

Dissertation
博士論文 (要約)

Realization and Evaluation of
Measurement Feedback Coherent Ising Machines
for Combinatorial Optimization Problems

（ 組合せ最適化問題のための
測定フィードバック型コヒーレント・イジングマシン
の実現と評価 ）

Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo

Yoshitaka Haribara
針原佳貴

December 2017

Abstract

The growth of computer performance substantiated our sophisticated society with a tremendous amount of data and numerical optimizations everywhere. Combinatorial optimization problems appear in many diverse fields such as route planning for logistics, resource allocation in wireless communication, and machine learning. Many research efforts have been made to define the hardness of problems and construct efficient algorithms. In recent trends in semiconductor technologies, Moore's law is slowing down mainly due to the limitation of micro-fabrication heat dissipation and communication bottleneck problems on a chip. More efforts to boost the processor performance directed to parallelized architectures including GPU, other multi/many-core processors, and neuromorphic hardware. In fact, many computationally heavy tasks such as deep learning and computational science run on GPU clusters or special purpose processors. Given the fact, recently it is attracting more attention to build a computer based on physical dynamics which cannot be efficiently simulatable, including quantum computers and quantum annealers, which are distinguished from von-Neumann computer architecture. Since 2011, coherent Ising machine (CIM), which are the optics-based special purpose computers, have been proposed to exploit a rapid physical convergence time for accelerating the solution search in hard optimization problems. The system was initially proposed with laser followed by non-classical optical parametric oscillator (OPO) implementations, where the proof of concept was demonstrated with small sized problems with $N \leq 16$.

However, the previously proposed system lacks its scalability. In the N OPO coupling with $N - 1$ optical delay lines need to be stabilized simultaneously, which is unrealistic in the current experimental system. It is necessary to scale up to 1-2 orders of magnitude larger problems to handle reasonable size of problems.

Hence, the measurement-feedback CIM is proposed in this dissertation. It only requires a single feedback circuit with optical measurement and coupling calculation, which can implement arbitrary topology of graphs, and scalable to the practical size in real-world problems. In this dissertation, the physical and mathematical models of the measurement-feedback CIM are presented. Its stochastic differential equations via truncated-Wigner representation are formulated. The models predict the performance of the scalable CIM and shows promising performance on dense graphs. The comparative approaches for the baseline of performance are calculated with the approximation algorithms based on semidefinite programming (SDP) and established heuristics – simulated annealing (SA).

The rapid computation is confirmed with experimental implementations. It is reported that the more than 10 times speedup is observed compared to SA implemented on a single thread of single core of CPU in the case of $N = 2000$. Since the idea of CIM originates from the quantum information science, it is necessary to be compared with the current state-of-the-art computer performance. This paper has shown the possibility of experimental implementation and relative speedup, which will definitely open the future of alternative physical computer architectures.

Acknowledgements

I could never finished this dissertation without many supports from people around. Behind the work, there are plenty of collaborations, discussions, and shared goals with them. Let me introduce my collaborators.

First of all, I would like to give the greatest gratitude to my principal advisor, Professor Kazuyuki Aihara. I learned a lot from his advice on research, his mind for bold direction, and his laboratory members in variety of fields. He is a great educator and his continuous encouragement was the key to my research success.

Next, I would like to appreciate the contribution of my dissertation committee in the department of mathematical informatics: Professor Satoru Iwata, Professor Yoshito Hirata, Professor Takaaki Ohnishi, and Professor Taiji Suzuki. Their careful reading and critical advise on the dissertation improved its quality by an order of magnitude.

I would also like to thank Professor Yoshihisa Yamamoto for inspiring the theme when I entered the graduate school. Not only how to develop an idea, I learned intuition in physics and an essential point of view in science.

I greatly respect the experimental effort of my collaborators. This dissertation would never be satisfactory without their successful results. I am very happy to work with such talented people. I want to thank Peter L. McMahon and Alireza Marandi in E. L. Ginzton Laboratory of Stanford University, and Takahiro Inagaki, Hiroki Takesue, Toshimori Honjo in NTT Basic Research Laboratories.

Insightful discussions about theory and simulation built a foundation of my research topic. Professor Shoko Utsunomiya, Shuhei Tamate, Kenta Takata, Daiki Maruo in NII, and Ryan Hamerly and Zhe Wang in E. L. Ginzton Laboratory of Stanford University gave me a chance to have wide range of discussions among quantum optics and Ising machines. Professor Ken-ichi Kawarabayashi, Tomohiro Sonobe supported my research in combinatorial optimization, graph theory, and software algorithms. Hitoshi Ishikawa in PEZY Computing helped me for porting the simulation and solver used in benchmark to many-core processors.

When designing a practical implementation of measurement-feedback circuit, I was helped by Professor Kyo Inoue and Professor Koji Igarashi in Osaka University. Other collaborators helped the OPO experiments: Carsten Langrock, Professor Robert L. Byer, Professor M. M. Fejer, Professor Hideo Mabuchi in E. L. Ginzton Laboratory of Stanford

University, and Takeshi Umeki, Koji Enbutsu, Osamu Tadanaga, Hirokazu Takenouchi in NTT Device Technology Laboratories.

The laboratory members are necessary from various points of view such as research discussion, university life, and thesis revision. I would like to thank Timothée Leleu Hiromasa Sakaguchi, Kyohei Kamiyama, Hiroyuki Yasuda, Atsushi Yamamura, Taime Shoji, Makito Oku, and Masashi Kajita.

Around the research project, many discussions reshaped and refined my ideas. I would like to thank Yasuo Saito in NCOS, Professor Shunsuke Saruwatari, Takefumi Miyoshi, Ichiro Ogawa in sdlab for FPGA architecture and design. I also appreciate discussions with Kai Yan, Ryoji Miyazaki, Yutaka Takeda in NII, Professor Yuichi Katori, Professor Hidetoshi Nishimori, Professor Toru Aonishi, Professor Kazushi Mimura, Professor Synge Todo.

Professor Kei Hiraki, Professor Hal Tasaki, Professor Satoshi Matsuoka, Professor Keisuke Fujii, and Professor Keiji Matsumoto left many useful comments on the web. I appreciate their discussions on Twitter.

I also thank the laboratory staffs for kind helping: Sanae Kimoto, Chiko Tsukamoto, Yuka Suzuki in Aihara group. Yukiko Sato, Michiru Kunisaki in JST, Chiemi Kubo, Risa Aihara, Yoko Shioda in NII, Yurika Peterman, Rieko Sasaki in Stanford also helped me.

My work is supported by the Impulsing Paradigm Change through Disruptive Technologies (ImPACT) Program of the Cabinet Office of Japan.

I also appreciate the undergraduate seminar program advised by Masaaki Wada in Department of Mathematics in Osaka University, which led me to work on computation with wave dynamics. Finally, I want to thank my parents, sister, grandparents, and wife for continuous support.

Contents

Acknowledgements	iii
Chapter 1 Introduction	3
1.1 Trends in Computer Architecture	3
1.2 NP-hard Combinatorial Optimization Problem	4
1.2.1 Combinatorial Optimization Problem	4
1.2.2 Computational Complexity Theory	5
1.2.3 Motivation to focus on NP-hard problem	5
1.3 Alternative Approaches to Computation	6
1.3.1 Quantum Computing	6
1.4 Ising Machine	8
1.4.1 Target Problem	8
1.4.2 Quantum Annealing	8
1.4.3 Other Approaches	10
1.4.4 Applications	10
1.5 Coherent Computing	10
1.6 Contribution	11
1.7 Dissertation Structure	11
Chapter 2 Maximum Cut Problem	13
2.1 Graph	13
2.1.1 Typical Topology of Graphs	13
2.2 Problem Definition of MAX-CUT	14
2.3 Approximation Algorithm	15
2.3.1 Goemans-Williamson Algorithm (GW)	15
2.3.2 Greedy algorithm of Sahni-Gonzalez (SG) and its Variants	17
2.3.3 Other Approximation Algorithms	17
2.3.4 Efficiently Solvable Classes	17
2.4 Metaheuristic	18
2.4.1 Derandomized Hopfield Network (HN)	18
2.4.2 Simulated Annealing (SA)	19
2.4.3 Hopfield-Tank Neural Network (HTNN)	20

2.4.4	Breakout Local Search (BLS)	20
2.4.5	Other Metaheuristics	21
2.4.6	Off-the-Shelf Optimizer	21
2.5	Hardness of Approximation	21
2.5.1	Interactive Proof	21
2.5.2	Probabilistically Checkable Proof (PCP) Theorem	22
2.5.3	Unique Games Conjecture (UGC)	22
Chapter 3	Degenerate OPO	23
3.1	Basics of Degenerate Optical Parametric Oscillators	23
3.2	System Hamiltonian	24
3.3	Master Equation	24
3.4	Fokker-Planck Equation	25
3.4.1	Phase-Space Methods	26
3.4.2	In the P Representation	29
3.4.3	In the Positive P Representation	29
3.4.4	In the Truncated Wigner Representation	30
3.5	Langevin Equation	31
3.6	Typical Properties of DOPO	32
3.6.1	Pump-Signal Relation	32
3.6.2	Pitchfork Bifurcation	33
3.6.3	Analytical Solution without Noise	33
3.6.4	Steady State Distribution	33
3.7	DOPO in the Optical Fiber	34
Chapter 4	Coherent Ising Machine	37
4.1	Overview	37
4.2	Quantum Formulation of CIM	39
4.2.1	Positive P representation	40
4.2.2	Truncated Wigner Representation	41
4.2.3	Measurement Effect in Measurement Feedback CIM	42
4.2.4	Computational Principle	44
4.3	Discrete Feedback Simulation with Ring Cavity	44
4.4	FPGA Design for Feedback Circuit	45
Chapter 5	Numerical Simulation	49
5.1	Implementation of Numerical Analysis	49
5.1.1	Parameter Optimization	49
5.2	Single Run	50
5.3	Performances on Benchmark Instances	54
5.3.1	Parameter Sweep	54

5.3.2	Performance on G-set	54
5.3.3	Computation Time Scaling on Complete Graph	59
5.4	Summary	59
Chapter 6	Experimental Results and Validation	61
6.1	Setup: Machine with $n = 100$	61
6.2	Results	63
6.2.1	Single Run	63
6.2.2	Energy Distribution	63
6.2.3	Statistics with Cubic Graph	66
6.2.4	Finite Size Scaling with Random Graph	68
6.3	Summary	72
Chapter 7	Performance Evaluation with Larger Problems	73
7.1	Setup: Machine with $n = 2048$	73
7.2	Benchmark results against SA	74
7.2.1	CUT Histogram	74
7.3	Hardware Configurations for Comparison	74
7.3.1	CPU (for SA and HN)	74
7.3.2	MIMD Many-Core Processor (for HTNN)	76
7.4	Time To Target against Neural Network Based Heuristics	77
7.5	Summary	77
Chapter 8	Conclusion	81
8.1	Experimental Validity and Computational Performance	81
8.2	Possibility of Other Implementations	81
8.3	Practical Issues	82
8.4	Discussion	82
8.4.1	Related Quantum Nature	82
8.4.2	Further Discussion	83
8.5	Perspectives and Concluding Remarks	83
Appendix A	Application Examples	85
A.1	Boxed-Constrained Quadratic Programming (BCQP)	85
A.2	Truss Topology Optimization	85
A.2.1	Weight Minimization	85
A.2.2	Case Studies	86
A.3	Electronic Circuit Design	86
A.3.1	VLSI Circuit Design	86
A.3.2	Versatile Place and Route	86
A.4	Application to Various Problems	87
A.4.1	3-SAT	87
A.4.2	List of Problems	89

AppendixB	Multimode Effects in a DOPO Ising Machine	91
B.1	Equations for Signal Fields and Hermite Function Expansion . . .	91
B.2	Numerical Results	93
B.2.1	Success Rate on $n = 16$ Instances	93
B.2.2	Statistics	93
B.2.3	Scaling	94
B.2.4	On Large Instances	94
AppendixC	Comments on the Processor Implementations	97
C.1	CPU Architecture Generation	97
C.2	Parallel Implementation on CPU	97
C.3	Cache Size	97
C.4	GPU Acceleration of Langevin Equations	99
AppendixD	Several Variants of Monte Carlo Methods	101
D.1	Order- N Monte Carlo	101
AppendixE	Extended Implementations	103
E.1	Temporal Bit Coding	103
E.1.1	Result	103
AppendixF	FPGA Implementation	105
F.1	FPGA Bit Resolution for Feedback Circuit	105
F.2	Optimal Bit Resolution	105
Bibliography		109

t

Chapter 1

Introduction

Since the invention of digital computers, its performance has dramatically increased, and many applications have changed the world. The improvement comes mainly from the two aspects, developing fast algorithms and improving hardware performance. Regarding the latter, recent data revealed different aspects of processor trends. This is primarily due to the limitation of energy consumption on the chip, hence an alternative architecture is required. In this chapter, we will outline recent trends and then overview the idea of alternative approaches.

1.1 Trends in Computer Architecture

Modern computers and architecture are beneficial to our lives. After the first operation of the electronic general-purpose digital computer ENIAC built by Eckert and Mauchly, von Neumann left his draft on the EDVAC in 1945. Since then, a single architecture known as the current von Neumann computer determined the processor design [1, 2]. It has a central processing unit (arithmetical and control part) and a memory, where the memory hierarchy design is basically employed to manage both access and memory capacity [1]. Though, as the processor performance grows, it came to be known as the von Neumann bottleneck, where the computational performance is limited by memory bandwidth.

The most famous empirical processor scaling will be Moore's law, predicting the number of transistor on a processor doubles in every 18 months [3]. The processor performance is growing in accordance with it, although, the growth rate slowed down around 2003 as shown in Figure 1.1 [1]. Hence, to compensate the ability of the single-core, recent trends of hardware architectures shifted to the multi-core and many-core systems including the graphics processing unit (GPU) as shown in Figure 1.1 [4, 5]. Recently, the general purpose GPU (GPGPU) is used from scientific simulations such as the molecular dynamics simulation [6], visualizing CT scan images [7] and astrophysical many-body systems [8] to artificial intelligence (AI) and machine learning (ML) [9, 10], which exploit the accelerated matrix operation with parallel single-instruction-multiple-data (SIMD) operations on GPU.

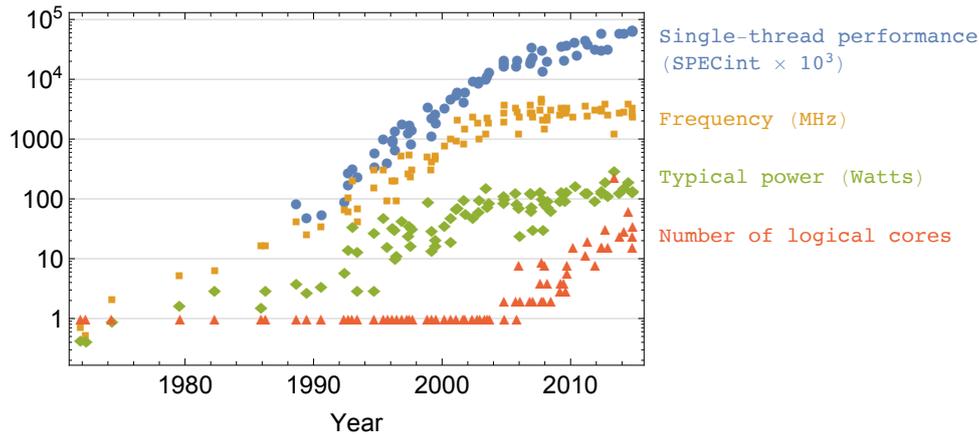


Figure 1.1. Relative performance of processors measured on the Standard Performance Evaluation Corporation (SPEC) benchmarks with integer arithmetic (SPECint) [14] is indicated with blue dots. It can be seen that the relative gain of performance is decreased to around 22% per year. It is mainly due to power limits and instruction-level parallelism [1]. In recent microprocessor trends, the number of cores on a single chip is exponentially increasing as indicated by the bottom red triangles. Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten. New plot and data collected for 2010-2015 by K. Rupp [5].

