Scaled Quantum Circuits Emulated with Room Temperature p-Bits

Kerem Y. Camsari, Shuvro Chowdhury and Supriyo Datta

1School of Electrical and Computer Engineering, Purdue University, IN, 47907

(Dated: October 17, 2018)

Exploiting a well-established mapping from a d-dimensional quantum Hamiltonian to a d+1-dimensional classical Hamiltonian that is commonly used in software quantum Monte Carlo algorithms, we propose a scalable hardware emulator where quantum circuits are emulated with room temperature p-bits. The proposed emulator operates with probabilistic bits (p-bit) that fluctuate between logic 0 and 1, that are suitably interconnected with a crossbar of resistors or conventional CMOS devices. One particularly compact hardware implementation of a p-bit is based on the standard 1 transistor/1 Magnetic Tunnel Junction (1T/1MTJ) cell of the emerging Embedded Magnetoresistive RAM (eMRAM) technology, with a simple modification: The free layer of the MTJ uses a thermally unstable nanomagnet so that the resistance of the MTJ fluctuates in the presence of thermal noise. Using established device models for such p-bits and interconnects simulated in SPICE, we demonstrate a faithful mapping of the Transverse Ising Hamiltonian to its classical counterpart, by comparing exact calculations of averages and correlations. Even though we focus on the Transverse Ising Hamiltonian, many other “stoquastic” Hamiltonians — avoiding the sign problem — can be mapped to the hardware emulator. For such systems, large scale integration of the eMRAM technology can enable the intriguing possibility of emulating a very large number of q-bits by room temperature p-bits. The compact and low-level representation of the p-bit offers the possibility of greater efficiency and scalability compared to standard software implementations of quantum Monte Carlo methods.

CONTENTS

I. Introduction 1
Scope 1
Organization 2

II. Transverse Ising Hamiltonian 3
Quantum Boltzmann Law 3
Averages and correlations 3

III. q-bit to p-bit 3
Behavioral equations of p-bits 4
PSL dynamics 5
PSL vs Quantum Boltzmann Law 5

IV. p-bit to Stochastic MRAM 6
Network parameters 6
Device models 6
Device operation 7
Stochastic MRAM-based p-bit vs Quantum Boltzmann Law 7

V. Conclusions 7
Acknowledgments 8
References 8

I. INTRODUCTION

The basic building block of conventional digital electronics is the CMOS (Complementary Metal Oxide Semiconductor) transistor that is used to represent deterministic bits, that are either 0 or 1. Quantum computing, on the other hand, is based on q-bits that are coherent, delicate superpositions of 0 and 1. It is possible to define an entity in between bits and q-bits that are classical but probabilistic, which we call “p-bits” [1]. Interestingly, in his celebrated paper that helped inspire quantum computing, Feynman started with the description of a probabilistic computer before contrasting it with a quantum computer [2].

It has been shown that just as the three-terminal transistors can be interconnected to hierarchically build complicated circuits to implement specific functions, a three terminal realization of the p-bit can be interconnected to perform useful functionalities, that are broadly relevant in the context of quantum computing and machine learning [3]. For example, p-circuits can be used to perform classical annealing in hardware [4], perform integer factorization by operating multipliers in an invertible mode [1, 5], just like quantum annealers that have been used for similar applications [6, 7]. In the machine learning context, p-bits can function as hardware accelerators of binary stochastic neurons [8] that can be used to become efficient inference engines [9, 10], or they can be used in an efficient calculation of correlations to accelerate learning algorithms [11], an application area also discussed in the context of quantum computing [12–15].
FIG. 1. q-bit to p-bit mapping: (a) A $d$-dimensional q-bit array described by the Transverse Ising Hamiltonian can be mapped to a $(d+1)$-dimensional p-bit array with $n$-replicas that are coupled in the vertical direction by the Trotter-Suzuki decomposition (The case for $d = 2$ is illustrated). In this scheme, the replicas are always connected with periodic boundary conditions such that $m_{i,n+1} = m_{i,n}$. The many-body quantum and the corresponding classical Hamiltonian are shown where the operators $\sigma_{z}, \sigma_{x}$ of the quantum system are replaced with binary p-bits in the classical system with $m_{i,j} \in \{-1, +1\}$. 

Corresponding coupling terms are $(J_{\parallel})_{i,j} = J_{i,j}/n$ and $J_{\perp} = -1/(2\beta)\log\tanh(\beta \Gamma_{x}/n)$. 

uated by a number of classical replicas that are interacting with each other [16] (FIG. 1). This approach is the basis of Path Integral Monte Carlo (PIMC) simulations of quantum systems in software or high-level hardware simulations [17–21]. By contrast, we show that p-bits realized as low-level hardware building blocks can emulate a quantum system by building p-bits as physical replicas operating at room temperature. Therefore, a p-bit based quantum emulator can overcome fundamental difficulties associated with the low temperature operation of quantum annealers [22], as well as naturally allowing all-to-all connectivity beyond nearest neighbor coupling without requiring any special encoding [23, 24]. Moreover, it is possible to engineer arbitrary $k$-body interactions, using conventional electronic devices such as GPUs or FPGAs to interconnect the p-bits. Such interactions are usually difficult to implement for $k > 2$ and it is customary to introduce ancillary bits to map the problem to one with 2-body interactions [25, 26].

