&\qquad - p_{k,\infty,r}) - (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})\big]^4 \\
&\qquad \cdot (\widehat{p}_{k,m,r} - p_{k,\infty,r})^2\big\} \\
&\quad = O(1/m).
\end{aligned}$$

Applying Chebyshev's inequality, we obtain

$$\begin{aligned}
&P\big(\big|A_1/\sqrt{n} - E(A_1)/\sqrt{n}\big| > am^{-1/2}\big) \\
&\quad \le \frac{\operatorname{Var}(A_1/\sqrt{n})}{m^{-1}a^2} = \frac{m\operatorname{Var}(A_1/\sqrt{n})}{a^2},
\end{aligned}$$

which can be arbitrarily small as $a \to \infty$. Hence,

$$\begin{aligned}
A_1/\sqrt{n} &= E(A_1)/\sqrt{n} + O_P\big(m^{-1/2}\big) \\
&= O\big(n^{1/2}m^{-1}\big) + O_P\big(m^{-1/2}\big).
\end{aligned}$$

Similarly, we can show

$$\begin{aligned}
A_2/\sqrt{n} &= E(A_2)/\sqrt{n} + O_P\big(m^{-1/2}\big) \\
&= O\big(n^{1/2}m^{-1}\big) + O_P\big(m^{-1/2}\big).
\end{aligned}$$

With the orders of A_1 and A_2 from (A.9), we arrive at

$$\begin{aligned}
&U_{mn} - (T-n)/\sqrt{2n} \\
&\quad = A_1/\sqrt{2n} + A_2/\sqrt{2n} + O_P\big(n^{1/2}m^{-1}\big) \\
&\quad = O_P\big(n^{1/2}m^{-1} + m^{-1/2}\big),
\end{aligned}$$

which together with $n^{1/2}m^{-1} \to 0$ implies that U_{mn} has the same limiting distribution as $(T-n)/\sqrt{2n}$.

Next, we will establish the limiting distribution for U_{0mn}. Again we need to show

$$U_{0mn} = \frac{T-n}{\sqrt{2n}} + O_P\big(n^{1/2}m^{-1} + m^{-1/2}\big). \tag{A.10}$$

We have

$$\begin{aligned}
&\arcsin\big(\sqrt{\widehat{p}_{k,m,r}}\big) - \arcsin\big(\sqrt{p_{k,\infty,r}}\big) \\
&\quad = \big[4p_{k,\infty,r}(1-p_{k,\infty,r})\big]^{-1/2}(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad - 2^{-3}(2p_{k,\infty,r}-1)\big[p_{k,\infty,r}(1-p_{k,\infty,r})\big]^{-3/2} \\
&\qquad \cdot (\widehat{p}_{k,m,r} - p_{k,\infty,r})^2 \\
&\qquad + O_P\big(m^{-3/2}\big), \\
&2m\big(\arcsin\big(\sqrt{\widehat{p}_{k,m,r}}\big) - \arcsin\big(\sqrt{p_{k,\infty,r}}\big)\big)^2 \\
&\quad = \frac{m(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2}{2p_{k,\infty,r}(1-p_{k,\infty,r})} \\
&\qquad - 2^{-3}(2p_{k,\infty,r}-1)\big[p_{k,\infty,r}(1-p_{k,\infty,r})\big]^{-2} \\
&\qquad \cdot m(\widehat{p}_{k,m,r} - p_{k,\infty,r})^3 + O_P\big(m^{-1}\big) \\
&\quad = \frac{m(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2}{2p_{k,\infty,r}(1-p_{k,\infty,r})} + B_{1r} + O_P\big(m^{-1}\big).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&2m\big(\arcsin\big(\sqrt{\widehat{p}_{\ell,m,r}}\big) - \arcsin\big(\sqrt{p_{\ell,\infty,r}}\big)\big)^2 \\
&\quad = \frac{m(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^2}{2p_{\ell,\infty,r}(1-p_{\ell,\infty,r})} \\
&\qquad - 2^{-3}(2p_{\ell,\infty,r}-1)\big[p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big]^{-2} \\
&\qquad \cdot m(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^3 + O_P\big(m^{-1}\big) \\
&\quad = \frac{m(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^2}{2p_{\ell,\infty,r}(1-p_{\ell,\infty,r})} + B_{2r} + O_P\big(m^{-1}\big),
\end{aligned}$$