In addition to general purpose processors, configurable logic can benefit from hardware acceleration especially when we have a high throughput, low latency or real-time demand. In such area, field programmable gate array (FPGA) is used to reduce the latency throughout from the data input to computation. While the FPGA is configurable with software, an application specific integrated circuit (ASIC) is also a choice when we need higher performance and energy efficiency, which is used in many fields including AI/ML [11, 12, 13]. Since hardware options are expanding, it is becoming more important to select and design the appropriate ones according to the problem to solve.

1.2 NP-hard Combinatorial Optimization Problem

1.2.1 Combinatorial Optimization Problem

A combinatorial optimization problem is everywhere in the real world. It is defined as a maximization or minimization problem on a discrete mathematical structure. For example, LSI circuit design in processor [15], sparse sensing in image processing [16], the frequency channel allocation in wireless communication [17], lead optimization in drug design [18], truss structure optimization in architecture [19], and more. Some of them belong to the difficult class of problems, to which no polynomial time algorithm is believed to exist. The difficulty usually originates from the combinatorial structure in a non-convex objective function.

1.2.2 Computational Complexity Theory

In the computational complexity theory, *computational problems* are classified according to the difficulty to solve the problems. Here in the computational problem, we are given an *input*, that is assumed to be a sequence of alphabets $\{0, 1\}^*$ without loss of generality, and return an *output*. Then the problem can be described by the property that the output should satisfy.

In a *decision problem*, given an input $x \in \{0, 1\}^*$ and we are asked if the output is YES or NO. It can also be stated that a decision problem can be specified by finding the subset of input $L \subset \{0, 1\}^*$ in which the answer is YES. The subset of $\{0, 1\}^*$ is called a *language*. Hence each decision problem is specified by a language $L \subset \{0, 1\}^*$ and *vice versa*.

Complexity classes can be defined with the above language or using an abstract model of computer, Turing machine. Both definitions are equivalent. Turing machines *read* and *write* symbols written on an infinite length *tape* with a moving *head* depending on its internal state, which is defined as follows.

Definition 1.1 (Turing Machine [20]). Let A the set of alphabets with an empty letter. A Turing machine Φ is the function with a tape of length $N \in \mathbb{Z}^+$:

$$\Phi : \{0, \dots, N\} \times A \rightarrow \{-1, \dots, N\} \times A \times \{-1, 0, 1\}. \quad (1.1)$$

To estimate and define how much resources the computational problems require, especially computation time, they are classified as the *computational complexity*. Here we define some of these classes on Turing machine which are important in the following arguments. P is the class of decision problems which are solvable in polynomial time. NP is the class of decision problems where once we can find the solution, we can check the feasibility of it in polynomial time. We can define NP-hard class, to which all the NP problems can be reduced in polynomial time. The intersection of NP and NP-hard is called NP-complete.

The solution of a NP problem can be verified in polynomial time by a deterministic Turing machine. Here, we can immediately know that $P \subset NP$ by the definition of a Turing machine. The language can be redefined by properties of prover P and verifier V . This formulation is called *interactive proof*, which will be the building blocks in theoretical bounds and presented in Section 2.5.

1.2.3 Motivation to focus on NP-hard problem

It is mostly believed that $P \neq NP$. Hence, there exists problems which cannot be solved in polynomial time by conventional computers (we refer to deterministic Turing machines) in NP-hard. By definition, any problems in NP can be reduced to the problem which exists in NP-hard. If we could find the optimal solution of any NP-hard problems efficiently, then we can obtain solutions to any problems in NP with additional polynomial

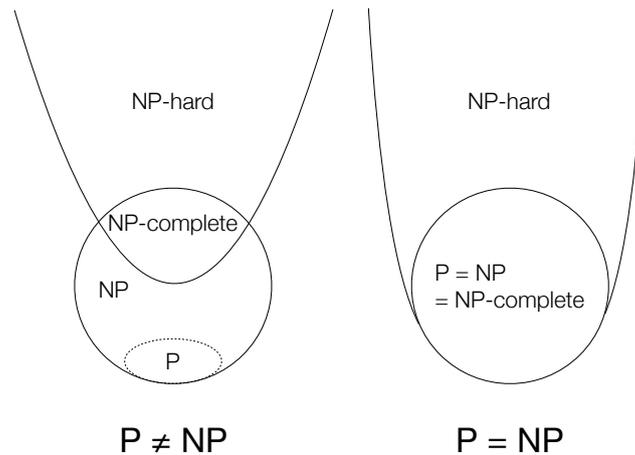


Figure 1.2. Relationship between the complexity classes. The left panel shows the case of $P \neq NP$. If $P = NP$, it becomes as the right panel.

time. This existence of polynomial-time mapping is the initial motivation to tackle NP-hard problems. Note that if we have an infinite precision analog computer, NP-hard (or precisely PSPACE) problems are efficiently solvable [21].

1.3 Alternative Approaches to Computation

Given the fact on the trends in the computer architecture, it would be reasonable to think of the architecture more deeply into the physical layer as redefining the way of computation. The point is to use the dynamics of physical devices which are not efficiently simulatable. Here, several approaches will be introduced, which exploit the physical property, especially for quantum mechanics.

1.3.1 Quantum Computing

While quantum computer is not directly related to the current research focus, we mention here to clarify our relative stance. The field of the quantum computer is recently attracting more and more attention both in the academia and industry [22, 23]. Quantum computers exploit entanglement in a quantum system to simulate processes in nature efficiently. There are known several *quantum algorithms* for intractable problems.

In the algorithm side, there have been known several famous successful examples, such as Deutsch-Jozsa [24] algorithm followed by Shor's prime factorization [25] and Grover's database search [26], which are proposed in the 1990s.

For example, the most used Rivest-Shamir-Adleman (RSA) cryptography system relies on the computational difficulty of RSA problem, which is solvable with integer factorization [27]. While the prime factorization is in class NP, Shor's algorithm solves the prime

factorization in polynomial time of the input. Note that there are several types of post-quantum cryptograph [28], which are secure against the attack using quantum computers, such as lattice-based cryptography [29, 30] and elliptic curve cryptography [31, 32].

More recently, the Harrow-Hassidim-Lloyd (HHL) algorithm for matrix inversion [33, 34], semidefinite programming [35] are proposed, which show exponential speed up. There are studies on quantum machine learning [36]. Note that the quantum computer is not expected to solve NP-hard problems since the quantum hard problems are distinct from the classical counterpart in the complexity theory.

Physical qubit implementations

Physical realization is recent hot topic since the successful implementation of error correcting superconducting qubits [37], originally invented by Nakamura, Pashkin, and Tsai [38, 22]. One of the near-term milestone in the experiment is making intermediate-scale quantum circuit around 50 qubits, which cannot be efficiently simulated by existing digital computers¹.

Quantum error correction and scalability

In practice, it needs quantum error correction towards large-scale *fault-tolerant* quantum computers, to implement quantum algorithms on assumed error-free qubits and gates [39]. The distillation when running Shor's algorithm takes most of (90% in 2048-bit and 75% in 4096-bit) resources of quantum computers. Then it will require 6,144 (application) + 66,564 (logical distillation) qubits, hence millions or billions of physical qubits and 1.81 days to solve 1024-bit factorization. In addition, the energy consumption is still high, so that the realization of the fault-tolerant quantum computer with enough number of qubits would be still far.

Hence, recently applications using bare qubits attract more attention such as simulating the electronic structure of a Fermionic system with quantum computers [40].

Quantum Simulation

The origin of the idea of using a quantum system as a computational resource is referred to Feynman's talk in 1981, where he mentioned the possibility of using a quantum system as a quantum simulator [41]. Quantum simulation is a method to directly simulate the difficult many-body problems in quantum physics or chemistry by a controllable quantum system, such as ultracold atoms in an optical lattice or quantum computers [42, 43, 44]. One of the focus is to find the ground state of the Fermi-Hubbard model, which is believed to be an important key of high-temperature superconductivity [45].

¹Google, Intel, IBM have recently announced 49-qubit, 49-qubit, and 50-qubit implementations based on superconducting circuits, respectively

1.4 Ising Machine

An Ising machine is a completely different approach but more reasonable in near-term practical implementation. It aims to exactly or approximately solve intractable problems using physical dynamics.

1.4.1 Target Problem

Ising problem

An Ising problem is to find the *ground state* of an Ising model of spin glass, which is studied in the field of statistical physics. In a (classical) Ising model, Ising spin $\sigma_i = \pm 1$ is associated with each site on the d -dimensional hypercubic lattice Λ . The hamiltonian of the system is defined on the collection of spins, *spin configuration* $\sigma = (\sigma_i)_{i \in \Lambda} \in \{1, -1\}^{|\Lambda|}$

$$H(\sigma) = - \sum_{i,j \in \Lambda} J_{ij} \sigma_i \sigma_j, \quad (1.2)$$

where $J_{ij} \in \mathbb{R}$ denotes Ising interactions. While the computational complexity of the Ising model is studied with lattice [46, 47], here we treat an equivalent problem in graph theory.

Maximum cut problem (MAX-CUT)

A maximum cut problem (MAX-CUT) on a graph is a binary quadratic programming without constraints. It is essentially equivalent to the Ising problem on the $\sigma_i = \pm 1$ variable, and also referred to unconstrained binary quadratic programming (UBQP) [48] or quadratic unconstrained binary optimization (QUBO) [49] on boolean binary variables $\{0, 1\}$. Details will be described in the next Chapter.

1.4.2 Quantum Annealing

Quantum annealing is recognized to be proposed by Kadowaki and Nishimori in 1998, which aimed to find the ground state in Ising model faster than the simulated annealing [50, 51]. Later, it is reinvented as the adiabatic quantum computation [52]. Recently, it attracts the big attention since experimental quantum annealers are implemented by D-Wave Systems Inc., initially with 128 qubits and currently with 2000 qubits. The quantum nature of the machine is investigated [53, 54] and recently it is believed to exploit the quantum tunneling as theory shows [55, 56, 57].

In the process of quantum annealing, an n -spin system is prepared with a superposition of up and down eigen states. They are represented by eigen values of σ_i^z (± 1) which are initialized with transverse field $\sum_{i=1}^n \sigma_i^x$. Here the Pauli matrices ²

$$\sigma_i^x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_i^z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

²They are related to the generator of the special unitary group of degree 2, SU(2).

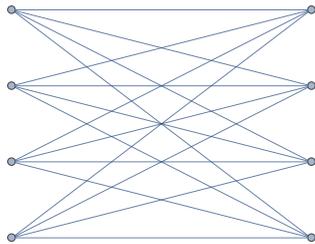


Figure 1.3. The component of a chimera graph is a complete bipartite graph $K_{4,4}$ defined in Section 2.3.4.

have eigen vectors $\{(1, 1)^\top/\sqrt{2}, (1, -1)^\top/\sqrt{2}\}$ and $\{|1\rangle := (1, 0)^\top, |-1\rangle := (0, 1)^\top\}$, respectively. The system keeps the ground states and reaches to the optimal solution while gradually releasing the field with $\Gamma(t) \geq 0$. The time dependent Hamiltonian is written as

$$\mathcal{H}_{\text{QA}} = - \sum_{i=1, i \neq j}^n J_{ij} \sigma_i^z \sigma_j^z - \Gamma(t) \sum_{i=0}^{n-1} \sigma_i^x. \quad (1.3)$$

One of popular methods to simulate the dynamics of quantum annealing is quantum Monte Carlo (QMC) simulation, which treats the classical expansion of (1.3). In the QMC algorithm, an imaginary dimension is introduced and spins are spread into that direction, which is called Suzuki-Trotter decomposition into $P \in \mathbb{N}$ slices [58]:

$$\mathcal{H}_{\text{QMC}} = - \sum_{k=1}^P \left(\sum_{i \neq j} J_{ij} s_i^k s_j^k + J^\perp \sum_{i=0}^{n-1} s_i^k s_i^{k+1} \right), \quad (1.4)$$

where new binary variables $s_i^k = \pm 1$ are introduced and the coupling between the slices is

$$J^\perp := - \frac{PT}{2} \log \tanh \frac{\Gamma}{PT} \quad (1.5)$$

with temperature T .

While the D-Wave machines implement QA and the performance is studied in against QMC, the shortage is its graph topology. Because of physical restriction, the qubits are coupled by the chimera structure, which consists of multiple complete bipartite graphs $K_{4,4}$ connected left side in the vertical and right side in the horizontal (see Figure 1.3). Since the chimera graph is very sparse, it needs $O(n^2)$ spins to map a complete graph of order n (or even an expander graph). This technique is called *minor embedding*, in which the graph minor of the chimera graph is considered and the desired graph topology is constructed. The order of the original chimera graph will shrink to the $O(\sqrt{n})$ by the graph minor operation in the worst case (complete graph or even sparse expander graph).

1.4.3 Other Approaches

Recently, several hardware designs of Ising problem-oriented chips are reported from Hitachi [59, 60] and Fujitsu [61, 62, 63]. They are based on the variant of SA, in which the energy evaluation and flip judgment process are implemented with hardware. It is worth mentioning that the quantum-oscillator-based optimization machine using Kerr nonlinearity is also proposed [64].

1.4.4 Applications

Not only the mathematical optimization problems, a lot of other efforts have been made to find a promising specific applications of such quantum computers, annealers, and Ising-type computers. For example: prime factorization by Ising model via Gröbner basis expansion [65], detecting the Higgs-boson-signal from background noise [66], optimizing the traffic flow [67]. Other examples related to our work are discussed in the Appendix A.

1.5 Coherent Computing

Since 2011, a series of research about the optical computer targeting to NP-hard combinatorial optimization problems, collectively referred to coherent Ising machine (CIM), has been activated. It was studied initially with spacially coupled laser, in which the binary Ising spin variable was represented by laser light polarization degree of freedom [68]. Later, the degenerate optical parametric oscillator (DOPO) was proposed to use [69]. Unlike the laser, which has phase-insensitive gain, DOPO has intrinsically bistable steady state in phase space. It contributes to the experimental stability and the demonstration of $n = 4$ spin network in a free-space cavity succeeded [70].

A major difference in the $n = 4$ DOPO experiment from the initial proposal was that one long optical cavity is designed to contain time-multiplexed pulsed DOPOs not to be spatially distributed to many cavities. This made the number of Ising pulses scalable followed by subsequent experiments of $n = 16$ in free-space [71] and $n \geq 10000$ in an optical fiber [72].

At this point, the scalability for the number of pulse itself is resolved, but the connectivity problem remains. The main difficulty of physically connecting pulses is the limitation of the topology. For example, the $n = 16$ system is implemented with a bow-tie ring cavity and 3 optical delay lines. Suppose t_{RT} is the cavity round trip time, then we can make a n -pulse cavity by pumping the gain medium with pulse repetition interval of t_{RT}/n . Then n equally spaced pulses are generated. We can put optical delay lines to connect pulses. Here, the length of the delay lines should be an integer multiple of t_{RT}/n . The delay line is shared by each pulse, so $n - 1$ delay lines can implement all-to-all connections between n pulses.

Experimentally, it is hard to stabilize all of the $n - 1$ delay lines. So, the setting in Reference [71] only has 3 delay lines. It implements 1, 8, and 15 pulses delay in $n = 16$

system hence cubic graph. It seems that the number of delay line is enough as far as we are interested in sparse graphs. But it is false. This is because, even if we implement random sparse graphs, the number of delay lines required is at least the maximum degree of the vertices in the graph. The delay lines in the physical system can only implement regular structure, hence a random sparse graph requires much more delay lines than the average graph degree.