Organization

In this paper, we limit our discussion to the Transverse Ising Hamiltonian [27, 28] commonly employed by quantum annealers [29], though the methodology of the mapping is general and can be used for other Hamiltonians, as long as they are stoquastic, a special class of Hamiltonians [30] that avoid the sign problem [31], thereby allowing efficient implementation of PIMC algorithms. In Section II we describe the Transverse Ising Hamiltonian and its exact quantum treatment that we use as a benchmark. In Section III, we show the mapping of the quantum model to the classical model, describing how the behavioral equations describing the dynamics of p-circuits can be used to solve the classical model, comparing the results of p-bits to exact averages and correlations of the quantum system. In Section IV we first describe a particularly compact implementation of the p-bit, consisting of 3 transistors and a 2-terminal Magnetic Tunnel Junction (MTJ), though non-magnetic and completely digital implementations of p-bits are also possible [5, 32]. The proposed hardware p-bit makes use of the standard 1T/MTJ cell of the existing Embedded Magnetoresistive RAM (eMRAM) technology that has been under development by a number of foundries [33–35]. Unlike the standard MRAM where a non-volatile MTJ is carefully engineered with a large energy barrier ($E_{B} \approx 40-60 \, k_{B}T$) so that the magnetization state is retained for a long time [36], the free layer of the MTJ for the p-bit is designed as a thermally unstable magnet ($E_{B} \approx 0 \, k_{B}T$) whose magnetization rapidly fluctuates in time in the presence of thermal noise [37]. In Section IV, using full device-level SPICE simulations corresponding to the p-bit and
a resistive interconnection matrix, we demonstrate that the correct quantum correlations can be obtained using a network of p-bits, establishing the mapping between the quantum and the classical system.

II. TRANSVERSE ISING HAMILTONIAN

The Transverse Ising Hamiltonian in 1D is written as [28]:

$$H_Q = -\left( \sum_{i} J_{i,i+1} \sigma^z_i \sigma^z_{i+1} + \Gamma_x \sum_{i} \sigma^x_i + \Gamma_z \sum_{i} \sigma^z_i \right)$$  \hspace{1cm} (1)

We assume periodic boundary conditions such that $\sigma^z_{i+1} = \sigma^z_i$. $\Gamma_x$ is the local transverse magnetic field and $\Gamma_z$ is a local z-directed magnetic field. Eq. 1 can be constructed by first writing each term, $\sigma^j$, as a $2^M \times 2^M$ matrix followed by ordinary matrix multiplication for each product term. These terms are written in terms of $2 \times 2$ Pauli spin matrices ($\sigma^{x,y,z}$) at the $j^{th}$ lattice point as $\sigma_j = I \otimes I \otimes \ldots \otimes \sigma \otimes \ldots \otimes I \otimes I$ where $I$ is the $2 \times 2$ identity matrix and $\sigma$ is the Pauli spin matrix at the $j^{th}$ term in the product.

Quantum Boltzmann Law

In principle, Eq. 1 can be exactly solved for any quantity of interest as a function of temperature and all other parameters $J, \Gamma$ from the principles of quantum statistical mechanics [38]:

$$\langle S \rangle = \frac{\text{Tr} [S_{op} \exp(-\beta H_Q)]}{\text{Tr} [\exp(-\beta H_Q)]}$$ \hspace{1cm} (2)

where $\beta \equiv 1/k_B T$ is the “inverse temperature” and we have chosen to use a unit system in which $k_B = 1$. $S$ is the quantity we wish to calculate with a corresponding operator $S_{op}$. In practice, directly solving Eq. 2 becomes intractable due to the exponential dependence of the Hamiltonian ($2^M \times 2^M$) to the size of the problem ($M$). Due to its similarity to the classical Boltzmann Law [39], we refer to Eq. 2 as the “Quantum Boltzmann Law” throughout this paper and solve it for small 1D systems. To obtain numerically stable results at low temperatures (high $\beta$), we first diagonalize the Hamiltonian and subtract the ground state energy from the diagonals, without changing any observable quantities.

Averages and correlations

In FIG. 2a we calculate the average $z$-magnetization of a 1D ferromagnetic ($J_{ij} = +2$) chain with $M = 8$ spins, as a function of a transverse magnetic field. The average $z$-magnetization, $\langle m_z \rangle$, is obtained by the operator $\sigma^z = \sum_j \sigma^z_j / M$ where $\sigma^z_j$ provides the net $z$-spin, $\langle \uparrow \rangle - \langle \downarrow \rangle$, at site $j$. To break the symmetry of $m_z = \pm 1$ at low temperatures ($\beta = 10$) we introduce a $+\hat{z}$-directed magnetic field. As the transverse magnetic field increases, $\langle m_z \rangle$ gradually decreases, while $\langle m_x \rangle$ (not shown) increases, as spins become aligned with the transverse magnetic field. Incidentally, the reverse process, starting from a large $\Gamma_x$ at a low temperature and slowly decreasing it to find the ground state of a complicated spin-glass, is commonly used in quantum annealing algorithms [18].