and

$$\begin{aligned}
&4m\big(\arcsin\big(\sqrt{\widehat{p}_{k,m,r}}\big) - \arcsin\big(\sqrt{p_{k,\infty,r}}\big)\big) \\
&\qquad \cdot \big(\arcsin\big(\sqrt{\widehat{p}_{\ell,m,r}}\big) - \arcsin\big(\sqrt{p_{\ell,\infty,r}}\big)\big) \\
&\quad = 4\big[4p_{k,\infty,r}(1-p_{k,\infty,r})\big]^{-1/2} \\
&\qquad \cdot \big[4p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big]^{-1/2} m(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad \cdot (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r}) \\
&\qquad - 4\big[4p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big]^{-1/2} \\
&\qquad \cdot 2^{-3}(2p_{k,\infty,r}-1)\big[p_{k,\infty,r}(1-p_{k,\infty,r})\big]^{-3/2} \\
&\qquad \cdot m(\widehat{p}_{k,m,r} - p_{k,\infty,r})^2(\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r}) \\
&\qquad - 4\big[4p_{k,\infty,r}(1-p_{k,\infty,r})\big]^{-1/2} 2^{-3}(2p_{\ell,\infty,r}-1) \\
&\qquad \cdot \big[p_{\ell,\infty,r}(1-p_{\ell,\infty,r})\big]^{-3/2} m(\widehat{p}_{k,m,r} - p_{k,\infty,r}) \\
&\qquad \cdot (\widehat{p}_{\ell,m,r} - p_{\ell,\infty,r})^2
\end{aligned}$$

$$
\begin{aligned}
&+ O_P(m^{-1}) \\
&= [p_{k,\infty,r}(1-p_{k,\infty,r})p_{\ell,\infty,r}(1-p_{\ell,\infty,r})]^{-1/2} \\
&\quad \cdot m(\widehat{p}_{k,m,r}-p_{k,\infty,r})(\widehat{p}_{\ell,m,r}-p_{\ell,\infty,r}) \\
&\quad - B_{3r} - B_{4r} + O_P(m^{-1}).
\end{aligned}
$$

Then

$$
\begin{aligned}
&2m\sum_{r=1}^{n}[(\arcsin(\sqrt{\widehat{p}_{k,m,r}}) \\
&\quad - \arcsin(\sqrt{p_{k,\infty,r}})) - (\arcsin(\sqrt{\widehat{p}_{\ell,m,r}}) \\
&\quad - \arcsin(\sqrt{p_{\ell,\infty,r}}))]^2 \\
&\quad \cdot m\sum_{r=1}^{n}\Bigg[\frac{(\widehat{p}_{k,m,r}-p_{k,\infty,r})^2}{2p_{k,\infty,r}(1-p_{k,\infty,r})} \\
&\quad + \frac{(\widehat{p}_{\ell,m,r}-p_{\ell,\infty,r})^2}{2p_{\ell,\infty,r}(1-p_{\ell,\infty,r})} \\
&\quad - \frac{(\widehat{p}_{k,m,r}-p_{k,\infty,r})(\widehat{p}_{\ell,m,r}-p_{\ell,\infty,r})}{[p_{k,\infty,r}(1-p_{k,\infty,r})p_{\ell,\infty,r}(1-p_{\ell,\infty,r})]^{1/2}}\Bigg] \\
&\quad + \sum_{r=1}^{n}(B_{1r}+B_{2r}-B_{3r}-B_{4r}) + O_P(nm^{-1}) \\
&= T + \sum_{r=1}^{n}(B_{1r}+B_{2r}-B_{3r}-B_{4r}) + O_P(nm^{-1}),
\end{aligned}
$$