1.6 Contribution

Implementing arbitrary topology of graphs requires dense connectivity due to non-regular edge structures of problem instances. In this thesis, the scalable architecture of CIM is proposed [73] and the computational performance of the model is estimated compared to existing algorithms [74, 75]. The linear coupling term is implemented with an external FPGA feedback circuit, along with a long optical fiber with nonlinear gain of DOPO pulses. To estimate its computational performance, the system is mathematically modeled and numerical simulation results are presented. It is reported with experimental results that CIM can implement arbitrary topology of graphs [76] up to $n = 2048$ [77]. To show the computational ability in contrast with an implementation on a many-core processor, we implemented several neural network based heuristics. It indicates relative speedup with $n = 2000$ (experiment) and $n = 20000$ (simulation) [78].

The system proposed here is interesting from three perspectives:

- It exploits physical degrees of freedom for optimization problem.
- It has a problem specific architecture and is scalable to thousands of pulses.
- It showed comparable performance to the heuristic algorithms run on current general-purpose processors.

1.7 Dissertation Structure

This dissertation is structured as follows. In the next chapter, the MAX-CUT problem, which is the main target problem, is defined and review existing algorithms and known theoretical results. In chapter 3, the physics of DOPO is introduced and the stochastic differential equations (SDEs) based on truncated Wigner representation in phase space is presented. Chapter 4 presents the computational system for MAX-CUT using DOPO is described and the scalable architecture, measurement-feedback scheme is proposed. The numerical simulation based on the SDEs is performed in Chapter 5 in order to estimate the computational performance of CIM. Then, the experimental demonstrations are performed in Chapters 6 and 7. In the former chapter, the results are compared to the simulation in Chapter 5. In the latter chapter, benchmark studies against existing algorithms are presented. Chapter 8 concludes the dissertation.

Chapter 2

Maximum Cut Problem

In this chapter we define the problem, review existing algorithms, and summarize theoretical results. Before that, we start from the preliminary knowledge of the graph.

2.1 Graph

Typical combinatorial optimization problems are defined on a graph. The graph is a discrete mathematical structure to express the relationship among any objects. It is introduced in the study of Leonhard Euler in 1735 and became more popular concept due to the emergence of the real-life applications such as search engine of web pages, social networks, and road networks [79, 80].

Definition 2.1 (Graph). A graph is a pair $G(V, E)$ of a set of vertices V and edges E . The number of vertices $n = |V|$ is called graph order and the number of edges $m = |E|$ is size of the graph.

A graph G can be described with an *adjacency matrix* $A \in \{0, 1\}^{n \times n}$. Unless otherwise noted, we assume the A is symmetric, which is the case that G is an *undirected graph*.

Definition 2.2 (Weighted Graph). A graph G is called a weighted graph when each edge $(i, j) = e \in E$ has the weight $w_{ij} \in \mathbb{R}$. Otherwise, the graph is called unweighted graph, whose weight matrix is identical to its adjacency matrix.

Definition 2.3 (Vertex Degree). The degree k of the vertex $v \in V$ in an unweighted graph is the number of edges whose one endpoint is v . If the G is weighted graph, the vertex degree of v_i is defined as

$$\sum_{j=1, j \neq i}^n w_{ij}.$$

2.1.1 Typical Topology of Graphs

Definition 2.4 (Complete Graph). The graph G is called a complete graph if all pairs of vertices are connected. The complete graph of order n is denoted by K_n .

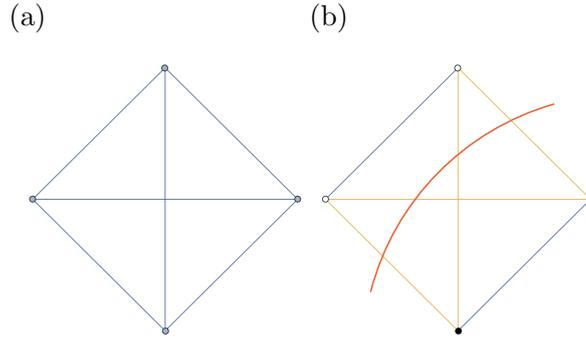


Figure 2.1. (a) A typical graph with 4 vertices and 6 edges (the complete graph K_4). It is the only cubic graph of $n = 4$. (b) A cut is demonstrated with red curve, which attains the maximum cut (defined in Def. 2.9) with 2 vertex signed as 1 (black) and the others as -1 (white).

Definition 2.5 (Cubic Graph). The graph G is called a k -regular graph if the vertex degree of all the vertices V are exactly k . Especially, a 3-regular graph is called the cubic graph. Note that the cubic graphs exist only when n is the even number.

Definition 2.6 (Cycle Graph). A cycle graph C_n of order n is the 2-regular graph where n vertices are connected by n edges.

The cycle graph is sometimes called (*ferromagnetic/anti-ferromagnetic Ising*) ring when we regard it as an Ising model. The length of the cycle is the number of edges contained.

Definition 2.7 (Möbius Ladder Graph). A Möbius ladder graph M_n is a cubic graph of order n ($n \in 2\mathbb{Z}$), which can be created from a C_n with adding edges connecting each opposite pair of node in the cycle.

Definition 2.8 (Erdős-Rényi Random Graph). Erdős-Rényi model is a model to generate random graphs with m edges, in which the edges are connected in probability m/nC_2 .

2.2 Problem Definition of MAX-CUT

One of the well-known examples of combinatorial optimization problems is a maximum cut problem (MAX-CUT) on a graph. It is essentially equivalent to finding the ground state of the Ising model in statistical mechanics [81, 82, 83].

Definition 2.9 (Cut). A cut in a graph $G(V, E)$ is a partition of vertices into two subsets $\{V_l, V_r\}$, where $V_l \cap V_r = \emptyset$ and $V_l \cup V_r = V$. The size of the cut $\{V_l, V_r\}$ is defined as the total weight of edges separated by the partition, i.e., edges which have one endpoint in V_l and the other in V_r :

$$\text{CUT}(\{V_l, V_r\}) = \sum_{v \in V_l, u \in V_r} w_{vu}.$$

Definition 2.10 (MAX-CUT). MAX-CUT is a problem to find the largest cut in a given undirected graph $G = (V, E)$.

If we put the binary values $x_i = \pm 1$ to express which side of the cut the vertex $i \in V$ belongs to, the judgement of edge whether it is cut or not can be done with checking the sign of x_i , namely $\frac{1-x_i x_j}{2}$. Then the objective function of MAX-CUT can be written as

$$\text{CUT}(x) = \sum_{1 \leq i < j \leq n} w_{ij} \frac{1 - x_i x_j}{2}, \quad (2.1)$$

where w_{ij} is the weight (or the adjacency matrix of the graph G). The MAX-CUT is a classically known example of NP-hard problems as in Reference [83].

2.3 Approximation Algorithm

In this section, approximate algorithms for MAX-CUT is presented. They have performance guarantee in terms of the solution quality, and run in polynomial time. The performance of an algorithm is described with approximation ratio and asymptotic behavior.

Definition 2.11 (Big- O Notation). Let f and g be functions defined on some subset of \mathbb{R} . If and only if $\exists M, x_0 > 0$ s.t. $|f(x)| < M|g(x)|$ ($\forall x > x_0$), we say f scales as $O(g)$ or simply write $f(x) = O(g(x))$.

Definition 2.12 (Polynomial Time Algorithm). An algorithm which finishes with $\exists k > 0$ s.t. $O(n^k)$ time operations is called polynomial-time algorithm.

We refer to a polynomial-time algorithm as *efficient*.

Definition 2.13 (Approximation Algorithm). An algorithm is an α -approximation for a maximization problem when it finds the solution of the value CUT in polynomial time which satisfies

$$\alpha \text{OPT} \leq \text{CUT} \leq \text{OPT}, \quad (2.2)$$

where $0 < \alpha < 1$ and OPT is the optimal solution of the problem.

2.3.1 Goemans-Williamson Algorithm (GW)

The Goemans-Williamson algorithm (GW) based on SDP is an α_{GW} -approximation algorithm for NP-hard MAX-CUT problems [84], where

$$\alpha_{\text{GW}} := \min_{0 < \theta < \pi} \frac{\theta/\pi}{(1 - \cos \theta)/2}. \quad (2.3)$$

The minimum value $\alpha_{\text{GW}} = \frac{2}{\pi \sin \theta^*} \approx 0.87856$ is achieved when $\theta = \tan(\theta/2)$.

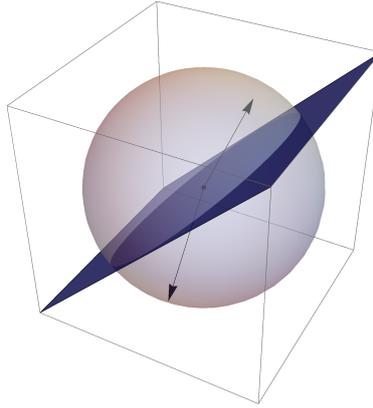


Figure 2.2. In this example, 2 (in n) vectors are separated by a hyper plane in SDP.

It achieves the best approximation ratio for MAX-CUT problems so far. With this algorithm, the original MAX-CUT problem is extended to a vector-valued optimization problem as maximizing

$$\frac{1}{2} \sum_{i < j} w_{ij} (1 - \vec{v}_i \cdot \vec{v}_j), \quad \vec{v}_i \in S^{k-1}, \quad (2.4)$$

where S^{k-1} is a unit sphere in \mathbb{R}^k and $k \leq n = |V|$. There exists a polynomial time algorithm to find the optimal solution of this extended problem (with error $\varepsilon > 0$), and its value is commonly called the SDP upper bound. A final solution to the original MAX-CUT problem is obtained by projecting the solution vector sets to randomly chosen one-dimensional Euclidean spaces, i.e., dividing the sphere by random hyperplanes, resulting in the geometric value of α_{GW} above.

There are three types of computational complexities of the best-known algorithms for solving the SDP relaxation. The matrix multiplicative weights method runs in almost linear time $O(m(\log n)^2 \varepsilon^{-4})$ when we have a graph with bounded degree (or more generally, weighted graph with bounded ratio between largest/smallest degree) [85], where ε represents the accuracy of the obtained solution. The Lagrangian relaxation method can be used for a graph with non-negative weights, which runs in $O(nm(\log n)^2 \varepsilon^{-3})$ time [86]. In general, when the graph has both positive and negative edge weights, the SDP problem is commonly solved using the interior point method, which scales as $O(n^{3.5} \log(1/\varepsilon))$ [87]. There are also faster general algorithms if only the matrices are sparse [88, 89]. In our computational experiments, the COPL-SDP, which is implemented with the primal-dual interior point method [90] was used. The SDP upper bound U_{SDP} and the solution C_{SDP} were obtained using the following parameters: interior point method was used until the relative gap $r_{\text{gap}} = 1 - P_{\text{obj}}/D_{\text{obj}}$ reached 10^{-3} , where P_{obj} and D_{obj} are the objective functions of the primal and dual of the SDP problem, respectively [91]. Random projection onto the one-dimensional space was executed linearly to the graph order, n times.

2.3.2 Greedy algorithm of Sahni-Gonzalez (SG) and its Variants

Sahni and Gonzalez constructed a greedy algorithm for MAX-CUT problems, which has 1/2-performance guarantee [92]. In this algorithm, vertices V are divided into two disjoint subsets $\{S_1, S_2\}$ sequentially. A vertex v_i is selected randomly and put into the subset to which less edges are connected from v_i . Later, several modifications are proposed and SG3 is one of them [93]. For each iterative process, the node with the maximum score is selected, and it is put into either set S_1 or S_2 so as to earn larger cuts. Here, the score function is defined as $x_i = |\sum_{j \in S_1} w_{ij} - \sum_{j \in S_2} w_{ij}|$ ($i = 1, \dots, n$). It stops when all the edges are evaluated to calculate the score function, thus SG3 scales as $O(n + m)$.

2.3.3 Other Approximation Algorithms

Spectral Partitioning

The variant of spectral partitioning achieves the 0.53112 approximation by calculating the minimum eigen value of the adjacency matrix iteratively [94].

Random Walk

The algorithm based on a number of random walks and aggregation gives $(0.5 + \delta)$ -approximation with a positive constant $\delta > 0$ [95].

2.3.4 Efficiently Solvable Classes

To see the specific subset of graphs which is solvable in polynomial time, we additionally define several type of graphs.

Definition 2.14 (Bipartite Graph). Let V_1, V_2 are two disjoint sets of vertices. A graph $G = (V_1 \cup V_2, E)$ is bipartite when endpoints of all edges $(i, j) \in E$ belong to the different set $i \in V_1, j \in V_2$.

Definition 2.15 (Planar Graph). A planar graph is a graph which can be embedded in a plane without any edge intersection.

A graph H is called *graph minor* of a graph G if it can be constructed by edge deletion, vertex deletion, and edge contraction from G , or equivalently called, the graph G has its minor H . For example, the graph planarity can be checked by checking the graph minor: a graph is planar iff it does not have $K_{3,3}, K_5$ -minor [96, 97].

Definition 2.16 (Weakly Bipartite Graph). A signed graph is a pair (G, Σ) , where $G = (V, E)$ is an undirected graph and $\Sigma \subset E$. A signed graph is called weakly bipartite if each vertex of the polyhedron in \mathbb{R}^E determined by $w_e \geq 0$ ($e \in E$), and $\sum_{e \in C} w_e \geq 1$ (C is an odd circuit) is integer.

The MAX-CUT problem generally belongs to NP-hard, though there exists a special subset of problems which can be solvable in polynomial time. For example, on the weakly

Table 2.1. Classical neural-network approaches for combinatorial optimization problems.

	Deterministic	Stochastic
Binary	Derandomized Hopfield network (HN)	Simulated annealing (SA)
Analog	Hopfield-Tank neural network (HTNN)	

bipartite graphs [98] (including the planar graphs [99, 100] and the graphs without K_5 -minors [101] studied earlier), the positively weighted graphs without a long odd cycle C_n ($n \in 2\mathbb{Z} + 1$) [102], and the graphs with integer edge weight bounded by n and fixed genus [103], MAX-CUT problems are reduced to the class P.

2.4 Metaheuristic

For many practical applications, heuristic algorithms are more convenient to use. It is because the degree of the polynomial in time complexity is concerned in practical. For example, the GW scales as $O(n^{3.5})$ in general, requires long computation time despite its complexity class of P when the number of vertices n increases. As a metaheuristic algorithm, the CIM can be interpreted as a mathematical model to solve combinatorial optimization problems using recurrently updated neurons with nonlinear activation function in each unit (with linear growth and nonlinear saturation of amplitudes [104]). From this point of view, there have been related and interesting approaches using mathematical models of the neurons (e.g., [105, 106]) and their networks (e.g., [107]). Hopfield developed the optimization algorithm by using such neural networks [108]. Then Hopfield and Tank extended it to the continuous-valued model to improve the performance and applied it to the combinatorial optimization problems [109, 110]. Simulated annealing (SA) is proposed in the same period [111].

We describe in this section the classical neural network models to solve the same combinatorial optimization problems, which are summarized in Table 2.1.