FIG. 2b shows the probabilities of correlated states at a given temperature and transverse field expressed as decimal numbers. This is done by first converting the states to binary numbers such that $\uparrow$ denotes +1 and $\downarrow$ denotes 0 and then converting the full state into a decimal number, for example the all down state $|\downarrow \downarrow \ldots \downarrow \rangle$ corresponds to 0, and the all up state $|\uparrow \uparrow \ldots \uparrow \rangle$ corresponds to 255 and so on. There are $2^8 = 256$ such states, each with a given probability obtained from Eq. 2. These correlated states are calculated by first constructing an operator for the probability of finding a $|\uparrow \rangle$ state at a given site, $P_j(|\uparrow \rangle) = (I + \sigma^z_j)/2$ where $I$ is the $2^M \times 2^M$ identity matrix. Similarly, $P_j(|\downarrow \rangle) = (I - \sigma^z_j)/2$. Using these operators, any correlation of the form $|\uparrow \ldots \uparrow \rangle$ can be calculated from the corresponding composite operator:

$$P(|\downarrow \ldots \uparrow \rangle) = P(|\downarrow \rangle) P(|\uparrow \rangle \ldots P(|\uparrow \rangle = \prod_{k=1}^{M} P_k$$ \hspace{1cm} (3)

There are 256 such operators and Eq. 2 can be used for each of them to obtain a probability for each state for any $J, \Gamma, \beta$. FIG. 2b shows these probabilities at a chosen parameter combination and they are in agreement with results obtained from a simulation of p-bits, as we next explain in Section III. Note that this joint probability density contains all statistical information in the system, as averages and other correlations of interest can be calculated from it, for example one can obtain $\langle m_z \rangle$ by weighting each state by the net $z$-spin they contribute to the average.

III. Q-BIT TO P-BIT

Since the seminal work of Suzuki [16], it is well-known that a $d$-dimensional quantum many-body Hamiltonian can be mapped to a $d+1$-dimensional classical Hamiltonian applying the so-called Trotter-Suzuki decomposition [16, 40], which is used as a basis for PIMC methods to simulate quantum annealing using classical computers [17]. This decomposition results in the quantum system being mapped to a classical system with $n$ replicas that are coupled to each other. For the 1D transverse Ising Hamiltonian, the mapping produces the following classi-
FIG. 2. Exact quantum solution of a 1D-Transverse Ising Hamiltonian vs Probabilistic Spin Logic (PSL): (a) A 1D ferromagnetic linear chain \((J_{ij} = +2)\) with \(M = 8\) spins (nearest neighbor with periodic boundary conditions) described by the quantum transverse Ising Hamiltonian (Eq. 1) is solved exactly, as a function of the transverse magnetic field \((\Gamma_x)\) at an inverse temperature of \(\beta = 10\). A symmetry breaking magnetic field in the \(+\hat{z}\) direction is used, \(\Gamma_z = 1\), so that at \(\Gamma_x = 0\), all spins are pointing in the \(+\hat{z}\) direction. The green dashed line is obtained by evaluating Eq. 2 as a function of \(\Gamma_x\). The red dots represent 100 different PSL runs obtained with different RNGs, each running for \(t_f = 2000\) time steps. The blue triangles represent the average of PSL simulations and closely match the exact solution, establishing the accuracy of the quantum to classical mapping, with \(n = 250\) replicas. (b) A probability histogram of correlations of the form \(|\downarrow\downarrow\ldots\downarrow\rangle = 0, |\uparrow\uparrow\ldots\uparrow\rangle = 255\) are obtained from PSL and Quantum Boltzmann Law at \(\Gamma_x/J = 2.5\) that corresponds to the last point of the \(x\)-axis in (a). Only a portion of the states are shown for clarity, states in between show essentially identical agreement.

Behavioral equations of p-bits

The classical system expressed by Eq. 4 can be represented by p-circuits that are built out of p-bits. There are two central equations that are used to describe p-bit networks [1]:

\[
m_i(t + 1) = \text{sgn}[r + \tanh \beta I_i(t)]
\]  

(5a)

where \(t\) is dimensionless time that is incremented one at a time, \(r\) is a random number uniformly distributed between \(-1\) and \(+1\) and \(r\) at each time step is uncorrelated with the \(r\) chosen at the previous step. \(\beta I_i\) is the dimensionless current to each p-bit, where \(\beta\) is the inverse temperature as previously defined. \(I_i\) in general, is calculated according to,

\[
I_i(t) = -\frac{\partial H_C}{\partial m_i}
\]

(5b)

which in the present case, becomes:

\[
I_i(t) = \left( b_i + \sum_j W_{ij} m_j(t) \right)
\]

where \(W_{ij}\) is the interconnection matrix and \(b_i\) is the bias term. We refer to Eq. 5 as Probabilistic Spin Logic (PSL) equations and note that these equations are essentially the same as those discussed in the context of stochastic neural networks such as Boltzmann Machines, developed by Hinton and colleagues [8].