$$
U_{0rm} = \frac{T-n}{\sqrt{2n}}
$$

$$
\begin{aligned}
(\text{A.11}) \quad &+ (2n)^{-1/2}\sum_{r=1}^{n}(B_{1r}+B_{2r}-B_{3r}-B_{4r}) \\
&+ O_P(n^{1/2}m^{-1}).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&E[m(\widehat{p}_{k,m,r}-p_{k,\infty,r})^3] \sim p(1-p)(1-2p)m^{-1}, \\
&\operatorname{Var}[m(\widehat{p}_{k,m,r}-p_{k,\infty,r})^3] \\
&\quad \sim \{15[p(1-p)]^3 - [p(1-p)(1-2p)]^2\}m^{-1}, \\
&E[m(\widehat{p}_{k,m,r}-p_{k,\infty,r})^2(\widehat{p}_{\ell,m,r}-p_{\ell,\infty,r})] \\
&\quad = mE[(\widehat{p}_{k,m,r}-p_{k,\infty,r})^2]E[(\widehat{p}_{\ell,m,r}-p_{\ell,\infty,r})] \\
&\quad = 0, \\
&\operatorname{Var}[m(\widehat{p}_{k,m,r}-p_{k,\infty,r})^2(\widehat{p}_{\ell,m,r}-p_{\ell,\infty,r})] \\
&\quad = m^2\operatorname{Var}[(\widehat{p}_{k,m,r}-p_{k,\infty,r})^2] \\
&\qquad \cdot \operatorname{Var}[(\widehat{p}_{\ell,m,r}-p_{\ell,\infty,r})] \\
&\quad \sim 2[p_{k,\infty,r}(1-p_{k,\infty,r})]^2 p_{\ell,\infty,r}(1-p_{\ell,\infty,r})m^{-1}.
\end{aligned}
$$

We find

$$
\begin{aligned}
E\left(n^{-1/2}\sum_{r=1}^{n}B_{1r}\right) &= n^{-1/2}\sum_{r=1}^{n}E(B_{1r}) \\
&= O(n^{1/2}m^{-1}), \\
E\left(n^{-1/2}\sum_{r=1}^{n}B_{3r}\right) &= n^{-1/2}\sum_{r=1}^{n}E(B_{3r}) = 0, \\
\operatorname{Var}\left(n^{-1/2}\sum_{r=1}^{n}B_{1r}\right) &= \sum_{r=1}^{n}\frac{\operatorname{Var}(B_{1r})}{n} = O(m^{-1}), \\
\operatorname{Var}\left(n^{-1/2}\sum_{r=1}^{n}B_{3r}\right) &= \sum_{r=1}^{n}\frac{\operatorname{Var}(B_{3r})}{n} = O(m^{-1}).
\end{aligned}
$$

An application of Chebyshev's inequality leads to

$$
\begin{aligned}
n^{-1/2}\sum_{r=1}^{n}B_{1r} &= O_P(n^{1/2}m^{-1}+m^{-1/2}), \\
n^{-1/2}\sum_{r=1}^{n}B_{3r} &= O_P(n^{1/2}m^{-1}+m^{-1/2}).
\end{aligned}
$$

Similarly, we can show

$$
\begin{aligned}
n^{-1/2}\sum_{r=1}^{n}B_{2r} &= O_P(n^{1/2}m^{-1}+m^{-1/2}), \\
n^{-1/2}\sum_{r=1}^{n}B_{4r} &= O_P(n^{1/2}m^{-1}+m^{-1/2}).
\end{aligned}
$$

Substituting these results into (A.11) we prove (A.10).

A.3 Proof of Theorem 3

As a U-shape is decreasing and then increasing, a upside down flip over of a unimodal shape, similar to Hartigan and Hartigan (1985) using the order restricted inference we can easily show that the U-shape distribution function estimator $\tilde{F}_n$ is equal to the least concave majorant and then the largest convex minorant of $F_n(x)$ with the turning point chosen to minimize the maximum distance between F_n and all possible least concave majorant and then the largest convex minorant; the dip test statistic $D_0(F_n)$ is equal to the maximum distance between the empirical distribution function $F_n(x)$ and $\tilde{F}_n$; the uniform distribution on [0, 1] is the asymptotic least favorable U-shape distribution, and as $n \to \infty$, under the uniform distribution, $\sqrt{n}D_0(F_n)$ converges in distribution to $D_0(W)$, where W is a standard Brownian bridge on [0, 1], and $D_0(W)$ is the maximum difference between W and the concave and then convex function on [0, 1] that minimizes the maximum difference over all such concave-convex functions on [0, 1]. To show $D_0(W)$ and $D(W)$

