VI. Coherent Ising Machines

In this Chapter, we will discuss the operational principles of two types of optical neural networks operating at quantum limit ($\hbar\omega \gg k_B T$), which are the optical delay line coupling QNN and measurement feedback QNN, and compare their performances with those of classical neural networks (CNN) operating at the opposite thermal limit ($k_B T \gg \hbar\omega$). In the case of optical delay line coupling scheme, the quantum noise correlation formed among DOPOs at pump rates below the threshold is an important preparation step for the decision making process at the DOPO phase transition point. On the other hand, in the case of measurement feedback scheme based on local operation and classical communication (LOCC), such quantum correlation does not exist in the system. Instead, another important preparation step can be identified for the decision making process at the DOPO threshold, which is the contextuality in quantum measurements. Finally, the correlated symmetry breaking and the bosonic final state stimulation provide a crucial mechanism to amplify a weak quantum solution to a strong classical solution.

Quantum neural networks can solve various hard problems by mapping them to either the NP-hard Ising problems or the NP-complete satisfiability (SAT) problems. According to the computational complexity theorem, the three-dimensional Ising model and the two-dimensional Ising model with a dc field belong to the NP-hard class [1]. Various combinatorial optimization problems, such as maximum clique problems, graph coloring problems and many others, can be directly mapped to the Ising model [2] and thus they are universally solved by the coherent Ising machine.

6.1 Correlated spontaneous symmetry breaking in delay line coupling QNN

When two DOPOs are coupled with an anti-ferromagnetic interaction, $J_{12} < 0$, the in-phase amplitudes x_1 and x_2 of the two DOPOs should be negatively correlated not only at above the oscillation threshold but also at below the oscillation threshold. This means that the anti-squeezed in-phase amplitude noise $\triangle x_1$, and $\triangle x_2$ should be negatively correlated by the mutual coupling. Figure 1(a) and (b) show the numerical simulation results, which demonstrate such negative quantum noise correlations [3]. The positive $P(\alpha, \beta)$ representation for the density operator was used in the master equation and the resulting c-number

stochastic differential equations (CSDE) were numerically integrated for this study. The details of the particular quantum theory were already given in Chapter III. In Fig. 1(a), the external pump rate is linearly increased from $p = 0$ at $\tau = 0$ to $p = 1.5$ (1.5 times threshold pump rate) at $\tau = 200$. The signal photon decay rate γ_s for the two DOPOs are varied from $\gamma_s = 0.05$ to $\gamma_s = 5$, while the coupling path photon decay rate γ_c is set equal to $2\gamma_s$. The DOPO threshold is reached at $\tau \simeq 30$, 70, 110, and 160 for $\gamma_s = 0.05 - 0.1$, 0.5, 1, and 5, respectively. The negative noise correlation in $\hat{x}_1$ and $\hat{x}_2$ normalized by the variance product $\sigma_{x1}\sigma_{x2}$ grows up to the maximum value of -1 as the pump rate approaches the threshold except for a very large γ_s value. This result means the two DOPOs realize the microscopic and macroscopic out-of-phase correlation in $\hat{x}_1$ and $\hat{x}_2$ at below and above the oscillation threshold, respectively. In Fig. 1(b), γ_s is fixed to a small value (0.01) and γ_c is varied from 0.1 to 10. It is noted that there is an optimum loss value γ_c for the mutual coupling path where the degree of negative correlation is maximized. If the mutual coupling path is too lossy ($\gamma_c = 10$), the information exchange about $J_{12} < 0$ between the two DOPOs is overwhelmed by the vacuum noise injected into the mutual coupling path. On the other hand, if the mutual coupling path loss rate is too small ($\gamma_c = 0.1$), the in-phase amplitude noise σ_{x1} and σ_{x2} grow very rapidly due to mutual noise injection so that the normalized correlation decreases [3].
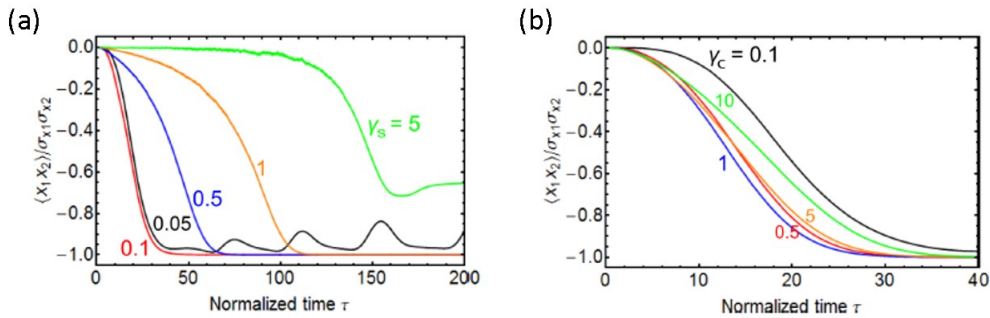

FIG. 1: (Color online) Time dependence of the normalized correlation functions for the in-phase amplitudes x_1 and x_2 of the two DOPOs. (a) The loss in both signal and mutual coupling path with $\gamma_c = 2\gamma_s$ and (b) only the loss in the mutual coupling path γ_c are varied. x_1 and x_2 are negatively and macroscopically correlated with the oscillation. 50 000 stochastic runs for each curve [3].

The quadrature-phase amplitudes p_1 and p_2 of the two DOPOs, on the other hand, feature the positive correlation as shown in Fig. 2(a) and (b), despite the two DOPOs are mutually coupled by anti-ferromagnetic (out-of-phase) interaction. This paradoxial result can be understood as a result of the gain saturation induced by the in-phase amplitudes.