It is important to note that while Eq. 5b is a linear synapse that typically arises from quadratic Hamiltonians with 2-body interactions, specially designed digital CMOS circuits can be used to implement more complicated interactions arising from cost functions such as generalized Hopfield models with \(k\)-body interactions [41, 42]. Such a flexibility of implementing complicated interactions could be a key advantage for hardware p-circuits.
FIG. 3. p-bit to Stochastic eMRAM: (a) Each replica in the classical system is represented by a hardware neural network involving p-bits (neurons), interconnected by a resistive network (synapse). The outputs of the p-bits are weighted by the resistive network to become inputs to each other. Bias terms are added as fixed external voltage sources. (b) Detailed circuit schematics of a given p-bit and synapse following Ref. [37]: The outputs are collected by a fast operational amplifier (assumed ideal in circuit simulations). Fixed layer ferromagnet (FM) is a stable magnet with a large energy barrier, while the free layer is a circular low-barrier magnet (LBM) \( E_b \approx 0 \ k_B T \) whose magnetization fluctuates in the presence of thermal noise. (c) SPICE simulations for the input-output characteristics of the p-bit: The results for 1000 p-bits where the input voltage is swept from \(-V_{DD}/2\) to \(+V_{DD}/2\), in \( t_{sim} = 1 \) ns (Inset shows the autocorrelation of the p-bit at \( V_{IN} = 0 \)). Each p-bit has a randomized resistance due to the random magnetization of the free layer, showing a range of outputs bounded by the parallel and anti-parallel resistance of the MTJ. Ensemble averaged output for 1000 samples at a given input voltage \( t_{sim} = 2 \) ns for each sample) shows a tanh(\( V_{IN}/V_0 \)) behavior. (d) The circuit model that self-consistently solves the stochastic LLG equation with the MTJ and transistor models. 

PSL dynamics

PSL equations can be updated to approximate the steady state joint probability density for any \( W \) matrix, symmetric or asymmetric. For symmetric \( W \) matrices, the joint probability density is simply expressed by the classical Boltzmann Law, \( \rho(\{m\}) \propto \exp[-\beta E(\{m\})] \), where \( E \) is the energy for a given configuration \( \{m\} \), \( E = 1/2 \ m^T [W] m \). There are two important conditions regarding the updating of Eq. 5. First, Eq. 5b needs to be calculated much faster than Eq. 5a for proper convergence [5], a requirement particularly relevant for hardware implementations. Second, Eq. 5a needs to be updated sequentially, as in Gibbs sampling [43]. The requirement of sequential updating prohibits parallelization in software implementations, except in special cases such as restricted Boltzmann machines where the lack of intralayer connections between “visible” and “hidden” layers allows each layer to be updated in parallel [44]. For asynchronous hardware implementations, however, a clockless operation seems to satisfy the requirement of sequential updating naturally [4, 5].

PSL vs Quantum Boltzmann Law

With this picture, the mapped classical Hamiltonian with \( n \) replicas described Eq. 4 is used to obtain a consolidated [\( W \)] matrix that is of size \((Mn) \times (Mn)\) to be used in Eq. 5. FIG. 2 shows the equivalence of the PSL implementation of the Transverse Ising Hamiltonian to the exact quantum many-body description for a 1D-chain with \( M = 8 \) spins. Note that the p-bit mapping can be applied to much larger spin systems, but an exact solution by Eq. 2 quickly becomes intractable. We investigate the average z-spin of this ferromagnetic chain at a constant temperature (\( \beta = 10 \)) as a function of the transverse magnetic field, \( \Gamma_x \). A symmetry breaking field (to favor a +1 order) of \( \Gamma_z = 1 \) is applied. As expected, the exact result shows how the average z-spin becomes disordered. The PSL results for a \( n = 250 \) replica system reproduce this behavior. The z-spin average is obtained by taking an average over the length of the chain, as well as over each replica. The final average (for a given red dot) is recorded at the end of \( t_f = 2000 \) dimensionless time steps. Since a single stochastic point is recorded at
the end of $t_f$, for each $\Gamma_z$ point, we observe a variance in the final results, however averaging over 100 different simulations for the same system, we get a very close match to the exact solution.

In FIG. 2b, the full joint probability density for the classical system is obtained from a PSL simulation that is run for $t_f = 10^5$ dimensionless time steps. The state of each replica with 8-spins is converted into a binary number at each time step, as in the exact solution, and then collected over all replicas. The striking agreement with PSL and the Quantum Boltzmann Law in FIG. 2 establishes the faithful mapping of the quantum system to the classical system, from the behavioral PSL equations Eq. 5.