have the same distribution, we note that Brownian bridge W is symmetric about the horizontal axis with $W(0) = W(1) = 0$, and $-W$ is also a standard Brownian bridge and has the same probabilistic behavior as W, thus $D_0(W)$ has the same distribution as $D_0(-W)$. On the other hand, the least concave majorant of W and the largest convex minorant of W are the reflections about the horizontal axis of the largest convex minorant of $-W$ and the least concave majortant of $-W$, respectively. Therefore, $D(W)$ is identically distributed as $D_0(-W)$, which in turn has the same distribution as $D_0(W)$.

ACKNOWLEDGEMENTS

The research of Yazhen Wang was supported in part by NSF Grants DMS-10-5635, DMS-12-65203 and DMS-15-28375. The authors thank the Editor, Associate Editor, and two anonymous referees for comments and suggestions, which led to improvements of the paper.

REFERENCES

AHARONOV, D., VAN DAM, W., KEMPE, J., LANDAU, Z., LLOYD, S. and REGEV, O. (2007). Adiabatic quantum computation is equivalent to standard quantum computation. *SIAM J. Comput.* **37** 166–194. MR2306288

ALBASH, T., RØNNOW, T. F., TROYER, M. and LIDAR, D. A. (2014). Reexamining classical and quantum models for the D-Wave One processor: The role of excited states and ground state degeneracy. Available at arXiv:1409.3827v1.

ASPURU-GUZIK, A., DUTOI, A. D., LOVE, P. J. and HEAD-GORDON, M. (2005). Simulated quantum computation of molecular energies. *Science* **309** 1704–1707.

BENJAMINI, Y. (2010). Discovering the false discovery rate. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 405–416. MR2758522

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

BERTSIMAS, D. and TSITSIKLIS, J. (1992). Simulated annealing. In *Probability and Algorithms* 17–29. Nat. Acad. Press, Washington, DC. MR1194437

BIAN, Z., CHUDAK, F., MACREADY, W. G., CLARK, L. and GAITAN, F. (2012). Experimental determination of Ramsey numbers. Available at arXiv:1201.1842.

BOIXO, S., ALBASH, T., SPEDALIERI, F. M., CHANCELLOR, N. and LIDAR, D. A. (2014a). Experimental signature of programmable quantum annealing. *Nature Comm.* **4**. 2067.

BOIXO, S., RØNNOW, T. F., ISAKOV, S. V., WANG, Z., WECKER, D., LIDAR, D. A., MARTINIS, J. M. and TROYER, M. (2014b). Evidence for quantum annealing with more than one hundred qubits. *Nature Physics* **10** 218–224.

BOIXO, S., SMELYANSKIY, V. N., SHABANI, A., ISAKOV, S. V., DYKMAN, M., DENCHEV, V. S., AMIN, M., SMIRNOV, A., MOHSENI, M. and NEVEN, H. (2015a). Computational role of collective tunneling in a quantum annealer. Available at arXiv:1411.4036v2.

BOIXO, S., SMELYANSKIY, V. N., SHABANI, A., ISAKOV, S. V., DYKMAN, M., DENCHEV, V. S., AMIN, M., SMIRNOV, A., MOHSENI, M. and NEVEN, H. (2015b). Computational role of multiqubit tunneling in a quantum annealer. Available at arXiv:1502.05754v1.

BORN, M. and FOCK, V. (1928). Beweis des Adiabatensatzes. *Zeitschrift Für Physik A* **51** 165–180.

BRITTON, J. W., SAWYER, B. C., KEITH, A., WANG, C.-C. J., FREERICKS, J. K., UYS, H., BIERCUK, M. J. and BOLLINGER, J. J. (2012). Engineered 2D Ising interactions on a trapped-ion quantum simulator with hundreds of spins. *Nature* **484** 489–492.