2

We also observed the positive correlation for $\langle \hat{x}_1 \hat{x}_2 \rangle / \sigma_{x1}\sigma_{x2}$ and negative correlation for $\langle \hat{p}_1 \hat{p}_2 \rangle / \sigma_{p1}\sigma_{p2}$ for the ferromagnetic interaction case [3].
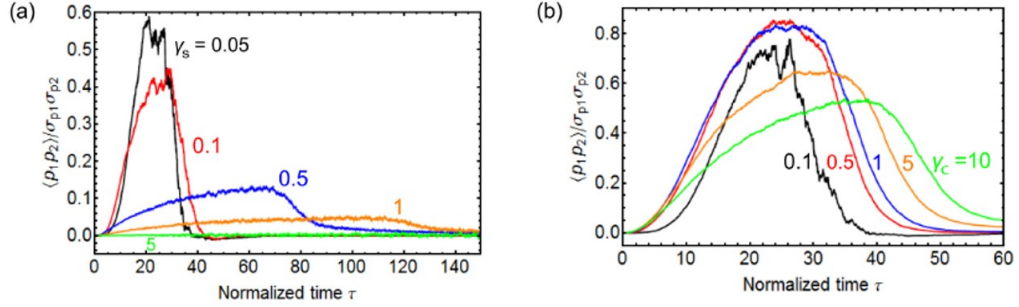

FIG. 2: (Color online) Time dependence of the correlation functions for the quadrature-phase amplitudes p_1 and p_2. (a) The loss in both signal cavity and mutual coupling path with $\gamma_c = 2\gamma_s$ and (b) only γ_c are changed. p_1 and p_2 are positively and microscopically correlated near the oscillation threshold. 50 000 stochastic runs for each curve [3].

The EPR-like operators $\hat{u}_+ = \hat{x}_1 + \hat{x}_2$ and $\hat{v}_- = \hat{p}_1 - \hat{p}_2$ commute, so that a simultaneous eigenstate of $\hat{x}_1 + \hat{x}_2$ and $\hat{p}_1 - \hat{p}_2$ should exist [4]. Such a simultaneous eigenstate of u_+ and v_- satisfies $\langle \triangle \hat{u}_+^2 \rangle = \langle \triangle \hat{v}_-^2 \rangle = 0$ but requires an infinite amount of energy to produce. If the total variance is non-zero but $\langle \triangle \hat{u}_+^2 \rangle + \langle \triangle \hat{v}_-^2 \rangle < 1$ is satisfied, such a state can be identified as an entangled state and produced with a finite energy. Figure 3 shows the total variance, $\langle \triangle \hat{u}_+^2 \rangle + \langle \triangle \hat{v}_-^2 \rangle$, of the DOPOs coupled with anti-ferromagnetic interaction as a function of normalized time. When both signal loss γ_s and the mutual coupling path loss γ_c are small, the total variance sharply rises to above the vacuum noise level, $\langle \triangle \hat{u}_+^2 \rangle + \langle \triangle \hat{v}_-^2 \rangle = 1$, in spite of the (positive) quantum correlation in p_1 and p_2, as seen in Fig. 2(a) and (b). Even though the variance of $\hat{v}_- = \hat{p}_1 - \hat{p}_2$ falls under the vacuum noise level of 0.5, that of $\hat{u}_+ = \hat{x}_1 + \hat{x}_2$ gets much larger than the vacuum noise level, $\langle \triangle \hat{u}_+^2 \rangle = 0.5$, due to the insufficient negative correlation among the amplified in-phase amplitude noise, as shown in Fig.1(a) and (b).

When the mutual coupling path loss γ_c is increased with keeping the signal loss γ_s small, the total variance $\langle \triangle \hat{u}_+^2 \rangle + \langle \triangle \hat{u}_-^2 \rangle$ is reduced to below the vacuum noise level. Thus, the two DOPOs have the entanglement in these parameter regions. The dissipation of the field in the mutual coupling path reduces the anti-squeezed in-phase amplitude noise but still maintain a coherent mutual injection between the two DOPOs. Such a lossy coupling channel can still convey some form of quantum information, i.e. it is not considered as a simple local operation and classical communication (LOCC). We note that the entanglement shown in
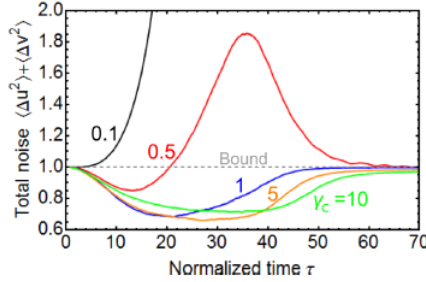
3

FIG. 3: (Color online) Time evolution of the total fluctuation in the EPR-type operator $\langle \triangle \hat{u}_+ \rangle + \langle \triangle \hat{v}_-^2 \rangle$ dependent on the loss in the injection path γ_c. $\langle \triangle \hat{u}_+^2 \rangle + \langle \triangle \hat{v}_-^2 \rangle < 1$ means the entanglement between the two DOPO fields. 50 000 stochastic runs for each curve [3].

Fig. 3 solely comes from the negative quantum correlation in p_1 and p_2, i.e. $\langle \triangle \hat{v}_-^2 \rangle < 0.5$ [3]. $\langle \triangle \hat{u}_+^2 \rangle$ is reduced to the vacuum noise level of 0.5, but is always larger than 0.5, which indicates the negative correlation in x_1 and x_2 does not penetrate into the quantum regime. However, if we inject a squeezed vacuum state from the open port of the out-coupler (see Fig. 9 in Chapter I), the correlation between x_1 and x_2 penetrates into the quantum regime ($\langle \triangle \hat{u}_+^2 \rangle < 0.5$) [3].

We also note that the quantum correlation in p_1 and p_2 manifests itself in the emergence of quantum discord [3]. A quantum discord is a measure of quantum correlation between sub-systems in a compound system and used for judging if a local measurement performed for a sub-system changes the state of the whole system [5]. The details of a quantum discord is given in Chapter III. Figure 4 shows the approximate quantum discord when the state is considered as a Gaussian state. When p_1 and p_2 are squeezed and maintain a positive correlation as shown in Fig. 2(b), the two DOPOs show a relatively large quantum discord. The final state of the two DOPOs at far above the threshold should be a statistical mixture of the negatively correlated coherent states:

$$\hat{\rho} = \frac{1}{2} |\alpha\rangle_1 |-\alpha\rangle_{2\,2} \langle -\alpha|_1 \langle \alpha| + \frac{1}{2} |-\alpha\rangle_1 |\alpha\rangle_{2\,2} \langle \alpha|_1 \langle -\alpha|, \tag{1}$$

where α is positive and real. The Gaussian discord for this state is not $D = 0$ but $D = 0.02356$ for $\alpha \gg 1$, which is in good agreement with the residual quantum discord shown in Fig. 4. The state given by (Eq.(1)) is a statistical mixture of coherent states but yet features a genuine quantum correlation.