2.4.1 Derandomized Hopfield Network (HN)

J. J. Hopfield implemented a classical neural network model solving combinatorial optimization problems in his 1982 paper [108], which is referred to the Hopfield network (HN). The neuron in this model has the discrete output values $x_i = \pm 1$ with a simple majority voting update rule:

$$x_i \leftarrow \operatorname{sgn}\left(\sum_{j=1}^n J_{ij}x_j\right) \quad (2.5)$$

which will execute asynchronously. The spin index i is selected randomly in the original paper but we derandomized it to enhance the speed, i.e., the spin indices from $i = 1$ to $i = n$ are updated sequentially. Simultaneous updates introduce the instability or

periodic solution into the system. Since the update is local and deterministic, the system will converge to the nearest local minimum, which is determined by the initial state. Note that the model is originally proposed with $\{0, 1\}$ -binary neurons, but for comparison, we use equivalent $\{+1, -1\}$ -valued neurons.

2.4.2 Simulated Annealing (SA)

Metropolis et al. introduced a simple algorithm that can be used to provide an efficient simulation of a collection of atoms in equilibrium at a given temperature [112]. Kirkpatrick et al. applied the algorithm to optimization problems by replacing the energy of the atomic system to the cost function of optimization problems and using spin configurations σ , which is called the simulated annealing algorithm (SA) [111]. While the HN will often get stuck at poor local minima, it can escape from the local minima by the thermal fluctuation. The spin index i is selected randomly while temperature $T \geq 0$ is gradually decreased. In each step of this algorithm, a system is given with a random spin flip of x_i and the resulting change

$$\Delta E_i = 2x_i \sum_{j=1}^n J_{ij}x_j \quad (2.6)$$

in the energy is computed. If $\Delta E_i \leq 0$, the spin-flip is accepted, and the configuration with the flipped spin is used as the starting point of the next step. If $\Delta E_i > 0$, the spin is treated probabilistically, i.e., the probability that the new configuration is accepted is

$$P(\Delta E_i) := \exp\left(-\frac{\Delta E_i}{k_B T}\right), \quad (2.7)$$

with a control parameter of system temperature T and Boltzmann constant k_B . This choice of $P(\Delta E_i)$ in this Metropolis-Hastings procedure results in the system evolving into an equilibrium Boltzmann distribution. Repeating this procedure, with the temperature T gradually lowered to zero for sufficiently long time, leads to convergence to the lowest energy state. In a practical case, with the finite time, the annealing schedule affects the quality of output values. Here in our numerical simulations, the temperature was lowered according to the logarithmic function [113, 114].

Multi-spin coding

When we deal with the complete graph with binary weight, it is faster to push into multiple spins in a container (variable) and use the bit instruction. It is called multi-spin coding and then the energy calculation can be reduced to $w_{ij} \oplus \sigma_j$, where \oplus denote the bit-wise exclusive or (XOR) followed by the bit population count which can be done by the SIMD operation in parallel.

Algorithm 1 Simulated Annealing

```

1: procedure SA( $J$ )                                ▷ solve the Ising problem with coupling matrix  $J$ 
2:    $T \leftarrow \infty$                                 ▷ temperature
3:    $x \leftarrow$  random state                            ▷ initial spin configuration
4:   for  $i = 1, \dots, t_{\text{step}}$  do
5:      $x' \leftarrow$  neighbor of  $x$ 
6:      $\Delta E \leftarrow$  calculate energy difference between  $(x, x')$ 
7:     if  $\Delta E \leq 0$  then
8:        $x \leftarrow x'$                                 ▷ update
9:     else
10:       $x \leftarrow x'$  in probability  $P(\Delta E)$         ▷ probabilistic update
11:       $T \leftarrow T_{\text{new}}$                                 ▷ cooling  $T$ 
12:   return  $x$                                           ▷ spin configuration

```

2.4.3 Hopfield-Tank Neural Network (HTNN)

Hopfield and Tank proposed another neural network approach using an analog valued neuron $x_i \in [-1, 1]$, which is referred to the Hopfield-Tank neural network (HTNN) [110]. The time evolution of the HTNN is described by ordinary differential equations (ODE):

$$\frac{dx_i}{dt} = -\alpha x_i + \beta \sum_{j=1}^n J_{ij} f(x_j), \quad (2.8)$$

where $f(x)$ is a nonlinear sigmoid function. In this study, $\tanh(x)$ is used as $f(x)$. The key observation is that, in the extremely high linear gain limit, i.e., when the slope of the sigmoid function around 0 is steep, the energy function of the system becomes proportional to the original objective function of the problem. The parameters in later section are optimized as the neuron decay rate $\alpha = 6$ and the synaptic connection strength $\beta = 0.1$ to achieve the best performance for the given MAX-CUT problems. The numerical integration of eq (2.8) is performed by the Euler method with the discrete time step $\Delta t = 0.01$.

2.4.4 Breakout Local Search (BLS)

The power of breakout local search (BLS) appears in the benchmark result for G-set graphs [115]. It updated the best solutions ever found in G-set with the specialized data structure for sorting and dedicated procedure to escape from local minima. The algorithm is combination of steepest descent and forced spin flipping: after being trapped by a local minima as a result of steepest descent procedure, three types of forced spin flipping (single, pair, and random) are probabilistically executed according to the vertex influence list (i.e., which vertex would increase the number of cut most when it is flipped) on each subset of partition. Its efficient data structure stores the gain list for each side of the partition and vertex address list, which enables constant $O(1)$ time spin configuration update.

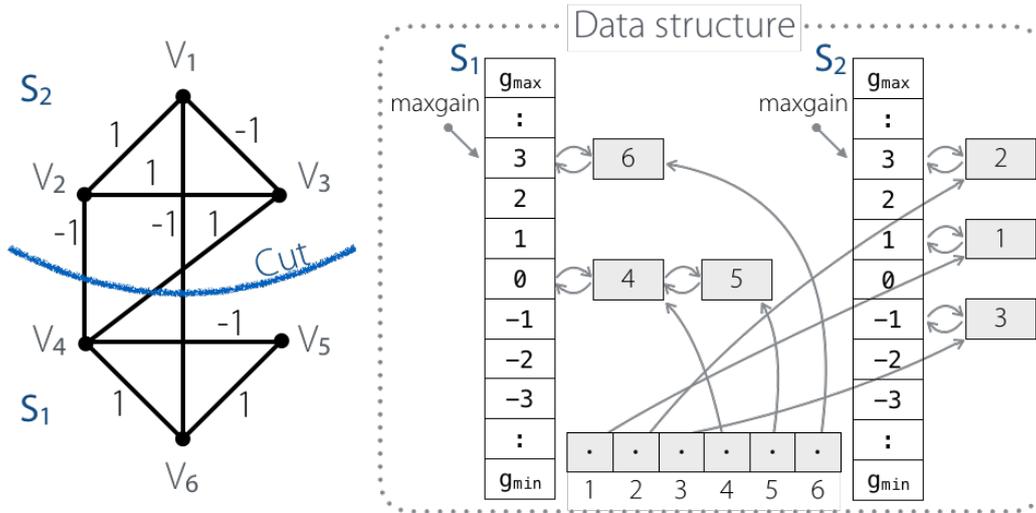


Figure 2.3. The bucket sorting data structure used in BLS on a graph of $n = 6$.

2.4.5 Other Metaheuristics

Genetic Algorithm

Inspired by the natural evolution, genetic algorithm implements selection, crossover, and mutation of solutions [116].

Population Annealing

Population annealing is a variant of simulated annealing which uses a population of replicas resampled at each temperature step [117, 118].

2.4.6 Off-the-Shelf Optimizer

It is able to use an off-the-shelf optimizer to solve the problems. For example, it is known that the IBM CPLEX mixed-integer linear programming (MILP) can solve the unconstrained binary quadratic programming (UBQP) formulation of MAX-CUT [119, 120]. The CPLEX MILP package uses a branch-and-cut algorithm with LP relaxation. It is reported that this method performs well in a specific topology of graphs [119, 120].

2.5 Hardness of Approximation

2.5.1 Interactive Proof

The complexity class defined in Section 1.2.2 will be restated with a prover P and a verifier V .

Definition 2.17 (NP). The language L is NP if and only if there is a polynomial time deterministic verifier V and an arbitrary powerful prover P which satisfy:

- $\forall x \in L, P$ can write a proof of length $\text{poly}(|x|)$ that V accepts.

- $\forall x \notin L$, no matter what $\text{poly}(|x|)$ -length proof P writes, V rejects.

Definition 2.18 (NP-hard). A problem H is NP-hard if every problem L in NP has a polynomial-time reduction from L to H .

Definition 2.19 (NP-complete). A decision problem C belongs to NP-complete if

- C is in NP-hard.
- Every problem in NP is reducible to C in polynomial time.

2.5.2 Probabilistically Checkable Proof (PCP) Theorem

Many attempts have been made to approximately solve NP-hard MAX-CUT problems. There is a known theoretical bound for approximating MAX-CUT.

Definition 2.20 (Probabilistically Checkable Proof (PCP)). A computational complexity class $\text{PCP}(r, q)$ has a proof system with input $x \in \{0, 1\}^n$. A prover P writes a proof of $\text{poly}(x)$ bit, then verifier V selects q bits in the proof and run a polynomial-time test with random number generator of r -bit, and determine whether to accept or reject the proof, which should guarantee:

- (Completeness) if $x \in L$ then P can construct a proof that V accepts with probability 1.
- (Soundness) for every $x \notin L$, no matter what proof P writes, V accepts with probability at most $1/2$.

Theorem 1 (PCP Theorem). $\text{NP} \subset \text{PCP}(\log n, 1)$.

A proof of Theorem 1 can be found in Reference [121]. The PCP theorem states that no polynomial time algorithms can approximate MAX-CUT problems better than $16/17 \approx 0.94118$ if $\text{P} \neq \text{NP}$ [122, 123, 124]. The value comes from the *gap amplification* which is used in the proof [121, 124].

2.5.3 Unique Games Conjecture (UGC)

The best approximation currently known is the approximation rate of α_{GW} in the algorithm of Goemans and Williamson explained in Section 2.3.1. Unique games conjecture (UGC) by Khot claims that the α_{GW} is optimal [125], although it is still an open question.

The conjecture is stated as below [125]: For any $\eta, \gamma > 0$ there exists a constant $M = M(\eta, \gamma)$ such that it is NP-hard to distinguish whether the unique label cover problem with the label set of size M has optimum at least $1 - \eta$ or at most γ . Please see Reference [125] for the details.

Chapter 3

Degenerate OPO

In this chapter, the key physical component for CIM is introduced and formulated mainly according to References [126, 127].

3.1 Basics of Degenerate Optical Parametric Oscillators

In optical system, nonlinear mediums show passive parametric process such as periodically poled lithium niobate (PPLN) crystal and optical fiber, which are namely $\chi^{(2)}$ (e.g., LiNbO_3) and $\chi^{(3)}$ nonlinear susceptibility. Below the oscillation threshold, a photon is split into a two entangled photon pair, that is known as spontaneous parametric down conversion. Optical parametric oscillator (OPO) exploits it to obtain the energy from the pump. Especially for the $\chi^{(2)}$ case, three fields interact – pump, signal, and idler as in Figure 3.1. Denoting these frequency $\omega_p, \omega_s, \omega_i$, respectively, the relation becomes $\omega_p = \omega_s + \omega_i$ (energy conservation), with its wave vectors $\vec{k}_p, \vec{k}_s, \vec{k}_i$ matched as $\vec{k}_p = \vec{k}_s + \vec{k}_i$ (phase matching). Practically, it is enabled by the technique called quasi-phase matching in the periodic structure in the nonlinear medium, then the signal mode gets a large gain from the pump and enabled to oscillate in a cavity [128].

In the special case of OPO frequency at degeneracy, i.e., $\omega_s = \omega_i$ hence $\omega_p = 2\omega_s$, it is called degenerate OPO (DOPO). In this case, maximum signal gain is attained at the phase relation

$$\phi_p = 2\phi_s + \pi/2 + 2m\pi \quad (m \in \mathbb{Z}), \quad (3.1)$$

where ϕ_p and ϕ_s denote the pump and signal phase, respectively. The coherence between pump and signal is established (self-phase-locking) and the signal phase state is stable either in $\phi_s \in \{0, \pi\}$.

In the context of quantum optics, OPO shows various quantum phenomena including squeezing, macroscopic superposition [129] and entanglement [130, 131]. The OPO can also be used as a physical random number generator [132]. In this chapter, the formulation of DOPO is described in terms of the phase space method especially for positive P representation and truncated Wigner representation.

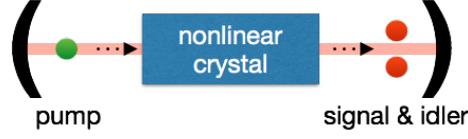


Figure 3.1. A schematic picture of an optical parametric oscillator cavity. A pump photon is divided into 2 photons; signal and idler, which are indistinguishable under the degeneracy condition.

3.2 System Hamiltonian

An open quantum system can be described by a *Hamiltonian* $\hat{\mathcal{H}}$ including the *reservoir* (or *bath*) interaction term $\hat{\mathcal{H}}_{\text{SR}}$, here *hat* ($\hat{\cdot}$) is set to emphasize the operator. In this way, the total Hamiltonian of a single DOPO is described as follows [133]:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{\text{free}} + \hat{\mathcal{H}}_{\text{param}} + \hat{\mathcal{H}}_{\text{pump}} + \hat{\mathcal{H}}_{\text{SR}}, \quad (3.2)$$

where

$$\hat{\mathcal{H}}_{\text{free}} = \hbar\omega_s \hat{a}_s^\dagger \hat{a}_s + \hbar\omega_p \hat{a}_p^\dagger \hat{a}_p, \quad (3.3a)$$

$$\hat{\mathcal{H}}_{\text{param}} = i\hbar \frac{\kappa}{2} (\hat{a}_s^\dagger{}^2 \hat{a}_p - \hat{a}_p^\dagger \hat{a}_s^2), \quad (3.3b)$$

$$\hat{\mathcal{H}}_{\text{pump}} = i\hbar(\varepsilon \hat{a}_p^\dagger - \varepsilon^* \hat{a}_p), \quad (3.3c)$$

$$\hat{\mathcal{H}}_{\text{SR}} = \hbar(\hat{a}_s \hat{\Gamma}_{\text{Rs}}^\dagger + \hat{\Gamma}_{\text{Rs}} \hat{a}_s^\dagger + \hat{a}_p \hat{\Gamma}_{\text{Rp}}^\dagger + \hat{\Gamma}_{\text{Rp}} \hat{a}_p^\dagger). \quad (3.3d)$$

First, (3.3a) is the free field Hamiltonian for signal and pump fields, whose photon creation/annihilation operators are denoted by $\hat{a}_s^\dagger, \hat{a}_s$ and $\hat{a}_p^\dagger, \hat{a}_p$, respectively, with commutation relations $[\hat{a}_k, \hat{a}_k^\dagger] := \hat{a}_k \hat{a}_k^\dagger - \hat{a}_k^\dagger \hat{a}_k = 1$ ($k \in \{s, p\}$). With the Planck constant \hbar ($= h/2\pi$), the single photon energy $\hbar\omega_k$ is multiplied by the photon number $\hat{a}_k^\dagger \hat{a}_k$. Next, the parametric interaction of two fields is expressed in (3.3b) with the parametric coupling constant κ . Here, $\hat{a}_s^\dagger{}^2 \hat{a}_p$ describes the parametric gain, where a single pump photon is absorbed and then two signal photons are created, while $\hat{a}_p^\dagger \hat{a}_s^2$ is its reverse process. DOPOs are pumped by an external laser exactly at the cavity resonance frequency (i.e., with no detuning), whose field is $\varepsilon = F_p e^{-i\omega_p t}$ in (3.3c). Finally, the system-reservoir coupling is taken into (3.3d), where $\hat{\Gamma}_{\text{R}k}$ ($k \in \{s, p\}$) are the reservoir field operators.