BROOKE, J., BITKO, D., ROSENBAUM, T. F. and AEPPLI, G. (1999). Quantum annealing of a disordered magnet. *Science* **284** 779–781.

BROWNE, D. (2014). Model versus machine. *Nature Physics* **10** 179–180.

BRUMFIEL, G. (2012). Simulation: Quantum leaps. *Nature* **491** 322–324.

CAI, T., KIM, D., WANG, Y., YUAN, M. and ZHOU, H. H. (2016). Optimal large-scale quantum state tomography with Pauli measurements. *Ann. Statist.* **44** 682–712. MR3476614

CLARKE, J. and WILHELM, F. K. (2008). Superconducting quantum bits. *Nature* **453** 1031–1042.

DECHTER, R. (1999). Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence* **113** 41–85. MR1724112

DENCHEV, V. S., BOIXO, S., ISAKOV, S. V., DING, N., BABBUSH, R., SMELYANSKIY, V., MARTINIS, J. and NEVEN, H. (2016). What is the computational value of finite range tunneling? Available at arXiv:1512.02206v4.

DEUTSCH, D. (1985). Quantum theory, the Church–Turing principle and the universal quantum computer. *Proc. Roy. Soc. London Ser. A* **400** 97–117. MR0801665

DEVORET, M. H., WALLRA, A. and MARTINIS, J. M. (2004). Superconducting qubits: A short review. Available at arXiv:cond-mat/0411174v1.

DICARLO, L., CHOW, J. M., GAMBETTA, J. M., BISHOP, L. S., JOHNSON, B. R., SCHUSTER, D. I., MAJER, J., BLAIS, A., FRUNZIO, L., GIRVIN, S. M. and SCHOELKOPF, R. J. (2009). Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460** 240–244.

DIVINCENZO, D. P. (1995). Quantum computation. *Science* **270** 255–261. MR1355956

FARHI, E., GOLDSTONE, J., GUTMANN, S. and SIPSER, M. (2000). Quantum computation by adiabatic evolution. Available at arXiv:quant-ph/0001106v1.

FARHI, E., GOLDSTONE, J., GUTMANN, S., LAPAN, J., LUNDGREN, A. and PREDA, D. (2001). A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science* **292** 472–476. MR1838761

FARHI, E., GOLDSTONE, J., GUTMANN, S. and SIPSER, M. (2002). Quantum adiabatic evolution algorithms versus simulated annealing. Available at arXiv:quant-ph/0201031v1.

FEYNMAN, R. P. (1981/82). Simulating physics with computers. *Internat. J. Theoret. Phys.* **21** 467–488. MR0658311

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.

HAJEK, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** 311–329. MR0942621

HARTIGAN, J. A. and HARTIGAN, P. M. (1985). The dip test of unimodality. *Ann. Statist.* **13** 70–84. MR0773153

HEN, I., JOB, J., ALBASH, T., RØNNOW, T. F., TROYER, M. and LIDAR, D. A. (2015). Probing for quantum speedup in spin glass problems with planted solutions. Available at arXiv:1502.01663v2.

HOLEVO, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*. *North-Holland Series in Statistics and Probability* **1**. North-Holland, Amsterdam. MR0681693

IRBACK, A., PETERSON, C. and POTTHAST, F. (1996). Evidence for nonrandom hydrophobicity structures in protein chains. *Proc. Natl. Acad. Sci. USA* **93** 533–538.

JOHNSON, M. W., AMIN, M. H. S., GILDERT, S., LANTING, T., HAMZE, F., DICKSON, N., HARRIS, R., BERKLEY, A. J., JOHANSSON, J., BUNYK, P., CHAPPLE, E. M., ENDERUD, C., HILTON, J. P., KARIMI, K., LADIZINSKY, E., LADIZINSKY, N., OH, T., PERMINOV, I., RICH, C., THOM, M. C., TOLKACHEVA, E., TRUNCIK, C. J. S., UCHAIKIN, S., WANG, J., WILSON, B. and ROSE, G. (2011). Quantum annealing with manufactured spins. *Nature* **473** 194–198.