Each DOPO pulse will select either 0-phase or π-phase oscillation randomly at the oscillation threshold, if the two DOPO pulses are decoupled. However, if the anti-ferromagnetic
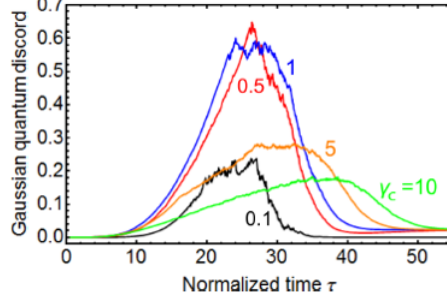
4

FIG. 4: (Color online) Quantum discord when the state is approximated as a bipartite Gaussian state. Squeezing in the DOPOs below the threshold and the mutual injections give a large discord. Coherent fields above the threshold in them and a coherent communication lead to a small but finite discord $D = 0.02$. 50 000 stochastic runs for each curve [3].

(out-of-phase) coupling is turned on between the two DOPOs, the anti-squeezed in-phase amplitude noises of the two DOPOs are negatively correlated at pump rates below oscillation threshold, as shown in Fig. 5(a). This is a crucial preparation stage for the correlated spontaneous symmetry breaking which will emerge at the threshold. The two DOPOs realize the spontaneous symmetry breaking in a correlated way at above threshold as shown in Fig. 5(b). If the DOPO1 selects the 0-phase (π-phase), the DOPO2 selects the π-phase (0-phase) due to the existence of the negative correlation between x_1 and x_2.
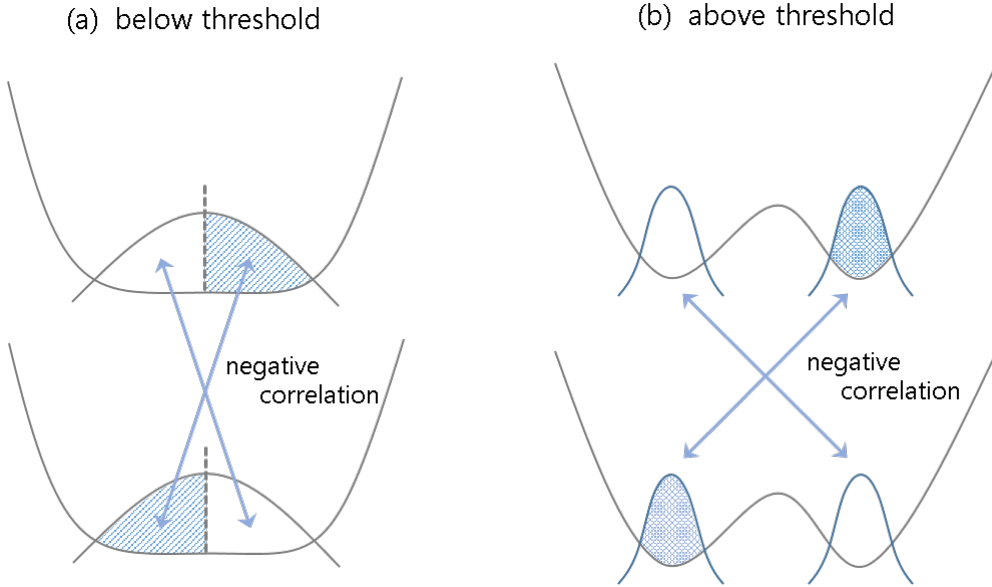

FIG. 5: The potentials of the two DOPOs at (a) below the threshold and (b) above the threshold. The two DOPOs exhibit the spontaneous symmetry breaking in a correlated way.

5

6.2 Contextuality in measurement feedback QNN

It is a fundamental nature of quantum mechanics that a measurement does not reveal a pre-existing value of the measured quantity, except for a special case that the measured system is prepared in an eigenstate of the measured observable. The measurement result is rather created by a joint interaction of a measured system and a measuring apparatus. Quantum mechanics is simply silent for what particular measurement result is obtained for a single event. The lack of causality in quantum measurements is an inherent property of quantum systems and quoted as the famous Heisenberg's "transition from the possible to the actual". This quantum doctrine demands such a weak link between our knowledge and the object.

However, we have shown in Chapter IV that it is not necessary for us to stop there but we can penetrate this subtle issue a bit further by discussing the contextual nature of quantum measurements [6–8]. Contextuality in quantum measurements means two things: 1) A present measurement result depends on the history of the previous measurement results that were performed for the same system, and 2) A present measurement result depends on the future measurement results that will be performed for a same system. We will describe in this section the dynamics of a particular quantum system (measurement feedback QNN) under repeated measurement and feedback control in this context. In particular, it is shown that such a contextual nature of the system makes a transition from truly stochastic measurement results to eventually deterministic measurement results.

Figure 6 shows the system configuration of a measurement feedback QNN, in which a part of the j-th DOPO pulse is picked off by the output coupler, the in-phase amplitude x_j is measured by homodyne detectors, the feedback pulse amplitude, $\sum_j J_{ij}\tilde{x}_j$, to the i-th DOPO pulse is computed in the field programmable gate array (FPGA), and the feedback pulse is produced by an optical phase/amplitude modulator based on the value of $\sum_j J_{ij}\tilde{x}_j$ and injected into the i-th DOPO pulse [9, 10]. In this way, if the system has N DOPO pulses, a single measurement feedback circuit can implement $\sim O\left(N^2\right)$ coupling terms in one round trip time.