IV. P-BIT TO STOCHASTIC MRAM

We now show how the behavioral p-bit model can be represented by a stochastic neural network in hardware (FIG. 3a). Each replica in the classical system consists of p-bits (neurons) that are interconnected to each other with a resistive network (synapse), a typical architecture often used in many hardware neural networks [45, 46] though for more complicated systems involving $k$-body interactions ($k > 2$), standard electronic devices such as FPGA’s could also be used for this purpose, for example as in Ref. [47]. We assume that the weighted summation is carried out by ideal operational amplifiers. The replicas are also connected in the vertical direction (not shown in FIG. 3) with nearest neighbor coupling according to the coupling coefficient $J_{\perp}$.

In the case of quantum annealing, the vertical resistors need to be reconfigurable, therefore they need to be designed differently compared to the fixed resistors that represent the transverse coupling ($J_{\perp}$). In our device level examples, we use fixed resistors in order to establish the equivalence between the classical and quantum systems and have not performed annealing.

Network parameters

The device equations for the synapse and the p-bit shown in FIG. 3 are given as [9, 37]:

$$V_{\text{OUT}ij}/(V_{\text{DD}/2}) = \text{sgn}(r + \tanh(V_{\text{IN}ij}/V_0)) \quad (6)$$

$$V_{\text{IN}ij} = \sum_i \frac{R_{\text{ref}}}{R_{ji}} V_{\text{OUT}ij} \quad (7)$$

Eq. 6 and Eq. 7 are combined with the PSL equations, Eq. 5, to obtain the following equations that map the behavioral PSL equations to physical parameters:

$$m_j = \frac{V_{\text{OUT}ij}}{(V_{\text{DD}/2})}, \quad W_{ji} = \frac{R_0}{R_{ji}}, \quad \beta = \frac{V_{\text{DD}} R_{\text{ref}}}{2 V_0 R_0} \quad (8)$$

where $R_0$ is a unit resistor that is used to electrically change the inverse temperature $\beta$, and $V_0$ is a transistor dependent parameter ($\approx 40 \text{ mV}$) that defines the stochastic window of the p-bit (FIG. 3c). Depending on the sign of the interconnection, $W_{ji}$, the non-inverted output $V_{\text{OUT}ij}$ or the inverted output $\overline{V}_{\text{OUT}ij}$ is used for the synaptic connections.

Device models

The 1T/1MTJ p-bit is modeled by combining a 14 nm-High Performance FinFET model from the open source Predictive Technology Models (PTM) [48] with a stochastic Landau-Lifshitz-Gilbert (sLLG) solver implemented in SPICE [49], following the design described in [37] (FIG. 3d). The MTJ is modeled as a simple conductor whose conductance depends on the instantaneous magnetization $m_z(t)$, provided by the sLLG such that

$$G_{\text{MTJ}}(t) = G_0 \left[1 + m_z(t) \frac{R_{AP} - R_P}{R_{AP} + R_P}\right] \quad (9)$$

where $R_P$ and $R_{AP}$ are the parallel and antiparallel resistance of the MTJ and $G_0 = (R_{AP}^{-1} + R_P^{-1})/2$. We use an experimentally measured value for the tunneling magnetoresistance (TMR) = $(R_{AP} - R_{P})/R_P = 110\%$ after Ref. [33]. $G_0$ is set equal to the transistor resistance at $V_{\text{IN}} = 0$ to produce a symmetric sigmoid with no offsets, in this case $G_0^{-1} = 23.4 \Omega$. The free layer is assumed to be a circular low barrier nanomagnet [50, 51] with a diameter of 22 nm and thickness of 2 nm and a saturation magnetization of $M_s = 1100 \text{emu/cc}$, with a damping coefficient $\alpha = 0.01$, typical parameters for CoFeB [52].
The time dependent magnetization is obtained by solving sLLG equation in the monodomain approximation [53]:

\[
(1 + \alpha^2) \frac{d\hat{m}}{dt} = -|\gamma| \left( \hat{m} \times \hat{H} \right) + \alpha |\gamma| \left( \hat{m} \times \hat{m} \times \hat{H} \right) + \frac{1}{qN} \left( \hat{m} \times \hat{I}_S \times \hat{m} \right) + \frac{\alpha}{qN} \left( \hat{m} \times \hat{I}_S \right)
\]

(10a)

\( \gamma \) is the electron gyromagnetic ratio, \( q \) is electron charge and \( N \) is the number of Bohr magnetons (\( \mu_B \)) in the volume of the magnet, \( N = M_s V_{\text{Vol}}/\mu_B \). \( \hat{H} \) contains the external magnetic and internal anisotropy fields of the magnet as well as the noise field. In the case of a circular nanomagnet without an easy-axis anisotropy, the total internal magnetic anisotropy becomes \( \hat{H}_{m} = -4\pi M_s m_z \hat{x} \), where \( \hat{x} \) and \( \hat{y} \) are the easy plane of the magnet. The thermal noise is added in three directions (\( \hat{x}, \hat{y} \) and \( \hat{z} \)) with zero mean and \( \langle \hat{H}_{\text{noise}}^2 \rangle = 2\alpha k_B T/[(1 + \alpha^2)(\gamma M_s V_{\text{Vol}})] \) in units [Oe^2/s] [54].