3.3 Master Equation

The master equation formulates all of the density matrix elements including the off-diagonal parts, analogous to the classical master equations which only have the diagonal elements. The quantum state is represented by *state vectors* $|\psi_s\rangle$ ($s \in \Lambda$) in a complex Hilbert space¹ and its statistical mixture, *density operator* $\hat{\rho} = \sum_{s \in \Lambda} p_s |\psi_s\rangle \langle \psi_s|$, where

¹Hilbert space is a vector space which is complete with respect to the norm induced by the inner product defined on it.

each state $|\psi_s\rangle$ among the index set $s \in \Lambda \subset \mathbb{N}$ is weighted by a probability $0 \leq p_s \leq 1$ and we use *bra-ket* notation $\langle \psi_s | = |\psi_s\rangle^\dagger$ where *dagger* (\cdot^\dagger) denotes Hermitian conjugate. The equation of motion of the density matrix $\hat{\rho}$ follows from the time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} |\psi_s\rangle = \hat{\mathcal{H}} |\psi_s\rangle \quad (3.4)$$

which leads to

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{\mathcal{H}}, \hat{\rho}]. \quad (3.5)$$

This is called the *Liouville von-Neumann equation*

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\hat{\mathcal{H}}_{\text{param}}, \hat{\rho}] + [\hat{\mathcal{H}}_{\text{pump}}, \hat{\rho}] + [\hat{\mathcal{H}}_{\text{SR}}, \hat{\rho}] \quad (3.6)$$

in the interaction picture.

The quantum master equation for the DOPO is obtained by applying Born-Markov approximation to tracing out the reservoir field with assuming $\hbar = 1$ [126]

$$\begin{aligned} \frac{\partial \hat{\rho}}{\partial t} = & -i\omega_s [\hat{a}_s^\dagger \hat{a}_s, \hat{\rho}] - i\omega_p [\hat{a}_p^\dagger \hat{a}_p, \hat{\rho}] \\ & + i\frac{\kappa}{2} [\hat{a}_s^{\dagger 2} \hat{a}_p - \hat{a}_p^\dagger \hat{a}_s^2, \hat{\rho}] - i[\varepsilon \hat{a}_p^\dagger + \varepsilon^* \hat{a}_p, \hat{\rho}] \\ & + \frac{\gamma_s}{2} (2\hat{a}_s \hat{\rho} \hat{a}_s^\dagger - \hat{a}_s^\dagger \hat{a}_s \hat{\rho} - \hat{\rho} \hat{a}_s^\dagger \hat{a}_s) + \frac{\gamma_p}{2} (2\hat{a}_p \hat{\rho} \hat{a}_p^\dagger - \hat{a}_p^\dagger \hat{a}_p \hat{\rho} - \hat{\rho} \hat{a}_p^\dagger \hat{a}_p) \\ & + \gamma_s \bar{n}_s (\hat{a}_s \hat{\rho} \hat{a}_s^\dagger + \hat{a}_s^\dagger \hat{\rho} \hat{a}_s - \hat{a}_s^\dagger \hat{a}_s \hat{\rho} - \hat{\rho} \hat{a}_s \hat{a}_s^\dagger) \\ & + \gamma_p n_p (\hat{a}_p \hat{\rho} \hat{a}_p^\dagger + \hat{a}_p^\dagger \hat{\rho} \hat{a}_p - \hat{a}_p^\dagger \hat{a}_p \hat{\rho} - \hat{\rho} \hat{a}_p \hat{a}_p^\dagger), \end{aligned} \quad (3.7)$$

where γ_s and γ_p are the decay rate of signal and pump field, respectively. \bar{n}_s is the mean signal photon number. When the pump decay rate is much higher than signal, which is called a *singly-resonant* OPO, the adiabatic elimination is applied to reduce the term for pump field [127]:

$$\begin{aligned} \frac{\partial \hat{\rho}}{\partial t} = & -i\omega_s [\hat{a}_s^\dagger \hat{a}_s, \hat{\rho}] + i\frac{\kappa}{2} [\hat{a}_s^{\dagger 2} \hat{a}_p - \hat{a}_p^\dagger \hat{a}_s^2, \hat{\rho}] \\ & + \frac{\gamma_s}{2} (2\hat{a}_s \hat{\rho} \hat{a}_s^\dagger - \hat{a}_s^\dagger \hat{a}_s \hat{\rho} - \hat{\rho} \hat{a}_s^\dagger \hat{a}_s) + \gamma_s \bar{n}_s (\hat{a}_s \hat{\rho} \hat{a}_s^\dagger + \hat{a}_s^\dagger \hat{\rho} \hat{a}_s - \hat{a}_s^\dagger \hat{a}_s \hat{\rho} - \hat{\rho} \hat{a}_s \hat{a}_s^\dagger). \end{aligned} \quad (3.8)$$

3.4 Fokker-Planck Equation

Since it is computationally hard to directly simulate the quantum master equation, the density matrix will be expanded into a phase space via one of the following (quasi-)probability distributions. Especially, positive P or truncated Wigner representations are used when we treat non-classical DOPOs. After reviewing each representation, we see how to obtain Fokker-Planck equations from (3.8).

3.4.1 Phase-Space Methods

In the quantum mechanics, that concerns the operator order, the classical correspondence is derived by determining which default operator order to use:

- Normal Order: The creation operator \hat{a}^\dagger should be placed on the left, and annihilation \hat{a} on the right, e.g., $\hat{a}^\dagger \hat{a}$.
- Symmetric Order: The sum of symmetrical order of both operators is used, e.g., $\hat{a}^\dagger \hat{a} + \hat{a} \hat{a}^\dagger$

This is expressed by *characteristic functions*.

Glauber-Sudarshan P Representation

In Glauber-Sudarshan P representation (or simply P representation), the normal ordered operator is adopted, whose characteristic function can be defined as

$$\chi_N(z, w) := \text{tr}(\rho e^{iw\hat{a}^\dagger} e^{iz\hat{a}}) \quad (z, w \in \mathbb{C}). \quad (3.9)$$

From this notation, the normal-ordered operator averages are calculated by the derivatives

$$\begin{aligned} \langle \hat{a}^{\dagger p} \hat{a}^q \rangle &= \text{tr}(\hat{\rho} \hat{a}^{\dagger p} \hat{a}^q) \\ &= \frac{\partial^{p+q}}{\partial (iw)^p \partial (iz)^q} \chi_N(z, w) \Big|_{z=w=0} \quad (p, q \in \mathbb{Z}), \end{aligned} \quad (3.10)$$

which is a holomorphic function. Now the P representation is the Fourier transform of (3.10). The density matrix is expanded as

$$\hat{\rho} = \int d^2\alpha |\alpha\rangle \langle \alpha| P(\alpha), \quad (3.11)$$

where $|\alpha\rangle$ is a *coherent state*, i.e., the right eigen state of the annihilation operator \hat{a} as $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$ (Figure 3.2). The integral is taken over the complex plane, i.e., $\int d^2\alpha := \int_{\mathbb{C}} dz = \int_{\mathbb{R}} \int_{\mathbb{R}} dx dy$ with $\alpha = x + iy$. Coherent states are not orthogonal but complete (the state is normalized):

$$\int d\alpha^2 |\alpha\rangle \langle \alpha| = 1. \quad (3.12)$$

Here, $\hat{\rho}$ can be expressed as a diagonal sum of coherent states since the coherent states form the overcomplete basis in the phase space.

But this formulation breaks down in the DOPO case. To see this, in brief, let us consider stochastic differential equations formally derived from the Fokker-Planck equation (the detail will be explained later with the other representation):

$$dx = -\gamma_s(1-p)x dt + \frac{\gamma_s}{2} \sqrt{\gamma_p} dW_x, \quad (3.13a)$$

$$dy = -\gamma_s(1+p)y dt + i \frac{\gamma_s}{2} \sqrt{\gamma_p} dW_y, \quad (3.13b)$$

where $p > 0$ is injected pump rate. Now, the fluctuation term in the second equation is pure imaginary while the y is a real number. Although there is a contradiction, the formalism is valid for the quantum averages. So we will modify it by introducing redundant complex variables with keeping average values the same.

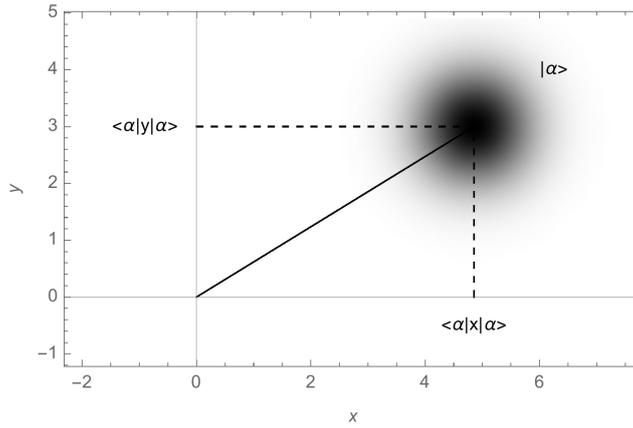


Figure 3.2. Schematic picture of a coherent state $|\alpha\rangle$ of the amplitude $|\alpha|$ in a phase space (x, y) .

Positive P Representation

The positive P representation is the generalization of the Glauber-Sudarshan P representation by Drummond and Gardiner [134]. It enabled to deal with the phase-space method even when the case of the diffusion matrix is not positive semidefinite such as DOPO [135]. In this formulation, the density operator is expanded as

$$\hat{\rho} = \iint d^2\alpha d^2\beta P(\alpha, \beta) \frac{|\alpha\rangle\langle\beta^*|}{\langle\beta^*|\alpha\rangle}. \quad (3.14)$$

The distribution function P becomes always positive. In the quantum optics, a state which cannot be represented by a sum of coherent states is called non-classical. Hence, the positive P representation can express non-classical states of light such as squeezed states.

The idea comes from the modification of (3.13a) and (3.13b) using complex conjugate variables $z = x + iy$, $z^* = x - iy$, then

$$dz = -\gamma_s(z - pz^*) dt + \gamma_s \sqrt{(1/2\gamma_p)\lambda} dW_z, \quad (3.15a)$$

$$dz^* = -\gamma_s(z^* + pz) dt + \gamma_s \sqrt{(1/2\gamma_p)\lambda} dW_{z^*} \quad (3.15b)$$

with new independent Wiener increments

$$dW_z = \frac{1}{\sqrt{2}}(dW_x - dW_y), \quad dW_{z^*} = \frac{1}{\sqrt{2}}(dW_x + dW_y). \quad (3.16)$$

At this point, there is still an inconsistency because two Wiener increments should be independent ($dW_{z^*} \neq (dW_z)^*$). Hence in the Positive P representation we introduce two independent complex variables α and β , which replace z and z^* , respectively, with keeping the mean becomes complex conjugate, i.e., $\langle\alpha\rangle = \langle\beta\rangle^*$, or it originates from $\langle\hat{a}\rangle = \langle\alpha\rangle$ and $\langle\hat{a}^\dagger\rangle = \langle\beta\rangle$.

Wigner Representation

While the P and positive P representations use the normal order of operators, Wigner representation comes from the symmetric order, whose characteristic function is

$$\chi_S(z, z^*) = \text{tr}(\rho e^{iz^* \hat{a}^\dagger + iz \hat{a}}). \quad (3.17)$$

The expansion of the exponential in (3.17) will be

$$\begin{aligned} e^{iz^* \hat{a}^\dagger + iz \hat{a}} &= \sum_{m=0}^{\infty} \frac{1}{m!} (iz^* \hat{a}^\dagger + iz \hat{a})^m \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{(iz^*)^n (iz)^m}{n! m!} (\hat{a}^{\dagger n} \hat{a}^m)_S, \end{aligned} \quad (3.18)$$

where $(\hat{a}^{\dagger n} \hat{a}^m)_S$ denotes the average of operator product in symmetric order, e.g.,

$$(\hat{a}^\dagger \hat{a})_S = \frac{1}{2} (\hat{a}^\dagger \hat{a} + \hat{a} \hat{a}^\dagger), \quad (3.19a)$$

$$(\hat{a}^{\dagger 2} \hat{a})_S = \frac{1}{3} (\hat{a}^{\dagger 2} \hat{a} + \hat{a}^\dagger \hat{a} \hat{a}^\dagger + \hat{a} \hat{a}^{\dagger 2}), \quad (3.19b)$$

⋮

Then the symmetric order operator average will be

$$\begin{aligned} \langle (\hat{a}^{\dagger n} \hat{a}^m)_S \rangle &= \text{tr}(\hat{\rho} (\hat{a}^{\dagger n} \hat{a}^m)_S) \\ &= \frac{\partial^{p+q}}{\partial (iz^*)^p \partial (iz)^q} \chi_S(z^*, z) \Big|_{z=z^*=0}. \end{aligned} \quad (3.20)$$

The Wigner quasi-distribution $W(\alpha, \alpha^*)$ is the Fourier transform of $\chi(z, z^*)$:

$$W(\alpha, \alpha^*) = \frac{1}{\pi^2} \int d^2 z \chi_S(z, z^*) e^{-iz^* \alpha^*} e^{-iz \alpha}, \quad (3.21)$$

$$\chi_S(z, z^*) = \int d^2 \alpha W(\alpha, \alpha^*) e^{iz^* \alpha^*} e^{iz \alpha}, \quad (3.22)$$

which is normalized as $\int d^2 \alpha W(\alpha, \alpha^*) = 1$. Substituting (3.22) gives

$$\langle (\hat{a}^{\dagger n} \hat{a}^m)_S \rangle = \int d^2 \alpha W(\alpha, \alpha^*) \alpha^{*p} \alpha^q. \quad (3.23)$$

In the Wigner case, $\hat{\rho}$ is expanded by the Wigner function $W(\alpha)$:

$$\hat{\rho} = \int d\lambda e^{\lambda^* \hat{a} - \lambda \hat{a}^\dagger} \left\{ \int d\alpha e^{\lambda \hat{a}^* - \lambda^* \hat{a}} W(\alpha) \right\}. \quad (3.24)$$

This representation can treat non-classical states as well as the positive P . Note that the relation between the two characteristic functions is given by the Baker-Hausdorff theorem as

$$\chi_S(z, z^*) = e^{-\frac{1}{2}|z|^2} \chi_N(z, z^*). \quad (3.25)$$

Theorem 2 (Baker-Hausdorff). If \hat{O}_1 and \hat{O}_2 are two non-commuting operators that both commute with their commutator, then

$$e^{\hat{O}_1 + \hat{O}_2} = e^{\hat{O}_1} e^{\hat{O}_2} e^{-\frac{1}{2}[\hat{O}_1, \hat{O}_2]} = e^{\hat{O}_2} e^{\hat{O}_1} e^{\frac{1}{2}[\hat{O}_1, \hat{O}_2]}. \quad (3.26)$$

3.4.2 In the P Representation

The master equation for the DOPO signal (3.8) can be also described by *Lindblad form*

$$\dot{\hat{\rho}} = -i\omega_s[\hat{a}_s^\dagger\hat{a}_s, \hat{\rho}] + \kappa(2\hat{a}_s\hat{\rho}\hat{a}_s^\dagger - \hat{a}_s^\dagger\hat{a}_s\hat{\rho} - \hat{\rho}\hat{a}_s^\dagger\hat{a}_s) + 2\kappa\bar{n}_s(\hat{a}_s\hat{\rho}\hat{a}_s^\dagger + \hat{a}_s^\dagger\hat{\rho}\hat{a}_s - \hat{a}_s^\dagger\hat{a}_s\hat{\rho} - \hat{\rho}\hat{a}_s\hat{a}_s^\dagger). \quad (3.27)$$