Note that the homodyne measurement of the in-phase amplitude $\tilde{x}_j$ of the j-th DOPO pulse partially collapses its wavefunction and the feedback pulse injection, based on the measurement result $\tilde{x}_j$, into the i-th DOPO pulse establishes a correlation between the two
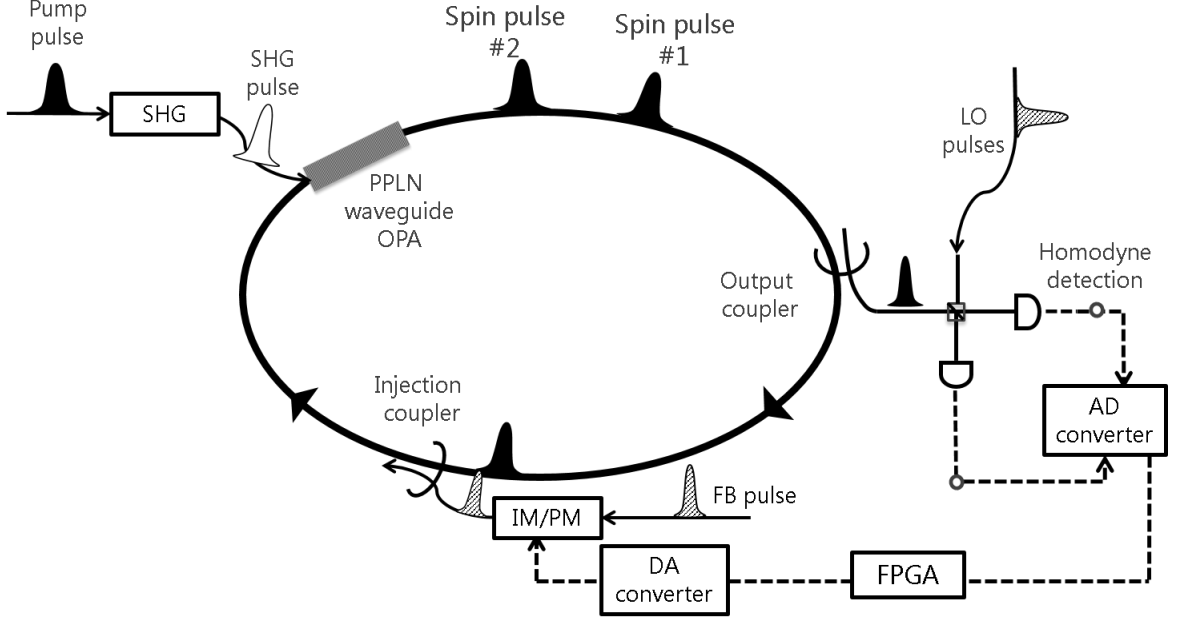
FIG. 6: Quantum neural network (QNN) based on time-division-multiplexed (TDM) degenerate optical parametric oscillator (DOPO) pulses with measurement and feedback control. Both local oscillator (LO) pulses and feedback (FB) pulses are taken from the pump laser. A parametric gain is provided by a periodically-poled LiNbO3 (PPLN) waveguide device and an optical ring cavity is formed by a fiber with ~ 1 km length [9, 10].

wavepackets. The intra-cavity DOPO pulse at the output of the PSA is a squeezed vacuum state with enhanced quantum noise along in-phase amplitude. If a measurement result $\tilde{x}_1$ for x_1 is positive, the center of the wavepacket $\langle x_1 \rangle$ is shifted to a positive value and the variance $\langle \triangle x_1^2 \rangle$ decreases. The positive $\tilde{x}_1$ value pushes the center of the wavepacket $\langle x_2 \rangle$ of the other DOPO pulse to a negative value, if the mutual coupling is an anti-ferromagnetic type. In this way, a single measurement feedback process already produces a negative correlation between $\langle x_1 \rangle$ and $\langle x_2 \rangle$, which is the precursor for a final anti-ferromagnetic order in a macroscopic level.

Figure 7(a) shows the time evolution of $\langle x_1 \rangle$ and $\langle x_2 \rangle$ for the anti-ferromagnetically coupled two DOPOs when the external pump rate is linearly increased from below to above the threshold [11]. Immediately after the pump is switched on, the negative correlation between $\langle x_1 \rangle$ and $\langle x_2 \rangle$ is formed. However, because of the randomness of the measurement results and the noise associated with a signal loss process, the system points toward one ground state $|\uparrow\downarrow\rangle$ at one time but switches back to the other ground state $|\downarrow\uparrow\rangle$ at another time. This correlated random motion of $\langle x_1 \rangle$ and $\langle x_2 \rangle$ gradually increases its magnitude (degree of negative correlation) and eventually stops at a normalized time $\tau = 0.85$, where

one of the ground states $|\uparrow\downarrow\rangle$ is selected permanently. It is worth mentioning that the average photon number of the DOPO pulse is $\langle n \rangle \simeq 1$ at this bifurcation point below the oscillation threshold. Note that this numerical simulation result is a single trajectory post-selected by the final state $|\uparrow\downarrow\rangle$. If we do not post-select the quantum trajectory but simply ensemble average over many non-conditional trajectories, the average amplitudes are always $\langle x_1 \rangle = \langle x_2 \rangle = 0$ for all time. The contextual nature of a quantum measurement feedback system is clearly seen in this numerical simulation. That is, the present center positions $\langle x_1 \rangle$ and $\langle x_2 \rangle$ depend on the previous measurement results and also depend on future measurement results. The actual measurement results reported by homodyne detectors shown in Fig. 7(b) never manifest such a correlated random motion due to a large measurement error. Only at a later time $\tau \geq 0.9$, the correlated symmetry breaking indeed happens at a macroscopic level and we know the specific ground state $|\uparrow\downarrow\rangle$ was indeed selected by the system.
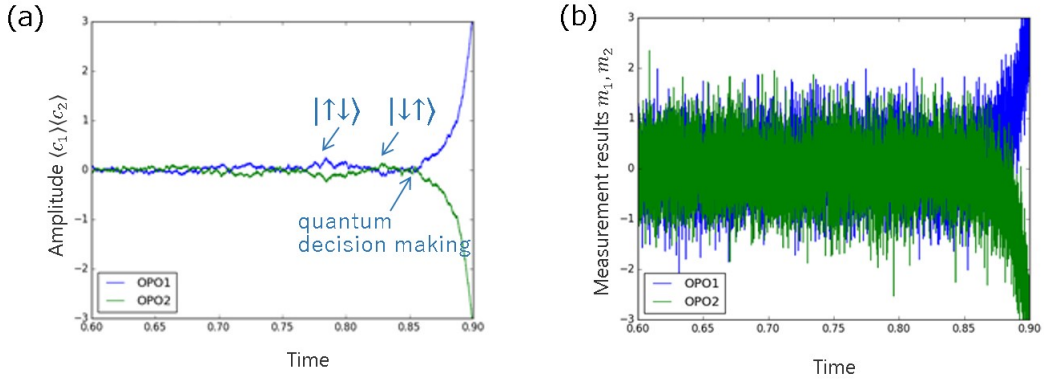

FIG. 7: (a) Time-evolution of the average in-phase amplitudes $\langle x_1 \rangle$ and $\langle x_2 \rangle$ of the anti-ferromagnetically coupled two DOPO pulses. (b) Time-evolution of the actual measurement results reported by homodyne detectors [11].

6.3 Quantum parallel search, correlated symmetry breaking and bosonic final state stimulation

In this section we will discuss how the correlation existing in the quantum noise produced among the DOPO states enhances the success probability to find a right solution. Let us consider a one-dimensional ring consisting of $N = 16$ DOPOs, in which neighboring DOPOs are coupled via anti-ferromagnetic interactions, i.e. $J_{ij} < 0$.