JONES, N. (2013). D-Wave is pioneering a novel way of making quantum computers—but it is also courting controversy. *Nature* **498** 286–288.

KATO, T. (1978). Trotter's product formula for an arbitrary pair of self-adjoint contraction semigroups. In *Topics in Functional Analysis* (*Essays Dedicated to M. G. KreĭN on the Occasion of His* 70*th Birthday*). *Adv. in Math. Suppl. Stud.* **3** 185–195. Academic Press, New York. MR0538020

KATZGRABER, H. G., HAMZE, F. and ANDRIST, R. S. (2014). Glassy Chimeras could be blind to quantum speedup: Designing better benchmarks for quantum annealing machines. *Phys. Rev. X* **4** 021008.

KECHEDZHI, K. and SMELYANSKIY, V. N. (2015). Open system quantum annealing in mean field models with exponential degeneracy. Available at arXiv:1505.05878v1.

KIRKPATRICK, S., GELATT, C. D. JR. and VECCHI, M. P. (1983). Optimization by simulated annealing. *Science* **220** 671–680. MR0702485

LANTING, T., PRZYBYSZ, A. J., SMIRNOV, A. YU., SPEDALIERI, F. M., AMIN, M. H., BERKLEY, A. J., HARRIS, R., ALTOMARE, F., BOIXO, S., BUNYK, P., DICKSON, N., ENDERUD, C., HILTON, J. P., HOSKINSON, E., JOHNSON, M. W., LADIZINSKY, E., LADIZINSKY, N., NEUFELD, R., OH, T., PERMINOV, I., RICH, C., THOM, M. C., TOLKACHEVA, E., UCHAIKIN, S., WILSON, A. B. and ROSE, G. (2014). Entanglement in a quantum annealing processor. *Phys. Rev. X* **4** 021041.

MAJEWSKI, J., LI, H. and OTT, J. (2001). The Ising model in physics and statistical genetics. *Am. J. Hum. Genet.* **69** 853–862.

MARTIN-MAYOR, V. and HEN, I. (2015). Unraveling quantum annealers using classical hardness. Available at arXiv:1502.02494v1.

MARTOŇÁK, R., SANTORO, G. E. and TOSATTI, E. (2002). Quantum annealing by the path-integral Monte Carlo method: The two-dimensional random Ising model. *Phys. Rev. B* **66** 094203.

MCGEOCH, C. C. (2014). *Adiabatic Quantum Computation and Quantum Annealing*. *Synthesis Lectures on Quantum Computing*. Morgan & Claypool Publisher, San Rafael, CA.

MORITA, S. and NISHIMORI, H. (2008). Mathematical foundation of quantum annealing. *J. Math. Phys.* **49** 125210, 47. MR2484341

NEUMANN, P., MIZUOCHI, N., REMPP, F., HEMMER, P., WATANABE, H., YAMASAKI, S., JACQUES, V., GAEBEL, T., JELEZKO, F. and WRACHTRUP, J. (2008). Multipartite entanglement among single spins in diamond. *Science* **320** 1326–1329.

NIELSEN, M. A. and CHUANG, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge Univ. Press, Cambridge. MR1796805

O'GORMAN, B., BABBUSH, R., PERDOMO-ORTIZ, A., ASPURU-GUZIK, A. and SMELYANSKIY, V. (2014). Bayesian network structure learning using quantum annealing. Available at arXiv:1407.3897v2.

PERDOMO-ORTIZ, A., DICKSON, N., DREW-BROOK, M., ROSE, G. and ASPURU-GUZIK, A. (2012). Finding low-energy conformations of lattice protein models by quantum annealing. *Sci. Rep.* **2** 571.

PERDOMO-ORTIZ, A., FLUEGEMANN, J., NARASIMHAN, S., BISWAS, R. and SMELYANSKIY, V. N. (2014). A quantum annealing approach for fault detection and diagnosis of graph-based systems. Available at arXiv:1406.7601v2.

PUDENZ, K. L., ALBASH, T. and LIDAR, D. A. (2014). Error corrected quantum annealing with hundreds of qubits. *Nature Comm.* **5** 3243.