Substituting (3.11) into the above equation, we get the equation of motion for the time dependent P distribution

$$\begin{aligned} \int d^2\alpha |\alpha\rangle\langle\alpha| \frac{\partial}{\partial t} P(\alpha, t) &= \int d^2\alpha P(\alpha, t) [(\hat{a}_s^\dagger\hat{a}_s|\alpha\rangle\langle\alpha| - |\alpha\rangle\langle\alpha|\hat{a}_s^\dagger\hat{a}_s) \\ &+ \kappa(2\hat{a}_s|\alpha\rangle\langle\alpha|\hat{a}_s^\dagger - \hat{a}_s^\dagger\hat{a}_s|\alpha\rangle\langle\alpha| - |\alpha\rangle\langle\alpha|\hat{a}_s^\dagger\hat{a}_s) \\ &+ 2\kappa\bar{n}_s(\hat{a}_s|\alpha\rangle\langle\alpha|\hat{a}_s^\dagger + \hat{a}_s^\dagger|\alpha\rangle\langle\alpha|\hat{a}_s - \hat{a}_s^\dagger\hat{a}_s|\alpha\rangle\langle\alpha| - |\alpha\rangle\langle\alpha|\hat{a}_s\hat{a}_s^\dagger)]. \end{aligned} \quad (3.28)$$

We can find the replacing rule of the operators

$$\hat{a}|\alpha\rangle\langle\alpha|\hat{a}^\dagger = \alpha|\alpha\rangle\langle\alpha|\alpha^* = |\alpha|^2|\alpha\rangle\langle\alpha|, \quad (3.29a)$$

$$\hat{a}^\dagger\hat{a}|\alpha\rangle\langle\alpha| = \hat{a}^\dagger\alpha|\alpha\rangle\langle\alpha| = \alpha\hat{a}^\dagger|\alpha\rangle\langle\alpha| = \alpha\left(\frac{\partial}{\partial\alpha} + \alpha^*\right)|\alpha\rangle\langle\alpha|, \quad (3.29b)$$

$$|\alpha\rangle\langle\alpha|\hat{a}^\dagger\hat{a} = |\alpha\rangle\langle\alpha|\alpha^*\hat{a} = \alpha^*|\alpha\rangle\langle\alpha|\hat{a} = \alpha^*\left(\frac{\partial}{\partial\alpha^*} + \alpha\right)|\alpha\rangle\langle\alpha|, \quad (3.29c)$$

$$|\alpha\rangle\langle\alpha|\hat{a}\hat{a}^\dagger = \left(\frac{\partial}{\partial\alpha^*} + \alpha\right)|\alpha\rangle\langle\alpha|\hat{a}^\dagger = \left(\frac{\partial}{\partial\alpha^*} + \alpha\right)\alpha^*|\alpha\rangle\langle\alpha|, \quad (3.29d)$$

$$\hat{a}^\dagger|\alpha\rangle\langle\alpha|\hat{a} = \left(\frac{\partial}{\partial\alpha} + \alpha^*\right)|\alpha\rangle\langle\alpha|\hat{a} = \left(\frac{\partial}{\partial\alpha} + \alpha^*\right)\left(\frac{\partial}{\partial\alpha^*} + \alpha\right)|\alpha\rangle\langle\alpha|. \quad (3.29e)$$

Then the operator equation (3.8) can be represented in terms of complex numbers

$$\begin{aligned} \int d^2\alpha |\alpha\rangle\langle\alpha| \frac{\partial}{\partial t} P(\alpha, t) &= \int d^2\alpha P(\alpha, t) \left[-\left(\frac{\gamma}{2} + i\omega_s\right)\alpha\frac{\partial}{\partial\alpha} - \left(\frac{\gamma}{2} - i\omega_s\right)\alpha^*\frac{\partial}{\partial\alpha^*} + \gamma\bar{n}\frac{\partial^2}{\partial\alpha\partial\alpha^*} \right] |\alpha\rangle\langle\alpha| \quad (3.30) \\ &= \int d^2\alpha |\alpha\rangle\langle\alpha| \left[\left(\frac{\gamma}{2} + i\omega_s\right)\frac{\partial}{\partial\alpha}\alpha + \left(\frac{\gamma}{2} - i\omega_s\right)\frac{\partial}{\partial\alpha^*}\alpha^* + \gamma\bar{n}\frac{\partial^2}{\partial\alpha\partial\alpha^*} \right] P(\alpha, t), \end{aligned}$$

where the mean signal photon number is represented as \bar{n} . The integration (3.12) gives

$$\frac{\partial P}{\partial t} = \left[\left(\frac{\gamma}{2} + i\omega_s\right)\frac{\partial}{\partial\alpha}\alpha + \left(\frac{\gamma}{2} - i\omega_s\right)\frac{\partial}{\partial\alpha^*}\alpha^* + \gamma\bar{n}\frac{\partial^2}{\partial\alpha\partial\alpha^*} \right] P. \quad (3.31)$$

3.4.3 In the Positive P Representation

We can derive the Fokker-Planck equations for positive P by replacing the operators with the derivatives in (3.7). The derivatives operators from the left in (3.8)

$$\frac{\partial}{\partial\hat{a}_s} \rightarrow i\alpha_s, \quad \frac{\partial}{\partial\hat{b}_s} \rightarrow -i\beta_s, \quad \frac{\partial}{\partial\hat{a}_p} \rightarrow -i\alpha_p, \quad \frac{\partial}{\partial\hat{b}_p} \rightarrow -i\beta_p, \quad (3.32)$$

and variables

$$\hat{a}_s \rightarrow -i \frac{\partial}{\partial \alpha_s}, \quad \hat{b}_s \rightarrow -i \frac{\partial}{\partial \beta_s}, \quad \hat{a}_p \rightarrow -i \frac{\partial}{\partial \alpha_p}, \quad \hat{b}_p \rightarrow -i \frac{\partial}{\partial \beta_p}. \quad (3.33)$$

It leads to

$$\frac{\partial}{\partial t} P(\alpha_s, \beta_s, \alpha_p, \beta_p) = \mathcal{L}P(\alpha_s, \beta_s, \alpha_p, \beta_p), \quad (3.34)$$

where the operator to the pump and signal fields is

$$\begin{aligned} \mathcal{L} = & \frac{\partial}{\partial \alpha_s} \left(\frac{\gamma_s}{2} \alpha_s - \kappa \alpha_p \beta_s \right) + \frac{\partial}{\partial \beta_s} \left(\frac{\gamma_s}{2} \beta_s - \kappa \beta_p \alpha_s \right) \\ & + \frac{\partial}{\partial \alpha_p} \left(\frac{\gamma_p}{2} \alpha_p - F_p + \frac{\kappa}{2} \alpha_s^2 \right) + \frac{\partial}{\partial \beta_p} \left(\frac{\gamma_p}{2} \beta_p - F_p + \frac{\kappa}{2} \beta_s^2 \right) \\ & + \frac{1}{2} \left(\frac{\partial^2}{\partial \alpha_s^2} (\kappa \alpha_p) + \frac{\partial^2}{\partial \beta_s^2} (\kappa \beta_p) + 2\gamma_s \bar{n}_s \frac{\partial^2}{\partial \alpha_s \partial \beta_s} + 2\gamma_p \bar{n}_p \frac{\partial^2}{\partial \alpha_p \partial \beta_p} \right), \end{aligned} \quad (3.35)$$

where α_s, β_s and α_p, β_p denote the independent coherent eigenstates corresponding to the signal and pump field operators, respectively.

When the signal is singly resonant, namely $\gamma_p \gg \gamma_s$, the system is dominated by the signal field dynamics. This approximation is called adiabatic elimination, that makes the Fokker-Planck equation rather simple:

$$\frac{\partial}{\partial t} P(\alpha_s, \beta_s) = \mathcal{L}P(\alpha_s, \beta_s), \quad (3.36)$$

where the operator to the signal field contains

$$\begin{aligned} \mathcal{L} = & \frac{\partial}{\partial \alpha_s} [\alpha_s - (p - \alpha_s^2) \beta_s] + \frac{\partial}{\partial \beta_s} [\beta_s - (p - \beta_s^2) \alpha_s] \\ & + \frac{1}{2A_s} \left[\frac{\partial^2}{\partial \alpha_s^2} (p - \alpha_s^2) + \frac{\partial^2}{\partial \beta_s^2} (p - \beta_s^2) \right]. \end{aligned} \quad (3.37)$$

Here we normalized both sides by the saturation parameter for signal amplitude $A_s = \sqrt{\gamma_p \gamma_s / 2\kappa^2}$ (the steady state amplitude at the pump rate $p = 2$ to the threshold $F_{\text{th}} = \gamma_s \gamma_p / 4\kappa$) and time unit of signal decay rate $\gamma_s \tau / 2$ (twice the cavity photon lifetime) is used.

3.4.4 In the Truncated Wigner Representation

The Fokker-Planck equation in the Wigner representation is derived from (3.31) in the P representation with regard to the relation between the different orders of operator characteristic functions (3.25)

$$\begin{aligned} \frac{\partial W}{\partial t} = & \left\{ \frac{\partial}{\partial \alpha} \left[\left(\frac{\gamma_s}{2} + i\omega_s \right) - \kappa \alpha^* \beta \right] + \frac{\partial}{\partial \alpha} \left[\left(\frac{\gamma_s}{2} - i\omega_s \right) \alpha^* - \kappa \alpha \beta^* \right] \right. \\ & \left. + \gamma_s \left(\bar{n} + \frac{1}{2} \right) \frac{\partial^2}{\partial \alpha \partial \alpha^*} + \frac{1}{4} \frac{\kappa}{2} \left(\frac{\partial^3}{\partial \alpha^2 \partial \beta^*} + \frac{\partial^3}{\partial \alpha^* \partial \beta} \right) \right\} W. \end{aligned} \quad (3.38)$$

In the truncated Wigner representation, the derivatives are truncated at the second order.

Now the photon number and the half $\bar{n} + \frac{1}{2}$ appears while \bar{n} appeared in the Fokker-Planck equation in the P representation, where the factor of $1/2$ is due to the damped coherent state. The Wigner distribution for a damped coherent state is given by

$$W(\alpha, \alpha^*, t)_{\rho(0)=|\alpha_0\rangle\langle\alpha_0|} = \frac{1}{\pi[\frac{1}{2} + \bar{n}(1 - e^{-\gamma t})]} \exp \left[-\frac{|\alpha - \alpha_0 e^{-(\gamma/2)t} e^{-i\omega t}|^2}{\frac{1}{2} + \bar{n}(1 - e^{-\gamma t})} \right], \quad (3.39)$$

with the initial coherent state $W(\alpha, \alpha^*, 0)_{\rho(0)=|\alpha_0\rangle\langle\alpha_0|}$. The factor of $1/2$ is from the boson commutation relation by normal order $\hat{a}\hat{a}^\dagger = \frac{1}{2}(\hat{a}^\dagger\hat{a} + \hat{a}\hat{a}^\dagger) = \frac{2}{\pi}e^{-2|\alpha-\alpha_0|^2}$.

Note that the truncation in the Wigner representation is valid in the small quantum noise limit, it gives a good approximation so it can produce the same variance as the positive P representation only with statistical error originated from the finite number of samples [136]. The treatment of the third-order derivatives can also be found in the Reference [137].

3.5 Langevin Equation

The stochastic differential equations are derived from the Fokker-Planck equations for simulating sampling paths. Einstein's observation of Brownian motion is formulated by Langevin, et al., which has the form

$$dX_t = a(t, X_t)dt + b(t, X_t)\xi_t dt \quad (3.40)$$

with the deterministic drift term a and a noisy diffusive term $b\xi_t$, where the ξ_t is the standard Gaussian random variables for each t and b . This can be interpreted as the integral form

$$X_t(\omega) = X_{t_0}(\omega) + \int_{t_0}^t a(s, X_s(\omega))ds + \int_{t_0}^t b(s, X_s(\omega))\xi_s(\omega)ds \quad (3.41)$$

for each sample path. The two-time covariance $c(t) = E(\xi_s\xi_{s+t})$ has a constant spectral density, which means that all time frequencies are equally weighted in any Fourier transform of $c(t)$, that is the Gaussian white noise. When $a = 0, b = 1$, the process ξ_t should be pure Brownian motion, thus we can rewrite as

$$X_t(\omega) = X_{t_0}(\omega) + \int_{t_0}^t a(s, X_s(\omega))ds + \int_{t_0}^t b(s, X_s(\omega))dW_s(\omega). \quad (3.42)$$

Here, the Wiener process W_t is not differentiable, so the Ito integral is introduced as:

$$X_t(\omega) = \int_{t_0}^t f(s, \omega)dW_s(\omega), \quad (3.43)$$

which is \mathcal{A}_t measurable (we denote the $\{\mathcal{A}_t, t \leq 0\}$ is an increasing family of σ -additive class) and zero mean mean-square integrable and

$$E(X_t^2) = \int_{t_0}^t E(f, (s, \cdot)^2)ds. \quad (3.44)$$

Ito formula has the form

$$dY_t = \left\{ \frac{\partial U}{\partial t} + e \frac{\partial U}{\partial x} + \frac{1}{2} f^2 \frac{\partial^2 U}{\partial x^2} dt + f \frac{\partial U}{\partial x} dW_t \right\} \quad (3.45)$$

with the partial derivatives of U evaluated at (t, X_t) . When U is linear in x , $\frac{\partial^2 U}{\partial x^2} = 0$ is satisfied, hence the equation becomes

$$dY_t = \frac{\partial U}{\partial t}(t, U_t)dt + \frac{\partial U}{\partial x}(t, X_t)dX_t. \quad (3.46)$$

In general, a Fokker-Planck equation with the drift $f(x)$ and diffusion $D(x)$ terms

$$\frac{\partial}{\partial t} P(x, t) = \left[\frac{\partial}{\partial x} f(x) + \frac{\partial^2}{\partial x^2} D(x) \right] P(x, t) \quad (3.47)$$

can be expressed as the corresponding Langevin equation

$$dx = -\frac{df(x)}{dx}dt + \sqrt{2D(x)}dW. \quad (3.48)$$

The Langevin equation of a single DOPO is derived from the Fokker-Planck equations in the positive P (3.36) and Wigner representations (3.38). Note that the P representation (3.31) is not valid due to the diffusion matrix. Former is found at the beginning of the previous section (Section 3.4.1). The Langevin equation for the truncated Wigner representation becomes:

$$dc = (-1 + p - c^2 - s^2)c dt + \frac{1}{A_s} \sqrt{c^2 + s^2 + \frac{1}{2}} dW_c, \quad (3.50)$$

$$ds = (-1 - p - c^2 - s^2)s dt + \frac{1}{A_s} \sqrt{c^2 + s^2 + \frac{1}{2}} dW_s, \quad (3.51)$$

where c and s are in-phase and quadrature phase amplitudes, p is normalized pump rate, A_s is the steady state amplitude at $p = 2$ and dW_c, dW_s are independent Wiener processes. We basically use the above set of functions in the following analysis because it is simple and easy to treat. In the next chapter, the coupled and discretized formulation will be presented to fit the real experimental system. We can go back to the Fokker-Planck equations or the formulation in the other representations anytime we need the further analysis of the system. Note that the Langevin dynamics based on the above equation can converge to its global optimal states [138].

3.6 Typical Properties of DOPO

3.6.1 Pump-Signal Relation

The relation between the input pump rate and the signal photon number in the steady state is shown in Figure 3.3. When the DOPO is pumped below the threshold, the cavity has only a few (less than 1) photon. Then the number of photons dramatically increases at the threshold and the gap of this jump depends on the design of the cavity.