In Fig. 8 we plot the success probability of finding one of the two degenerate ground states $|\uparrow\downarrow\uparrow\downarrow \cdots \downarrow\rangle$ and $|\downarrow\uparrow\downarrow\uparrow \cdots \uparrow\rangle$. Here, we post-selected the trajectories conditioned by

8

the specific final result $|\uparrow\downarrow\uparrow\downarrow\cdots\downarrow\rangle$ [12]. The success probability by a brute force search is $P_0 = 1/2^{16} \sim 3 \times 10^{-5}$. Immediately after the pump is switched on at a normalized time $\tau = 0$, the probabilities of finding two degenerate ground states jump up and monotonically increase until the selected ground state $|\uparrow\downarrow\cdots\downarrow\rangle$ and the unselected ground state $|\downarrow\uparrow\cdots\uparrow\rangle$ depart at $\tau \simeq 70$. In this early stage of computation ($0 \leq \tau \leq 70$), the average photon number per DOPO pulse is less than one, so that the enhanced success probability is realized not by the coherent mean field but by the formation of the negative quantum noise correlation in x_j and x_{j+1}. This is a "quantum parallel search" mode.

At a time of $\tau \simeq 70$, the trajectories of the two degenerate ground states depart, i.e. the final state $|\uparrow\downarrow\cdots\downarrow\rangle$ continues to build-up but the unselected state $|\downarrow\uparrow\cdots\uparrow\rangle$ starts to diminish. This is a "correlated spontaneous symmetry breaking" mode, which is an educated decision making process driven by the internal noise.
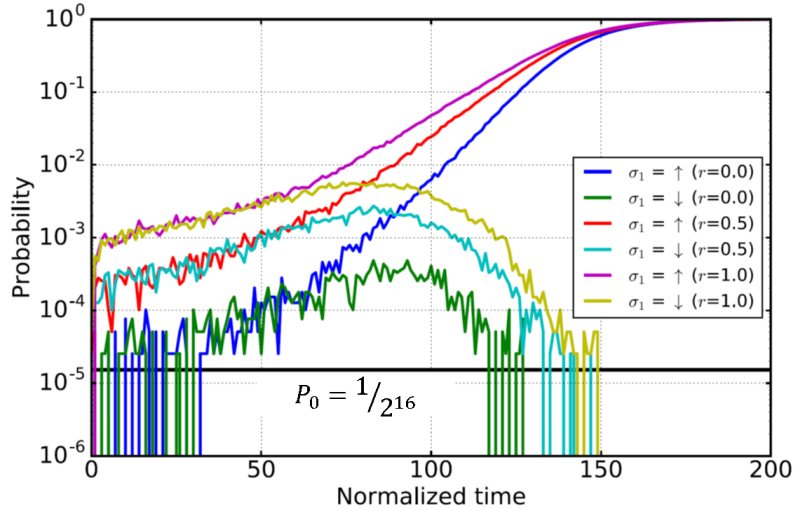

FIG. 8: Probabilities of finding the selected ground state $|\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\rangle$ ($\sigma_1 = \uparrow$) and the unselected ground state $|\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\rangle$ ($\sigma_1 = \downarrow$) for various squeezing parameters r of the input field to the open port of the output coupler. The system performs quantum search at $0 < \tau < 70$ by forming quantum noise correlation and correlated spontaneous symmetry breaking makes a decision at $\tau \simeq 70$. Finally, the amplitude of the DOPO field is amplified to a classical level to measure the result [12].

Subsequent exponential increase in the success probability is achieved by the single-mode oscillation at the selected ground state, which is a consequence of "bosonic final state stimulation". The average photon number increases rapidly to far above the threshold value $\langle n \rangle = 1/g^2$, where g is the saturation parameter [3, 6] . The quantum-to-classical transition is completed in this way simply by increasing the pump rate.

The quantum search at a time interval, $0 < \tau < 70$, is a crucial step for the QNN to

pick-up a correction solution out of $2^{16} \sim 3 \times 10^4$ candidates. Because of the noise correlation formed during this time interval, the probability of picking-up a wrong solution is efficiently suppressed. If we inject a strong noise field, such as a thermal field, into the open port of the out-coupler, the crucial noise correlation formed at below the threshold is easily buried in the external noise field. Numerical simulation results shown in Fig. 17 in Chapter I confirm this point. When $n_{th} \ll 1$ (quantum limit), the maximum success probability is achieved at just above the threshold. On the other hand, when $n_{th} \gg 1$ (classical limit), the maximum success probability is realized at well above the threshold. This is because only mean filed produced at well above the threshold can couple effectively neighboring DOPOs in the classical limit. However, such a strong mean field formed at well above the threshold necessarily produces a high potential barrier to stabilize the DOPO phase to either 0-phase or π-phase, so that the machine can be easily trapped by local minima.

6.4 MAX-CUT problems

One of the well-known combinatorial optimization problems is a maximum cut (MAX-CUT) problem on a graph, which is essentially equivalent to the Ising model without a dc field in statistical mechanics [14]. The task is to find the largest cut in a given graph $G = (V, E)$, where the number of edges at the boundary of a partition of vertices into two subsets is maximized. The size of the cut is defined as the total weight of edges separated by the cut, i.e., the edges which have each end point in the different sides of the cut. This objective function, which we hope to maximize, can be written as

$$\text{CUT}(x) = \sum_{1 \leq i \leq j \leq n} w_{ij} \frac{1 - x_i x_j}{2}, \tag{2}$$

where the graph order $n = |V|$ is the number of vertices, w_{ij} is the weight of the edge $(i, j) \in E$ and $x_i = \pm 1$ is a binary value indicating which side of the cut the vertex $i \in V$ belongs to.

To implement the above MAX-CUT problem on optical systems, the network of injection-locked lasers [15] and the network of degenerate optical parametric oscillators (DOPOs) [16] were proposed and named "coherent Ising machines". Such coherent Ising machines can be interpreted as a physical device to solve combinatorial optimization problems using

recurrently updated neurons with nonlinear activation function. From this point of view, there have been related approaches which employ mathematical models of the neurons (e.g., [17, 18]) and their networks (e.g., [19]). Hopfield developed the optimization algorithm in the discrete valued neural network model [20]. Later, Hopfield and Tank extended it to the continuous valued neural network to improve the performance and applied it to the combinatorial optimization problems [21, 22]. Simulated annealing (SA) is proposed to improve the performance of the discrete valued neural network approach by adding gradually decreasing thermal noise [23]. In the same spirit the continuous valued neural network can be supplemented by external noise injection, which can be described by the classical Langevin dynamics (CLD) [15, 16].