**Device operation**

The p-bit shown in FIG. 3c is a series-resistance controlled device where the transistor resistance can be made much smaller or much larger compared to the fluctuating MTJ resistance. Therefore, the operation of the p-bit does not require manipulating the magnetization of the free layer unlike in standard spin-transfer-torque MRAM cells. However, the current flowing through the fixed layer of the MTJ produces a spin-polarized spin current that can unintentionally torque the magnet. We assume that this current is given as \( \hat{I}_s = P I_{\text{MTJ}} \hat{z} \), where \( \hat{z} \) is the fixed layer orientation and \( P \) is an interface polarization that can be related to TMR [55]. This spin-current is fed back to the sLLG solver and fully accounted for in the calculation of magnetization in our simulations, however for the circular LBM with a large demagnetization field used here, its effects are negligible [56]. Using these models, FIG. 3c shows transient SPICE simulations of a single p-bit output, \( V_{\text{OUT}} \) for 1000 samples where \( V_{\text{IN}} \) is rapidly swept in 2 ns. The range of stochastic outputs is bounded by a distribution of resistances ranging from \( R_P \) to \( R_{AP} \). The ensemble average shows an approximate hyperbolic tangent behavior that allows the mapping shown in Eq. 6.

The insert of FIG. 3d shows the autocorrelation time of the circular in-plane magnet with a lifetime of \( \approx 100 \) ps. The fluctuations for a circular magnet is expected to be faster compared to a magnet with perpendicular anisotropy due to the strong demagnetizing field that keeps the magnetization vector in the easy plane of the magnet [57]. The very short lifetime of such a circular low barrier magnet could allow very fast and efficient sampling times, as long as the interconnection network operates faster than these timescales. In present simulations, the resistive network operates instantaneously with an ideal operational amplifier therefore this requirement is met naturally, however in real implementations the synapse needs to be designed carefully.

The second requirement, the need for sequential updating of each p-bit is met naturally since the probability of simultaneous flips among p-bits is extremely unlikely, therefore hardware p-bits go through an effectively random update order that does not affect their final distribution.

**Stochastic MRAM-based p-bit vs Quantum Boltzmann Law**

In FIG. 4, using full SPICE simulations for a 40 p-bit network we compute the joint probability density of a \( M = 4 \) spin ferromagnetic chain (\( J_{ij} = +1 \)) using 10 replicas, with \( \beta = 0.5 \) and \( \Gamma_x = 10 \). Unlike FIG. 2, no symmetry breaking field is applied and the network is asynchronously operated for \( t_{\text{sim}} = 250 \) ns, with a time step of 1 ps. All analog voltage values at the end of the SPICE simulation are thresholded (\( > 0 \) \( \equiv \) 1, \( < 0 \) \( \equiv \) −1) and a time-average is obtained similar to the PSL averaging after converting the state of each p-bit to binary and then to decimal. The results from the full device models seem to be in good agreement with the exact solution obtained from Eq. 2 and the behavioral PSL equations that are included for reference. Note the suppression of states \( 5 = (0101) \) and \( 10 = (1010) \) that correspond to the energetically unfavorable antiferromagnetic configurations \( |\uparrow\uparrow\downarrow\downarrow\rangle \) and \( |\downarrow\downarrow\uparrow\uparrow\rangle \), respectively. The agreement between the full SPICE models with the behavioral and exact solutions establishes the feasibility of the proposed quantum circuit emulator.

**V. CONCLUSIONS**

We have presented a scalable, room-temperature quantum emulator using stochastic p-bits that can be built by a simple modification of the existing 1T/1MTJ cell of the eMRAM technology. The proposed emulator uses physical replicas for repeated Trotter slices used in software Quantum Monte Carlo methods. Having physical replicas for each slice could enable better scaling properties for quantum annealing compared to classical annealing as discussed in [17], since choosing the optimal number of replicas or probing each replica separately to find better energy minima is possible in a physically engineered design, unlike in real quantum systems [18]. The electrical control of annealing parameters, inverse temperature (\( \beta \)) and transverse field (\( \Gamma_x \)), could allow a very large number of q-bits to be reliably emulated with room temperature p-bits. Using conventional electronic devices such as GPU’s or FPGA’s to implement the synapses, it should be possible to engineer complicated interactions that extend beyond nearest neighbors and/or involve \( k \)-body interactions (\( k > 2 \)). We note that even though the “sign problem” limits the universal use of our p-computer, a large number of practically relevant quantum systems
could be efficiently emulated by it, considering a large number of optimization problems have been mapped on to the Transverse Ising Hamiltonian [58]. Our results provide a method of emulating quantum systems with probabilistic hardware in advance of a scalable universal quantum computer.