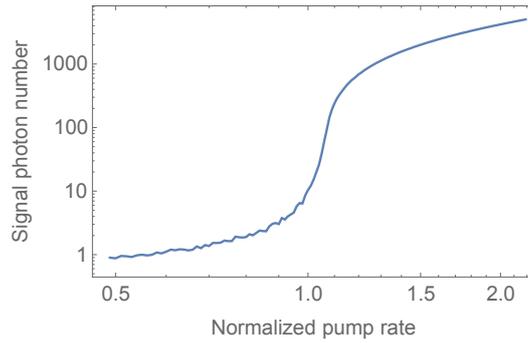


Figure 3.3. The number of signal photons as the function of the pump rate. It jumps around the threshold $p = p_{\text{th}}$, where the gap between the before and after the jump is determined by the saturation parameter $A_s^2 = \gamma_p \gamma_s / 2\kappa^2$.

3.6.2 Pitchfork Bifurcation

Without the noise term, steady states in (3.50) and (3.51) can be obtained with setting $dc/dt = ds/dt = 0$. Here, the pump rate p works as a bifurcation parameter. When $0 \leq p \leq 1$, the system has a steady state around $(c, s) = (0, 0)$. As the parameter increased to $p > 1$, the origin becomes unstable and two stable points appear $(c, s) = (\pm\sqrt{p-1}, 0)$. The equation is a typical case to show such (supercritical) pitchfork bifurcation.

3.6.3 Analytical Solution without Noise

If we ignore the noise terms in (4.19a1,2), we can find a sigmoid type function of gain I/O profile

$$g(c) = \frac{c}{\sqrt{1+c^2}}, \quad (3.52)$$

where c is the input amplitude. The time evolution is

$$c(t) = \text{sgn}(c_0) \sqrt{\frac{c_0^2}{c_0^2 + (1 - c_0^2)e^{-2t}}}, \quad (3.53)$$

where c_0 is the initial amplitude. Here, both equations are normalized with $p = 2$ (see Figure 3.4).

3.6.4 Steady State Distribution

The probability density function of the in-phase component x is obtained from the Fokker-Planck equation of positive P representation,

$$\text{Pb}(x) = \text{tr}(|x\rangle\langle x| \rho) = \sqrt{\frac{2}{\pi}} \iint d^2\alpha d^2\beta \exp \left[2 \left(x - \frac{\alpha + \beta}{2} \right) \right] P(\alpha, \beta), \quad (3.54)$$

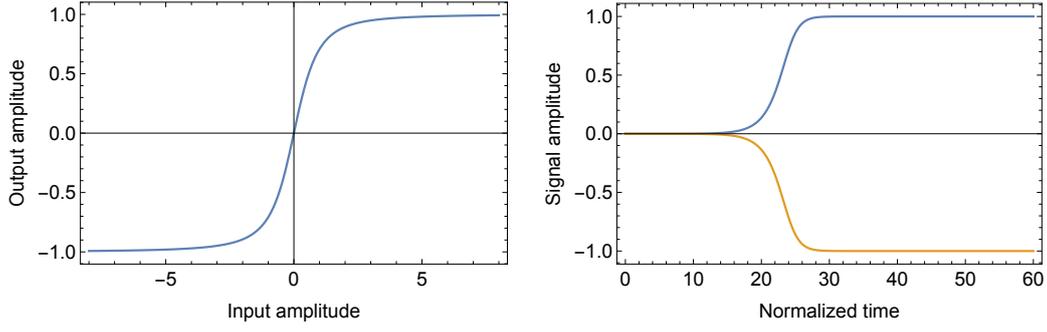


Figure 3.4. (a) Input-output relation in the DOPO gain process in (3.52) and the time evolution (3.53) when the pump rate $p = 2$ and the initial value $c_0 = 1/70^2$.

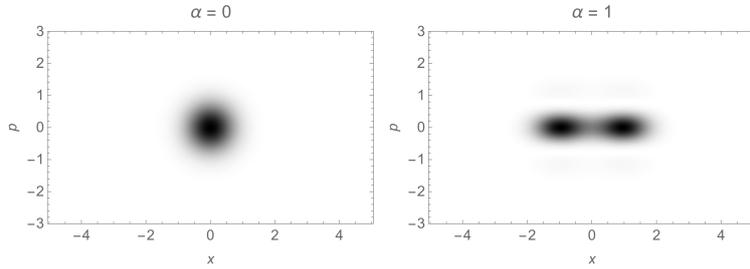


Figure 3.5. The signal field when the amplitudes are $\alpha = 0, 1$ described in the analytical form of (3.55a) and (3.55b). Note this parameter region is assumed to be the very low loss (high Q) cavity, so that the entanglement is observed as the interference in the quadrature amplitude direction p .

which can also be expanded as

$$P_\alpha(x) = \sqrt{\frac{2}{\pi}} \frac{1}{1 \pm \exp -2\alpha^2} \left[\pm e^{-2(x^2+\alpha^2)} + \frac{1}{2} e^{-2(x+\alpha)^2} + \frac{1}{2} e^{-2(x-\alpha)^2} \right], \quad (3.55a)$$

$$P_\alpha(y) = \sqrt{\frac{2}{\pi}} \frac{1}{1 \pm \exp -2\alpha^2} \exp^{-2y^2} [1 \pm \cos(4y\alpha)], \quad (3.55b)$$

as shown in Figure 3.5. If we ignore the multiplicative noise, the steady state will be the Boltzmann distribution [139]. In the very simple form, it only contains the in-phase component x with pump rate p normalized by the steady state amplitude A_s :

$$P(x, p) = \exp \left[\frac{1}{2A_s^2} (p - 1)x^2 - \frac{1}{4A_s^2} x^4 \right]. \quad (3.56)$$

3.7 DOPO in the Optical Fiber

The DOPO in our current system is generated in an optical fiber ring. Signal DOPOs are synchronously pumped by a pulsed laser which induces the $\chi^{(2)}$ parametric process in a nonlinear PPLN crystal [140]. Thanks to the low loss fiber (around the wavelength of

1.55 μm [140]) and phase stability by cavity control, recently a large number of optical pulses (more than 10000 pulses) can simultaneously oscillate in a long fiber [72]. This is a key for scaling up the optical part.

Chapter 4

Coherent Ising Machine

In this chapter, coherent Ising machine, a physical system which exploits collective behavior of optical oscillators, is described. After basic description based on previous research, the scalable implementation for CIM with measurement feedback will be introduced, which is first proposed in

- [73] Y. Haribara, S. Utsunomiya, and Y. Yamamoto. Computational principle and performance evaluation of coherent Ising machine based on degenerate optical parametric oscillator network. *Entropy*, 18(4):151, 2016.

The effect of using optical fiber instead of a free-space cavity is discussed in Appendix B.

4.1 Overview

We intend to solve combinatorial optimization problems by mapping the cost function eq (2.1) to the energy of an Ising spin system. CIM is initially proposed as an injection-locked laser system [68], followed by the proposal using a degenerate optical parametric oscillator (DOPO) system [69]. So far, several experimental machines are demonstrated with $n = 4, 16, 100, 2048$ -pulse systems [70, 71, 76, 77]. Since the original MAX-CUT has binary variables, we use a bistable optical device, DOPO at the output stage of computation, while an analog optical device, degenerate optical parametric amplifier (DOPA), at the solution search stage of computation.

Figure 4.1 depicts the schematic of the measurement feedback based CIM [76, 77], which is initially proposed in Reference [73]. Here we describe the typical experimental configurations in [77]. The DOPO part consists of a 1 km optical fiber (round trip time of 5 μ s) with an externally pumped periodically poled lithium niobate (PPLN) waveguide. The pulsed pump laser, at the 1 GHz repetition rate of 5000 times as the cavity circulation frequency, generates 5000 individual DOPO pulses in a single fiber ring cavity. A segment of them (2000 pulses) is used as the signal pulses for computation and the remaining portion (3000 pulses) is used for the cavity stabilization.

The feedback circuit stores the interaction strength for each pair of DOPO pulses. A portion of the optical pulse is picked-off by a beamsplitter (numbered as 1 in Figure 4.1)

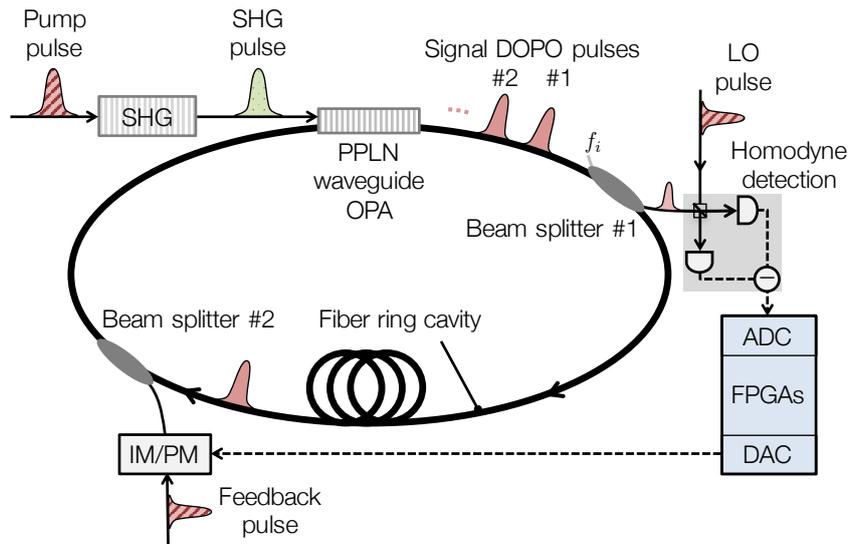


Figure 4.1. Experimental schematic of a measurement-feedback coherent Ising machine (MF-CIM) implemented on a fiber DOPO with an FPGA measurement feedback circuit. Small portion of each signal pulse is out-coupled through the beam splitter 1, and its in-phase component is measured by optical balanced homodyne detector, where LO pulse is directly obtained from the pump laser. Two detector outputs are converted to digital signals and input into an electronic digital circuit, where a feedback signal for i -th signal pulse is computed. Independently obtained feedback pulse from the master laser is modulated in its intensity and phase to achieve $\sum_j \xi_{ij} \tilde{c}_j$ and coupled into i -th signal pulse by directional coupler 2. Flows of optical fields and electrical signals are shown as solid and dashed lines, respectively.

and measured by balanced homodyne detectors. The measured values of DOPO pulse amplitudes are fed into an analog-digital converter (ADC), followed by FPGAs. Here, 1 GHz repetition rate of signal pulses is downclocked to 125 MHz (8 parallel) and the measured amplitudes \tilde{c}_i are sliced into the digital signals of 5 bits. Then 2 FPGAs sum up the coupling effect from the other vertices (in the given topology) $\sum_j J_{ij}\tilde{c}_j$ for the i th pulse. The feedback pulse train is modulated in intensity and phase by this output electrical signal after a digital-analog converter (DAC). The feedback pulse is injected to the signal DOPO pulse running through the main fiber ring cavity via a beamsplitter #2.

The DOPO is operated near the oscillation threshold by crossing the pump rate from below to above the threshold in the case of [77]. In the beginning, the DOPO is biased at below the threshold in which all phase configuration is established so as a superposition state and the quantum parallel search is implemented [136]. Then, the external pump rate (or the feedback) strength is gradually increased, and once the whole system reaches the oscillation threshold, it selects a particular phase configuration which corresponds to the near-optimal solution of the original optimization 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 fits for this purpose is the positive $P(\alpha, \beta)$ representation in terms of the off-diagonal coherent state expansion, $|\alpha\rangle\langle\beta|$. 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 [141] 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 [136]. We will use the latter approach in this thesis to evaluate the performance of the CIM.

4.2 Quantum Formulation of CIM

Extending (3.2), The total Hamiltonian of the coherent Ising machine (Figure 4.1) is described by multiple DOPO Hamiltonians with mutual coupling:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{\text{free}} + \hat{\mathcal{H}}_{\text{param}} + \hat{\mathcal{H}}_{\text{pump}} + \hat{\mathcal{H}}_{\text{couple}} + \hat{\mathcal{H}}_{\text{SR}} \quad (4.1)$$

$$\hat{\mathcal{H}}_{\text{free}} = \hbar\omega_s \sum_{j=1}^n \hat{a}_{sj}^\dagger \hat{a}_{sj} + \hbar\omega_p \sum_{j=1}^n \hat{a}_{pj}^\dagger \hat{a}_{pj} + \hbar\omega_s \sum_{j=1}^n \sum_{k \neq j} \hat{a}_{cjk}^\dagger \hat{a}_{cjk}, \quad (4.2a)$$

$$\hat{\mathcal{H}}_{\text{param}} = \frac{i\hbar\kappa}{2} \sum_{j=1}^n (\hat{a}_{sj}^{\dagger 2} \hat{a}_{pj} - \hat{a}_{pj}^\dagger \hat{a}_{sj}^2), \quad (4.2b)$$

$$\hat{\mathcal{H}}_{\text{pump}} = i\hbar \sum_{j=1}^n (\varepsilon \hat{a}_{pj}^\dagger - \varepsilon \hat{a}_{pj}), \quad (4.2c)$$

$$\hat{\mathcal{H}}_{\text{couple}} = i\hbar\zeta \sum_{j=1}^n \sum_{k \neq j} (\hat{a}_{cjk} \hat{a}_{sj}^\dagger - \hat{a}_{cjk}^\dagger \hat{a}_{sj} + \hat{a}_{sk} \hat{a}_{cjk}^\dagger e^{-ikcz} - \hat{a}_{sk}^\dagger \hat{a}_{cjk} e^{ikcz}), \quad (4.2d)$$

$$\hat{\mathcal{H}}_{\text{SR}} = \hbar \sum_{j=1}^n (\hat{a}_{sj} \hat{\Gamma}_{\text{Rs}j}^\dagger + \hat{\Gamma}_{\text{Rs}j} \hat{a}_{sj}^\dagger + \hat{a}_{pj} \hat{\Gamma}_{\text{Rp}j}^\dagger + \hat{\Gamma}_{\text{Rp}j} \hat{a}_{pj}^\dagger) + \hbar \sum_{j=1}^n \sum_{k \neq j} (\hat{a}_{cjk} \hat{\Gamma}_{\text{Rc}}^\dagger + \hat{a}_{cjk}^\dagger \hat{\Gamma}_{\text{Rc}}), \quad (4.2e)$$

where $\hat{\mathcal{H}}_{\text{free}}$ is the free field Hamiltonian for the signal, pump, and coupling fields, $\hat{\mathcal{H}}_{\text{param}}$ is the parametric interaction Hamiltonian, $\hat{\mathcal{H}}_{\text{pump}}$ is the external pumping Hamiltonian where ε is the real-number pump field amplitude, $\hat{\mathcal{H}}_{\text{couple}}$ is the coupling Hamiltonian among n DOPOs, $\hat{\mathcal{H}}_{\text{SR}}$ is the system-reservoir interaction Hamiltonian which describes any spurious dissipation processes for the signal, pump, and coupling fields. In (4.2d), the phase factors of the coupling field represent the in-phase or out-of-phase coupling from the j -th DOPO pulse to the k -th DOPO pulse. The ferromagnetic and anti-ferromagnetic couplings are realized when $e^{ikcz} = e^{-ikcz} = 1$ and $e^{ikcz} = e^{-ikcz} = -1$, respectively. The two dominating terms in the Hamiltonian are the parametric coupling term $\mathcal{H}_{\text{param}}$ and the mutual coupling term between different pulses $\mathcal{H}_{\text{couple}}$. The former term creates the squeezed vacuum state in each DOPO pulse, while the latter term modulates the effective loss according to the given problem so that the Ising Hamiltonian is mapped to the network loss.