Here, we compare the relative performance of the coherent Ising machines against a family of those classical neural network approaches [13].

6.5 Coherent Ising machines vs. classical neural networks

6.5.1 Coherent Ising machines (CIM)

We intend to solve MAX-CUT problems by mapping the cost function (Eq.(2)) to the energy of an Ising spin system. Since the MAX-CUT problem has a binary variables $x_i \in \{1, -1\}$ for each vertex state, we can use a bistable DOPO state (0-phase or π-phase) as the output results of computation, while an analog nature of the degenerate optical parametric amplifier (DOPA) at below the threshold can be used as the solution search mechanism of computation.

The DOPO is operated near the oscillation threshold by increasing the external pump rate from below to above the threshold. The DOPO is initially biased at below the threshold, where all phase configuration is prepared in a superposition state and the quantum parallel search is implemented [3, 12]. Then, the external pump rate is gradually increased and once the whole network reaches the oscillation threshold, it selects a particular phase configuration which corresponds to the near-optimal solution of the original MAX-CUT problem.

The dynamics of the CIM can be simulated by the quantum master equation. Instead of numerically integrating the master equation for the DOPO density operator, we can expand the density operator by the quasi-probability function in the phase space. One

quasi-probability function for this purpose is the positive $P(\alpha, \beta)$ representation in terms of the off-diagonal coherent state expansion, $|\alpha\rangle \langle\beta|$ [3]. The Fokker-Planck equation for $P(\alpha, \beta)$ is derived from the master equation and then the c-number stochastic differential equations for α and β are obtained using the Ito calculus (see [3] for detail). Another quasi-probability function used for this purpose is the truncated Wigner representation $W(\alpha)$. The corresponding c-number stochastic differential equations are derived in [12].

6.5.2 Classical neural networks

We describe in this section the four classical neural network models that can be used as algorithms in modern digital computers to solve the same MAX-CUT problems. Those four different kinds of classical neural network models are summarized in Table I.

TABLE I: Classical neural-network approaches for MAX-CUT problems [13].

	Deterministic	Stochastic
Binary	Hopfield Network (HN)	Simulated Annealing (SA)
Continuous	Hopfield-Tank Neural Network (HTNN)	Classical Langevin Dynamics (CLD)

A. Hopfield Network (HN)

J. J. Hopfield proposed a classical neural network model for solving combinatorial optimization problems in his 1982 paper [20, 21], which is referred to as the Hopfield network (HN). The neuron in this model has the discrete output values $\sigma_i = \pm 1$ with a simple majority voting update rule:

$$\sigma_i \leftarrow \text{sgn}\left(\sum_{j=1}^{n} J_{ij} f(\sigma_j)\right), \tag{3}$$

which should be executed asynchronously, that is, one neuron is updated at a time. The spin index i is selected randomly in the original paper but we can derandomize the original model to enhance the speed, i.e., the spin indices from $i = 1$ to $i = n$ are updated sequentially. Simultaneous updates introduce the undesired instability or periodic solution into the system. Since the update is local and deterministic, the system will converge into a nearest local minimum, which is determined by the initial state. Note that the model is originally proposed with $\{0, 1\}$-binary neurons, but we use equivalent $\{+1, -1\}$-valued neurons here.

B. Simulated Annealing (SA)

While the HN will often get stacked at poor local minima, Kirkpatrick, et al. introduced a stochastic spin update strategy in the so-called simulated annealing (SA) algorithm to mimic thermal annealing [23]. The probability of stochastic spin flip is governed by the Boltzmann factor in the Metropolis-Hastings procedure. If the energy difference by the ith spin flip,

$$\triangle E_i = 2x_i \sum_{j=1}^{n} J_{ij}x_j \tag{4}$$

is negative, such a spin flip event is always accepted. If such a spin flip makes the total energy increased, namely $\triangle E_i > 0$, such an event is still accepted by the probability: $P = \exp\left(-\triangle E_i/T\right)$. The spin index i is selected randomly, while the effective temperature T is gradually decreased.

C. Hopfield-Tank Neural Network (HTNN)

Hopfield and Tank proposed another neural network approach using an analog valued neuron $x_i \in [-1, 1]$ and continuous-time evolution, which is referred to as the Hopfield-Tank neural network (HTNN) [22]. The time evolution of the HTNN is described by ordinary differential equations (ODE):

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -\alpha x_i + \beta \sum_{j=1}^{n} J_{ij}f\left(x_j\right), \tag{5}$$

where $f\left(x\right)$ is a nonlinear sigmoid function shown by Fig.1 of Chapter I. In this numerical study, $\tanh\left(x\right)$ is used as $f\left(x\right)$. In the extremely high linear gain limit, i.e., when the slope of the sigmoid function around 0 is steep, this HTNN model becomes close to the HN model described above. The neuron decay rate α and the synaptic connection strength β must be chosen optimally to achieve the best performance for the given MAX-CUT problems, as is done in later discussions. The numerical integration of Eq.(5) is performed by the Euler method with the sufficiently small discrete time step $\triangle t$.

D. Classical Langevin Dynamics (CLD)

A stochastic version of HTNN is formulated using the physical model of classical oscillator network (Eq.(6)). In this model, an analog valued neuron $x_j \in [-1, 1]$ obeys the classical Langevin equation:

$$\mathrm{d}x_i = \left(-1 + p - x_i^2\right) x_i \mathrm{d}t + \beta \sum_{j \neq i} J_{ij} x_j + g\sqrt{x_i^2 + {}^{1}\!/_{2}}\,\mathrm{d}w_{x_i}, \tag{6}$$

where a time t is normalized by signal decay rate, p is a normalized pump rate which is linearly increased from below ($p < 1$) to above the oscillation threshold ($p > 1$), g is the saturation parameter, dw_{x_i} is the Langevin noise term [16]. β is the injection parameter. g and β can be optimized to realize the best performance.

6.5.3 Implementation of classical neural network algorithms

Here we describe the computer hardware configuration employed to implement the four classical neural network algorithms, which will be used in the benchmark study. Note that all codes are implemented with c++ ‡.

A. CPU (for SA and HN)

SA and HN are iterative updating algorithms for discrete variables. We can achieve CPU implementation efficiently by SIMD bitwise operations in parallel §. In this study, we mainly used Intel Xeon E3-1225 v3 @ 3.2 GHz (Haswell architecture shipped in 2013). The performance of SA is slightly improved from the older processor (Intel Xeon X5650 @ 2.67 GHz, Westmere architecture shipped in 2010) [13]. We did not use any accelerators for HN and SA in this study since it is already parallelized by SIMD operations in CPU and the cache hit rate is as high as 98.8%.