[1] Kerem Yunus Camsari, Rafatul Faria, Brian M Sutton, and Supriyo Datta, “Stochastic p-bits for invertible logic,” Physical Review X 7, 031014 (2017).
[2] Richard P Feynman, “Simulating physics with computers,” International journal of theoretical physics 21, 467–488 (1982).
[3] Kerem Y Camsari, Brian M Sutton, and Supriyo Datta, “p-bits for probabilistic spin logic,” arXiv preprint arXiv:1809.04028 (2018).
[4] Brian Sutton, Kerem Yunus Camsari, Behtash Behin-Aein, and Supriyo Datta, “Intrinsic optimization using stochastic nanomagnets,” Scientific Reports 7, 44370 (2017).
[5] Ahmed Zeeshan Pervaiz, Lakshmi Anirudh Ghantasala, Kerem Yunus Camsari, and Supriyo Datta, “Hardware emulation of stochastic p-bits for invertible logic,” Scientific reports 7, 10994 (2017).
[6] Roman Martońak, Giuseppe E Santoro, and Erio Tosatti, “Quantum annealing of the traveling-salesman problem,” Physical Review E 70, 057701 (2004).
[7] Xinhua Peng, Zeyang Liao, Nanyang Xu, Gan Qin, Xi-anyi Zhou, Dieter Suter, and Jiangfeng Du, “Quantum adiabatic algorithm for factorization and its experimental implementation,” Physical review letters 101, 220405 (2008).
[8] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski, “A Learning Algorithm for Boltzmann Machines,” Cognitive Science 9, 147–169 (1985).
[9] Rafatul Faria, Kerem Y Camsari, and Supriyo Datta, “Implementing bayesian networks with embedded stochastic mram,” AIP Advances 8, 045101 (2018).
[10] Ramtin Zand, Kerem Yunus Camsari, Steven D Pyle, Ibrahim Ahmed, Chris H Kim, and Ronald F DeMara, “Low-energy deep belief networks using intrinsic sigmoidal spintronic-based probabilistic neurons,” in Proceedings of the 2018 on Great Lakes Symposium on VLSI (ACM, 2018) pp. 15–20.
[11] Rafatul Faria, Jan Kaiser, Orchi Hassan, Kerem Yunus Camsari, and Supriyo Datta, “Accelerating machine learning using stochastic embedded mtj,” unpublished.
[12] Zhengbing Bian, Fabian Chudak, William G Macready, and Geordie Rose, “The ising model: teaching an old problem new tricks,” D-wave systems 2 (2010).
[13] Steven H Adachi and Maxwell P Henderson, “Application of quantum annealing to training of deep neural networks,” arXiv preprint arXiv:1510.06356 (2015).
[14] Jeremy Liu, Federico M Spedalieri, Ke-Thia Yao, Thomas E Potok, Catherine Schuman, Steven Young, Robert Patton, Garrett S Rose, and Gangotree Chamka, “Adiabatic quantum computation applied to deep learning networks,” Entropy 20, 380 (2018).

ACKNOWLEDGMENTS

KYC is grateful to Brian M. Sutton for helpful discussions. This work was supported in part by ASCENT, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.
[29] Mark W Johnson, Mohammad HS Amin, Suzanne Gildert, Trevor Lanting, Firas Hamze, Neil Dickson, R Harris, Andrew J Berkley, Jan Johansson, Paul Bunyk, et al., “Quantum annealing with manufactured spins,” Nature 473, 194 (2011).

[30] Tameem Albash and Daniel A. Lidar, “Adiabatic quantum computation,” Rev. Mod. Phys. 90, 015002 (2018).

[31] Matthias Troyer and Uwe-Jens Wiese, “Computational complexity and fundamental limitations to fermionic quantum monte carlo simulations,” Physical review letters 94, 170201 (2005).

[32] Ahmed Zeeshan Pervaiz, Brian M. Sutton, Lakshmi Anirudh Ghanatasala, and Kerem Y. Camsari, “Weighted p-bits for FPGA implementation of probabilistic circuits,” arXiv:1712.04166 [cs] (2017), arXiv: 1712.04166.

[33] CJ Lin, SH Kang, YJ Wang, K Lee, X Zhu, WC Chen, X Li, WN Hsu, YC Kao, MT Liu, et al., “45nm low power cmos logic compatible embedded stt-mram utilizing a reverse-connection 1t/1mtj cell,” in Electron Devices Meeting (IEDM), 2009 IEEE International (IEEE, 2009) pp. 1–4.

[34] YJ Song, JH Lee, HC Shin, KH Lee, K Suh, JR Kang, SS Pyo, HT Jung, SH Hwang, GH Koh, et al., “Highly functional and reliable 8mb stt-mram embedded in 28nm logic,” in Electron Devices Meeting (IEDM), 2016 IEEE International (IEEE, 2016) pp. 27–28.

[35] D. Shum, D. Houssameddine, S. T. Woo, Y. S. You, J. Wong, K. W. Wong, C. C. Wang, K. H. Lee, K. Yamane, V. B. Naik, C. S. Seet, T. Tahmasebi, C. Haf, H. W. Yang, N. Thiyagarajah, R. Chao, J. W. Ting, N. L. Chung, T. Ling, T. H. Chan, S. Y. Siah, R. Nair, S. Deshpande, R. Whig, K. Nagel, S. Aggarwal, M. DeHerrera, J. Janesky, M. Lin, H. J. Chia, M. Hossain, H. Lu, S. Iekawa, F. B. Mancoff, G. Shimom, J. M. Slaughter, J. J. Sun, M. Tran, S. M. Alam, and T. Andre, “Cmos-embedded stt-mram arrays in 2x nm nodes for gmcu applications.” in 2017 Symposium on VLSI Technology (2017) pp. T208–T209.