RIEFFEL, E., VENTURELLI, D., O'GORMAN, B., DO, M., PRYSTAY, E. and SMELYANSKIY, V. (2014). A case study in programming a quantum annealer for hard operational planning problems. Available at arXiv:1407.2887v1.

RIEGER, H. and KAWASHIMA, N. (1999). Application of a continuous time cluster algorithm to the two-dimensional random quantum Ising ferromagnet. *Eur. Phys. J. B* **9** 233–236.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester. MR0961262

RØNNOW, T. F., WANG, Z., JOB, J., BOIXO, S., ISAKOV, S. V., WECKER, D., MARTINIS, J. M., LIDAR, D. A. and TROYER, M. (2014). Defining and detecting quantum speedup. *Science* **25** 420–424. 6195.

SAKURAI, J. J. and NAPOLITANO, J. (2010). *Modern Quantum Mechanics*, 2nd ed. Addison-Wesley, Reading.

SANTORO, G. E., MARTONÁK, R., TOSATTI, E. and CAR, R. (2002). Theory of quantum annealing of an Ising spin glass. *Science* **295** 2427–2430.

SHANKAR, R. (1994). *Principles of Quantum Mechanics*, 2nd ed. Plenum Press, New York. MR1343488

SHIN, S. W., SMITH, G., SMOLIN, J. A. and VAZIRANI, U. (2014). How "quantum" is the D-Wave machine? Available at arXiv:1401.7087v1.

SHOR, P. W. (1994). Algorithms for quantum computation: Discrete logarithms and factoring. In 35*th Annual Symposium on Foundations of Computer Science* (*Santa Fe*, *NM*, 1994) 124–134. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1489242

SMOLIN, J. A. and SMITH, G. (2014). Classical signature of quantum annealing. *Front. Phys.* **2** 52.

STAUFFER, D. (2008). Social applications of two-dimensional Ising models. *Am. J. Phys.* **76** 470–473.

SUZUKI, M. (1976). Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations

with applications to many-body problems. *Comm. Math. Phys.* **51** 183–190. MR0425678

TROTTER, H. F. (1959). On the product of semi-groups of operators. *Proc. Amer. Math. Soc.* **10** 545–551. MR0108732

VENTURELLI, D., MANDRÁ, S., KNYSH, S., O'GORMAN, B., BISWAS, R. and SMELYANSKIY, V. N. (2014). Quantum optimization of fully-connected spin glasses. Available at arXiv:1406.7553v1.

VINCI, W., ALBASH, T., MISHRA, A., WARBURTON, P. A. and LIDAR, D. A. (2014). Distinguishing classical and quantum models for the D-Wave device. Available at arXiv:1403.4228v1.

WANG, Y. (1995). The L_1 theory of estimation of monotone and unimodal densities. *J. Nonparametr. Statist.* **4** 249–261. MR1366772

WANG, Y. (2011). Quantum Monte Carlo simulation. *Ann. Appl. Stat.* **5** 669–683. MR2840170

WANG, Y. (2012). Quantum computation and quantum information. *Statist. Sci.* **27** 373–394. MR3012432

WANG, Y. (2013). Asymptotic equivalence of quantum state tomography and noisy matrix completion. *Ann. Statist.* **41** 2462–2504. MR3127872

WANG, Y. and XU, C. (2015). Density matrix estimation in quantum homodyne tomography. *Statist. Sinica* **25** 953–973. MR3409732

WANG, L., RØNNOW, T. F., BOIXO, S., ISAKOV, S. V., WANG, Z., WECKER, D., LIDAR, D. A., MARTINIS, J. M. and TROYER, M. (2013). Comment on "Classical signature of quantum annealing." Available at arXiv:1305.5837v1.

WINKER, P. (2001). *Optimization Heuristics in Econometrics*: *Applications of Threshold Accepting*. Wiley, Chichester. MR1883767

XU, C. (2015). Statistical analysis of quantum annealing models and density matrix estimation in quantum homodyne tomography. Ph.D. thesis, Univ. Wisconsin-Madison.