‡ We used Ubuntu 16.04.4 with GCC 5.4.0 (CPU) and CentOS 7.1.1503 with GCC 4.8.3 (PEZY-SC)

§ The code is available here. https://github.com/haribara/SA-complete-graph

B. MIMD many core processor PEZY-SC (for HTNN and CLD)

Since HTNN and CLD are based on ordinary differential equations (a continuous-valued/continuous-time system) and require floating-point arithmetic, it is better to parallelize by accelerators. We used a MIMD many core processor PEZY-SC @ 733 MHz with 1024 cores and 8192 threads on a chip (the architecture is shown in Fig. 9), which is set in "Shoubu" supercomputer at Riken (Japan). We parallelized matrix-vector multiplication and neuron updates in 8192-thread parallel. The coupling matrix is efficiently stored as a 1-bit matrix (since $J_{ij} = \pm 1$ has no empty entry) and neuronal states as floating points (32-bit float). Note that it was 1.4 times faster than storing matrix values in 32 bits.
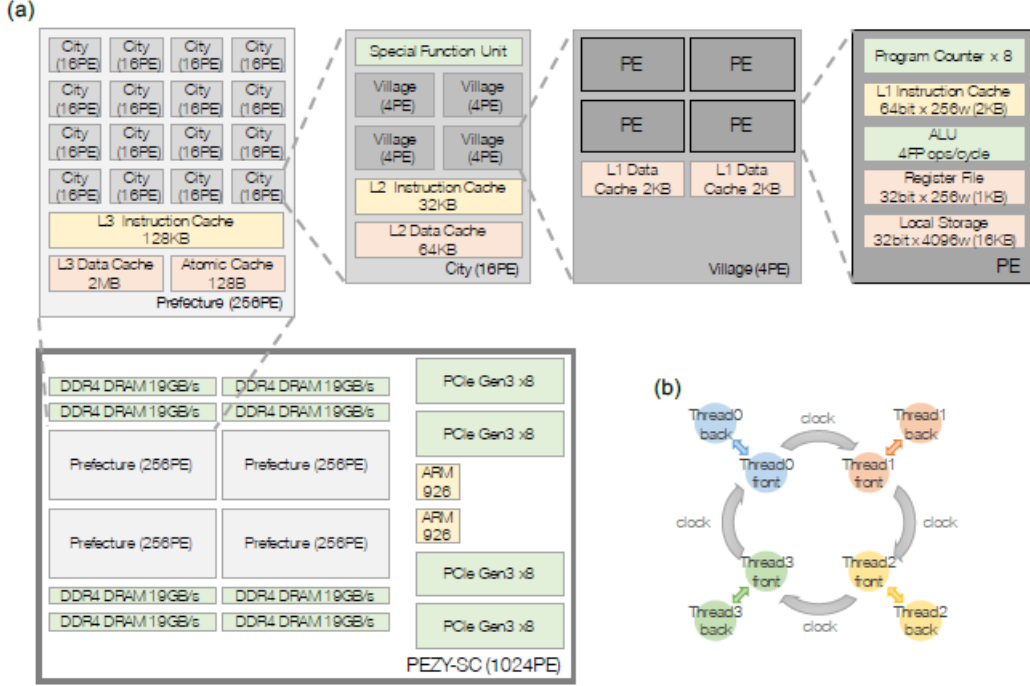
FIG. 9: (a) The hierarchical architecture of a PEZY-SC many core processor. There are 1024 processing elements (PE) packed in a single PEZY-SC chip. (b) Each PE core handles 8 threads independently [13].

6.6 Simulation results

We compared the performance of the four classical neural network algorithms (HN, SA, HTNN, and CLD) against the coherent Ising machine (CIM) by solving the MAX-CUT problems on a dense graph. The particular problem instance is a complete graph, in which all pairs of $N = 2000$ vertices are connected and edges are weighted by $\{+1, -1\}$ in uniform distribution (the identical instance as in Ref. [10]). Figure 10 shows the performance comparison, while the actual computation time to target and the hardware configurations are summarized in Table II.

We ran 100 different trials for the same problem instance (except for CIM, which consists of 26 experimental trials). Each solid line in the Fig. 10 indicates the ensemble average of all trials, while the lower and upper shaded lines indicate the best and worst case values, respectively. Here, the parameters for SA and HTNN are optimized to achieve the shortest computation time to the target which is obtained by the GW-SDP relaxation algorithm [24].

The computation time to the GW-SDP-produced target is shorter in the order of CIM, HN, SA, HTNN, CLD on the given problem instance. The data from CIM are noisy due to unavoidable experimental noise, but the CIM can find better solutions than the target in
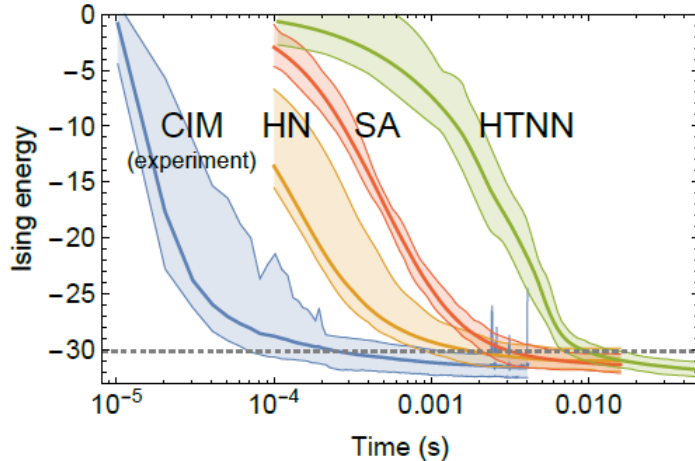
15

FIG. 10: Energy descent when solving a $\{+1,-1\}$-weighted $N = 2000$ complete graph [13]. Each thick line is the ensemble average of 100 trials (except for CIM experiment, which consists of 26 trials), while the lower and upper shaded error bars show the best and worst envelopes for each computational model. The gray dotted line is the target energy($= -60278/n$) which is obtained by the SDP relaxation algorithm [24].

TABLE II: Time to target and hardware configurations. The best (shortest) time to reach the target value and the ensemble averaged time to cross the target value are listed. Note that the target value is obtained by the GW-SDP relaxation algorithm which has performance guaranty to the 87% of the optimal value [13].