[36] Sabpreet Bhatti, Rachid Shaia, Atsufumi Hirohata, Hideo Ohno, Shunsuke Fukami, and SN Piranmanayagam, “Spintronics based random access memory: A review,” Materials Today (2017).

[37] Kerem Yunus Camsari, Sayeef Salahuddin, and Supriyo Datta, “Implementing p-bits with embedded mtj,” IEEE Electron Device Letters 38, 1767–1770 (2017).

[38] Leo P Kadanoff and Gordon A Baym, “Quantum statistical mechanics: Green’s function methods in equilibrium and nonequilibrium problems,” (1962).

[39] Richard P Feynman, Robert B Leighton, and Matthew Sands, The Feynman lectures on physics, Vol. I: The new millennium edition: mainly mechanics, radiation, and heat, Vol. 1 (Basic books, 2011).

[40] Hale F Trotter, “On the product of semi-groups of operators,” Proceedings of the American Mathematical Society 10, 545–551 (1959).

[41] E Gardner, “Multiconnected neural network models,” Journal of Physics A: Mathematical and General 20, 3453 (1987).

[42] Yuya Seki and Hidetoshi Nishimori, “Quantum annealing with antiferromagnetic transverse interactions for the hopfield model,” Journal of Physics A: Mathematical and Theoretical 48, 335301 (2015).

[43] Stuart Geman and Donald Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” IEEE Transactions on pattern analysis and machine intelligence, 721–741 (1984).

[44] Geoffroy E Hinton, “A practical guide to training restricted boltzmann machines,” in Neural networks: Tricks of the trade (Springer, 2012) pp. 599–619.

[45] Shimeng Yu, Yi Wu, Rakesh Jeyasingh, Duygu Kuzum, and H-S Philip Wong, “An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation,” IEEE Transactions on Electron Devices 58, 2729–2737 (2011).

[46] Miao Hu, John Paul Strachan, Zhiyong Li, Emmanuelle M Grafals, Norica Davila, Catherine Graves, Sity Lam, Ning Ge, Jianhua Joshua Yang, and R Stanley Williams, “Dot-product engine for neuromorphic computing: programming it1m crossbar to accelerate matrix-vector multiplication,” in Proceedings of the 53rd annual design automation conference (ACM, 2016) p. 19.

[47] Peter L McMahon, Alireza Marandi, Yoshihaka Haribara, Ryan Hamerly, Carsten Langrock, Shuhei Tamate, Takahiro Inagaki, Hiroki Takesue, Shoko Utsumi, Kazuyuki Aihara, et al., “A fully programmable 100-spin coherent ising machine with all-to-all connections,” Science 354, 614–617 (2016).

[48] “Predictive Technology Model (PTM) (http://ptm.asu.edu/)”,

[49] M. M. Torunbalci, P. Upadhyaya, S. A. Bhave, and K. Y. Camsari, “Modular compact modeling of mtj devices,” IEEE Transactions on Electron Devices, 1–7 (2018).

[50] R. P. Cowburn, D. K. Koltsos, A. O. Adeyeye, M. E. Welland, and D. M. Tricker, “Single-domain circular nanomagnets,” Physical Review Letters 83, 1042 (1999).

[51] Punyashloka Debashis, Rafatul Faria, Kerem Yunus Camsari, and Zhihong Chen, “Designing stochastic nanomagnets for probabilistic spin logic,” IEEE Magnetics Letters (2018).

[52] Jack C Sankey, Yong-Tao Cui, Jonathan Z Sun, John C Slonczewski, Robert A Buhrman, and Daniel C Ralph, “Measurement of the spin-transfer-torque vector in magnetic tunnel junctions,” Nature Physics 4, 67 (2008).

[53] Jonathan Z Sun, “Spin-current interaction with a monodomain magnetic body: A model study,” Physical Review B 62, 570 (2000).

[54] Z Li and S Zhang, “Thermally assisted magnetization reversal in the presence of a spin-transfer torque,” Physical Review B 69, 134416 (2004).

[55] Deepanjan Datta, Behtash Behin-Aein, Supriyo Datta, and Sayeef Salahuddin, “Voltage asymmetry of spin-transfer torques,” IEEE Transactions on Nanotechnology 11, 261–272 (2012).

[56] Rafatul Faria, Kerem Yunus Camsari, and Supriyo Datta, “Low-barrier nanomagnets as p-bits for spin logic,” IEEE Magnetics Letters 8, 1–5 (2017).

[57] L Lopez-Diaz, L Torres, and E Moro, “Transition from ferromagnetism to superparamagnetism on the nanosecond time scale,” Physical Review B 65, 224406 (2002).

[58] Andrew Lucas, “Ising formulations of many np problems,” Frontiers in Physics 2, 5 (2014).