	Best (ms)	Average (ms)	Hardware
CIM (experimental)	0.071	0.264	DOPO + FPGA
HN	0.924	1.84	CPU
SA	2.10	3.20	CPU
HTNN	7.04	9.67	PEZY-SC
CLD	100	120	PEZY-SC
CIM (virtual)	10^5		PEZY-SC

all 26 trials. HN is faster than SA since HN can be regarded as a deterministic version of SA. Note that in the worst case, HN cannot reach the target (it fails 3 times in 100 trials as it can be seen partly in the worst case in Fig. 10). It can be understood that HTNN and CLD perform much slower than HN and SA, since it solves ODEs which deals with the analog variables. However, HTNN and CLD achieve a lower energy than SA in Fig. 10 but the performance of SA heavily depends on temperature scheduling. We optimized the temperature scheduling to reach the target in a shortest time, but slower scheduling ends up with a lower energy in general.

6.7 Discussion

In this section, we will add the two discussions to justify the above results:

- Validity for the hardware selection.

- Optimization for PEZY-SC implementation for HTNN and CLD.

6.7.1 Validity for the hardware selection

HTNN and CLD are apparently efficient on PEZY-SC than on CPU, which is shown in Fig 11. On the other hand, we did not use any accelerator for HN and SA. This is because we do not expect significant speed-up since they are already parallelized by SIMD operations in CPU and the cache hit rate is as high as 98.8% (measured by perf command in Linux). Generally, asynchronous update in HN and SA seems to be not suitable for parallel implementation.
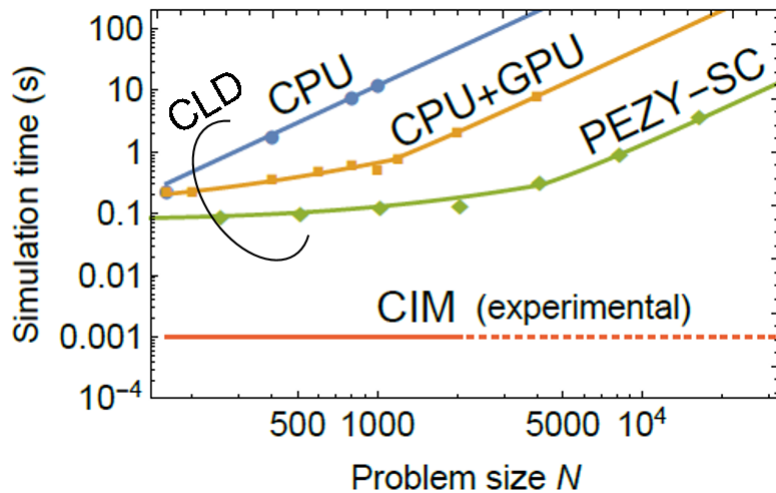

FIG. 11: Simulation time for the classical Langevin dynamics (200 round trips on complete graphs) [13].

6.7.2 Optimization for PEZY-SC implementation for HTNN and CLD

We tried to optimize the implementation by storing matrix data efficiently. Since the given adjacency matrix has only the 1-bit entry ($J_{ij} = \pm 1$), we packed each value in 1 bit. This contributes 1.4 times speed-up than having a 32-bit float matrix. However, putting

17

the data in local memory does not contribute to significant speed-up since its bottleneck in computation is not in memory transfer. There is a possibility of speed-up if we replace the multiplication by the selector, but it is possible to scale out for parallel distributed processing by using multiple PEZY-SC chips in "Shoubu" supercomputer, especially when the problem size is larger.

6.8 Conclusion

In this Chapter we compared the performance of the coherent Ising machine implemented on the network of DOPOs and FPGAs against the family of classical neural-network-based algorithms: HN, SA, HTNN, and CLD. To accelerate the performance of the classical neural networks, HN and SA are implemented on CPU with bit operations and HTNN and CLD are implemented on a many core processor PEZY-SC. It is shown that the CIM can achieve faster computational time than HN (13.0 times for the best case and 6.97 times for the average), SA (29.6 times for the best case and 12.1 times for the average), HTNN (99.2 times for the best case and 36.7 times for the average) and CLD (1400 times for the best case and 500 times for the average).

[1] F. Barahona, J. Phys. Math. Gen. 15, 3241 (1982).

[2] A. Lucas, Front. Phys. 2, 5 (2014).

[3] K.Takata et.al., Phys. Rev. A 92, 043821 (2015).

[4] L-M. Duan et al., Phys. Rev. Lett. 84, 2722 (2000).

[5] H. Ollivier and W. H. Zurek, Phys. Rev. Lett. 88, 017901 (2001).

[6] J. S. Bell, Rev. Mod. Phys. 38, 447 (1966).

[7] S. Kochan and E. P. Specker, J. Math. Mech. 17, 59 (1967).

[8] M. D. Menmin, Rev. Mod. Phys. 65, 803 (1993).

[9] P. McMahon et al., Science 354, 614 (2016).

[10] T. Inagaki et al., Science 354, 603 (2016).

[11] T. Shoji et al., Phys. Rev. A 96, 053833 (2017).

[12] D. Maruo et.al., Phys. Scr. 91, 083010 (2016).

[13] Y. Haribara et al., Quantum Sci. Tech. 2, 044002 (2017).

[14] M. R. Garey and D. S. Johnson, Computers and intractability 29 (W. H. Freeman, New York, 2002).

[15] S. Utsunomiya et al., Optics Express 19 18091 (2011).

[16] Z. Wang et al., Phys. Rev. A 88 063853 (2013).

[17] W. S. McCulloch and W. Pitts, The Bulletin of Mathematical Biophysics 5, 115 (1943).

[18] A. L. Hodgkin and A. F. Huxley, The Journal of Physiology 117, 500 (1952).

[19] K. Fukushima K 1980 Biological Cybernetics 36 ,193 (1980).

[20] J. J. Hopfield, Proceedings of the National Academy of Sciences 79, 2554 (1982).

[21] J. J. Hopfield, Proceedings of the National Academy of Sciences 81, 3088 (1984).

[22] J. J. Hopfield and D. W. Tank, Science 233, 625 (1986).

[23] S. Kirkpatrick et al., Science 220, 671 (1983).

[24] M. X. Goemans and D. P. Williamson, Journal of the ACM (JACM) 42, 1115 (1995).

Written by Y. Yamamoto

version 1