4.2.1 Positive P representation

We can derive the master equation for the total field density operator $\hat{\rho}$ using Equations (4.1) - (4.2e) and expand it in terms of the positive P representation in the same way [135]:

$$\hat{\rho} = \iint d^2\alpha d^2\beta P(\alpha, \beta) \frac{|\alpha\rangle\langle\beta^*|}{\langle\beta^*|\alpha\rangle}, \quad (4.3)$$

where

$$\alpha := (\alpha_{s1}, \dots, \alpha_{sn}, \alpha_{p1}, \dots, \alpha_{pn}, \alpha_{c12}, \dots, \alpha_{cn\ n-1})^\top$$

and

$$\beta := (\beta_{s1}, \dots, \beta_{sn}, \beta_{p1}, \dots, \beta_{pn}, \beta_{c12}, \dots, \beta_{cn\ n-1})^\top$$

are the vectors with complex eigenvalues,

$$|\alpha\rangle = |\alpha_{s1}\rangle \cdots |\alpha_{sn}\rangle |\alpha_{p1}\rangle \cdots |\alpha_{pn}\rangle |\alpha_{c12}\rangle \cdots |\alpha_{cn\ n-1}\rangle$$

and

$$|\beta\rangle = |\beta_{s1}\rangle \cdots |\beta_{sn}\rangle |\beta_{p1}\rangle \cdots |\beta_{pn}\rangle |\beta_{c12}\rangle \cdots |\beta_{cn\ n-1}\rangle$$

are the multimode coherent states. Here α_X and β_X are statistically independent, but their ensemble averaged values must satisfy $\langle \alpha_X \rangle = \langle \beta_X \rangle^*$. This off-diagonal $|\alpha\rangle \langle \beta|$ positive $P(\alpha, \beta)$ representation of the field density operator $\hat{\rho}$ allows to describe an arbitrary non-classical field, while the diagonal $|\alpha\rangle \langle \alpha|$ Glauber-Sudarshan $P(\alpha)$ representation [142] can describe only classical fields or statistical mixture of coherent states. We substitute (4.3) into the master equation to obtain the Fokker-Planck equation for $P(\alpha, \beta)$ [134]. Then we can derive the c-number SDE (CSDE) using the Ito rule [127]:

$$d\alpha_{sk} = (-\gamma_s \alpha_{sk} + \kappa \alpha_{pk} \beta_{sk} + \sum_{j \neq k} \zeta_{jk} \alpha_{cjk}) dt + \sqrt{\kappa \alpha_{pk}} dw_{\alpha_{sk}}(t), \quad (4.4a)$$

$$d\beta_{sk} = (-\gamma_s \beta_{sk} + \kappa \beta_{pk} \alpha_{sk} + \sum_{j \neq k} \zeta_{jk} \beta_{cjk}) dt + \sqrt{\kappa \alpha_{pk}} dw_{\beta_{sk}}(t), \quad (4.4b)$$

$$d\alpha_{pk} = (-\gamma_p \alpha_{pk} - \frac{\kappa}{2} \alpha_{sk}^2 + \varepsilon) dt, \quad (4.4c)$$

$$d\beta_{pk} = (-\gamma_p \beta_{pk} - \frac{\kappa}{2} \beta_{sk}^2 + \varepsilon) dt, \quad (4.4d)$$

$$d\alpha_{cjk} = (-\gamma_c \alpha_{cjk} - \zeta_{jk} \alpha_{sj} + \zeta_{jk} e^{i\theta} \alpha_{sk}) dt, \quad (4.4e)$$

$$d\beta_{cjk} = (-\gamma_c \beta_{cjk} - \zeta_{jk} \beta_{sj} + \zeta_{jk} e^{-i\theta} \beta_{sk}) dt. \quad (4.4f)$$

The positive P representation can be obtained by ensemble averaging over many trajectories generated by Monte Carlo numerical integration of (4.4a - 4.4f). Typically, we need 10^5 - 10^6 trajectories for obtaining reasonable convergence around the oscillation threshold. In Reference [141], quantum entanglement and quantum discord in gradually pumped two coupled DOPOs are studied in the positive P representation.

4.2.2 Truncated Wigner Representation

It is well-known that even though the positive- P representation method is rigorous and can treat arbitrary non-classical states, the convergence requires a huge computation time. If the quantum state of light in a given physical system is only slightly deviated from the Gaussian states, the truncated Wigner representation is an alternative approach with reasonable accuracy [126]. In this case, the field density operator is expanded by the Wigner function $W(\alpha)$:

$$\hat{\rho} = \int d\lambda e^{\lambda^* \hat{a} - \lambda \hat{a}^\dagger} \left\{ \int d\alpha e^{\lambda \hat{a}^* - \lambda^* \hat{a}} W(\alpha) \right\}, \quad (4.5)$$

where

$$\hat{a} = (\hat{a}_{s1}, \dots, \hat{a}_{sn}, \hat{a}_{p1}, \dots, \hat{a}_{pn}, \hat{a}_{c12}, \dots, \hat{a}_{cn\ n-1})^\top$$

and

$$\lambda = (\lambda_{s1}, \dots, \lambda_{sn}, \lambda_{p1}, \dots, \lambda_{pn}, \lambda_{c12}, \dots, \lambda_{cn\ n-1})^\top.$$

Here, $\chi(\lambda) = \int d\alpha e^{\lambda\hat{a}^* - \lambda^*\hat{a}} W(\alpha)$ is the symmetric correlation function, where $W(\alpha)$ and $\chi(\lambda)$ form a pair of Fourier transform. The Fokker-Planck equation for $W(\alpha)$ can be derived by truncating the third and higher-order derivatives, which gives another set of CSDEs:

$$d\alpha_{sk} = (-\gamma_s\alpha_{sk} + \kappa\alpha_{pk}\alpha_{sk}^* + \sum_{j \neq k} \zeta_{jk}\alpha_{cjk})dt + \sqrt{\gamma_s}dW_{sk}(t), \quad (4.6a)$$

$$d\alpha_{pk} = (-\gamma_p\alpha_{pk} - \frac{\kappa}{2}\alpha_{sk}^2 + \varepsilon)dt + \sqrt{\gamma_p}dW_{pk}(t), \quad (4.6b)$$

$$d\alpha_{cjk} = (-\gamma_c\alpha_{cjk} - \zeta_{jk}\alpha_{sj} + \zeta_{jk}e^{i\theta}\alpha_{sk})dt + \sqrt{\gamma_c}dW_{ck}(t). \quad (4.6c)$$

Here, $dW_X(t)$ is the complex valued Wiener process and corresponds to the quantum fluctuation injected into the system from coupled external reservoirs. In Reference [136], the degree of quantum entanglement is checked against the positive P representation. It partly confirmed that the truncation in the equation is valid within the statistical error introduced by the finite number of numerically generated trajectories in the small case of $n = 2$ and $n = 16$.

Now we ignore the equation of pump by adiabatic elimination and obtain the CSDE:

$$dc_i = \left[(-1 + p - c_i^2 - s_i^2)c_i + \sum_{j=1}^n \xi_{ij}c_j \right] dt + \frac{1}{A_s} \sqrt{c_i^2 + s_i^2 + \frac{1}{2}} dW_{c_i}, \quad (4.8)$$

$$ds_i = \left[(-1 - p - c_i^2 - s_i^2)s_i + \sum_{j=1}^n \xi_{ij}s_j \right] dt + \frac{1}{A_s} \sqrt{c_i^2 + s_i^2 + \frac{1}{2}} dW_{s_i}, \quad (4.9)$$

where c and s are in-phase and quadrature phase amplitudes, p is normalized pump rate, A_s is the steady state amplitude at $p = 2$ and dW_1, dW_2 are independent Wiener processes.

4.2.3 Measurement Effect in Measurement Feedback CIM

If we take the measurement effect explicitly, the quantum master equations should be modified as explained in References [143, 144]. While the comparison on the computational performance will be a future work and not fully considered in this thesis, the model overview is according to Reference [143].

The nonunitary reduction of a wave function by the homodyne measurement of x is described by Wiseman and Milburn [145]

$$\frac{d}{dt}\hat{\rho}_{\text{mes}} = \sum_{i=1}^n \left[\frac{\gamma_{\text{mes}}}{2} (2\hat{a}_{si}\hat{\rho}\hat{a}_{si}^\dagger - \hat{a}_{si}^\dagger\hat{a}_{si}\hat{\rho} - \hat{\rho}\hat{a}_{si}^\dagger\hat{a}_{si}) + \sqrt{\gamma_{\text{mes}}} \frac{dW_i}{dt} (\hat{a}_{si}\hat{\rho} + \hat{\rho}\hat{a}_{si}^\dagger - \langle c_i \rangle \hat{\rho}) \right], \quad (4.10)$$

where $\langle c_i \rangle = \langle \hat{a}_{si}^\dagger + \hat{a}_{si} \rangle$ is the expectation value of the in-phase amplitude of the signal field and dW_i is the Wiener increment, which satisfies

$$dW_i(t) \sim \mathcal{N}(0, dt), \quad \langle dW_i(t)dW_j(t') \rangle = 2\pi\delta_{ij}\delta(t-t'). \quad (4.11)$$

This model assumes the measured value has the stochasticity as

$$\tilde{c}_i dt = \langle \hat{a}_{si}^\dagger + \hat{a}_{si} \rangle dt + \frac{dW_i}{\sqrt{\gamma_{\text{mes}}}}. \quad (4.12)$$

Then the total master equation is the sum of (3.7) and (4.10).

Following the derived master equation, the corresponding partial differential equation in positive P representation becomes as

$$\begin{aligned} \frac{\partial P(\alpha, \beta)}{\partial t} = & \left[\sqrt{\gamma_{\text{mes}}}(\alpha_{si} + \beta_{si} - \langle \alpha_{si} + \beta_{si} \rangle) \frac{dW}{dt} \right. \\ & - \frac{\partial}{\partial \alpha_{si}} (-(\gamma_s + \gamma_{\text{mes}})\alpha_{si} + \kappa\beta_{si}\alpha_{pi} + \varepsilon_{si}) \\ & - \frac{\partial}{\partial \beta_{si}} (-(\gamma_s + \gamma_{\text{mes}})\beta_{si} + \kappa\alpha_{pi}\beta_{pi} + \varepsilon_{si}) \\ & - \frac{\partial}{\partial \alpha_{pi}} (-\gamma_p\alpha_{pi} + \frac{\kappa}{2}\alpha_{si}^2 + \varepsilon_{pi}) - \frac{\partial}{\partial \beta_{pi}} (-\gamma_p\beta_{pi} + \frac{\kappa}{2}\beta_{si}^2 + \varepsilon_{pi}) \\ & \left. + \frac{\partial^2}{\partial \alpha_{si}^2} \kappa\alpha_{pi} + \frac{\partial^2}{\partial \beta_{si}^2} \kappa\beta_{pi} \right] P(\alpha, \beta), \end{aligned} \quad (4.13)$$

where

$$\langle \alpha_{si} \rangle = \iint d^2\alpha d^2\beta \alpha_{si} P(\alpha, \beta), \quad (4.14a)$$

$$\langle \beta_{si} \rangle = \iint d^2\alpha d^2\beta \beta_{si} P(\alpha, \beta). \quad (4.14b)$$

The (4.13) is similar to the Fokker-Planck equation except for the first term in the right hand side. After the adiabatic elimination, the corresponding SDE becomes [143]:

$$da_i = [-(1 + \gamma')a_i + b_i(p - a_i^2) + f_i]dt + \frac{1}{A_s} \sqrt{p - a_i^2} dW_{a_i}, \quad (4.15)$$

$$db_i = [-(1 + \gamma')b_i + a_i(p - b_i^2) + f_i]dt + \frac{1}{A_s} \sqrt{p - b_i^2} dW_{b_i}, \quad (4.16)$$

where normalized amplitudes $a_i = \alpha_{si}/A_s$, $b_i = \beta_{si}/A_s$, normalized time $dt = \gamma_s dt'$, normalized pump rate $p = \kappa\varepsilon_i/\gamma_s\gamma_p$ and effective loss $\gamma' = \gamma_{\text{mes}}/\gamma_s$.

As a feature of this formulation, the first term in (4.13) is expressed as the replicator equation by extending the branching Brownian motion [146]. In this model, the change in $P(\alpha, \beta)$ is governed by

$$\frac{\partial P(\alpha)}{\partial t} = \lambda(\alpha, \beta)P(\alpha, \beta), \quad (4.17)$$

where the $\lambda(\alpha, \beta) = \tilde{c}_i - \langle c_i \rangle$ determines the replication or deletion probability of the Brownian particle at (α, β) , in the case of $\lambda > 0$ or $\lambda < 0$, respectively.

In this model, the back action of the measurement produces non-Gaussian states. It will be interesting to compare the performance on large-sized graph instances.

4.2.4 Computational Principle

The computational process of CIM is described. There are still some arguments and the unified view is absent. Note that, there is no evidence yet that the quantum effects intrinsically contribute to the computational process, which will be discussed in Section 8.4.1.

In the collective dynamics of quantum mechanical harmonic oscillators, the whole coupled DOPO system has the total loss function [69]

$$\Gamma = n - \sum_{i=1, i \neq j}^n \xi_{ij} c_i c_j + O\left(\frac{\epsilon n^4}{(p-1)^3}\right), \quad (4.18)$$

where $\epsilon = \max_{1 \leq i, j \leq n} |\xi_{ij}|$ appears from perturbation theory. The second term is an Ising Hamiltonian (1.2). Hence, (4.18) achieves the minimum loss when the system has the ground state configuration, assuming the pump rate is high enough $p \gg 1$ and all DOPOs have the unique amplitude.

Generally, the amplitude homogeneity breaks down in computational process. Though, CIM can find low energy solutions as the results in the following chapters show. Since the CIM embed binary spin values to continuous valued DOPO amplitudes, it can be interpreted as low-rank relaxation heuristic, followed by soft binalization by nonlinear gain saturation. While the dimension is not high enough as the SDP convex relaxation, a low dimensional relaxation is reported to show a good performance as a MAX-CUT heuristic [147].

4.3 Discrete Feedback Simulation with Ring Cavity

The above natural formulation assumes that the all physical process, e.g., OPA gain, fiber and beam splitter (BS) loss, feedback, occurs simultaneously. In the more realistic situation with time time-division multiplexing (TDM) scheme where the optical pulses are circulating in the main cavity, the BS loss and feedback should be occurred at the specific point namely BS 1 and BS 2, respectively, in the ring cavity. In our simulation, the OPO gain in the PPLN waveguide, out-coupling loss in the fiber BS for measurement circuit, and injection after the feedback calculation are taken in separately as

$$dc_i = [(p - c_i^2 - s_i^2)c_i] dt + \frac{1}{A_s} \sqrt{c_i^2 + s_i^2 + \frac{1}{2}} dW_{c_i}, \quad (4.19a1)$$

$$ds_i = [(-p - c_i^2 - s_i^2)s_i] dt + \frac{1}{A_s} \sqrt{c_i^2 + s_i^2 + \frac{1}{2}} dW_{s_i}, \quad (4.19a2)$$

$$c_i(t + \Delta t) \mapsto \sqrt{1 - T_{\text{mes}}} c_i(t) + \sqrt{T_{\text{mes}}} \frac{f_i}{A_s}, \quad (4.19b1)$$

$$s_i(t + \Delta t) \mapsto \sqrt{1 - T_{\text{mes}}} s_i(t) + \sqrt{T_{\text{mes}}} \frac{f_i}{A_s}, \quad (4.19b2)$$

$$c_i(t + \Delta t) \mapsto \sqrt{1 - T_{\text{inj}}} c_i(t) + \sqrt{T_{\text{inj}}} \xi \sum_{j=1}^n J_{ij} \tilde{c}_j, \quad (4.19c1)$$

$$s_i(t + \Delta t) \mapsto \sqrt{1 - T_{\text{inj}}} s_i(t), \quad (4.19c2)$$

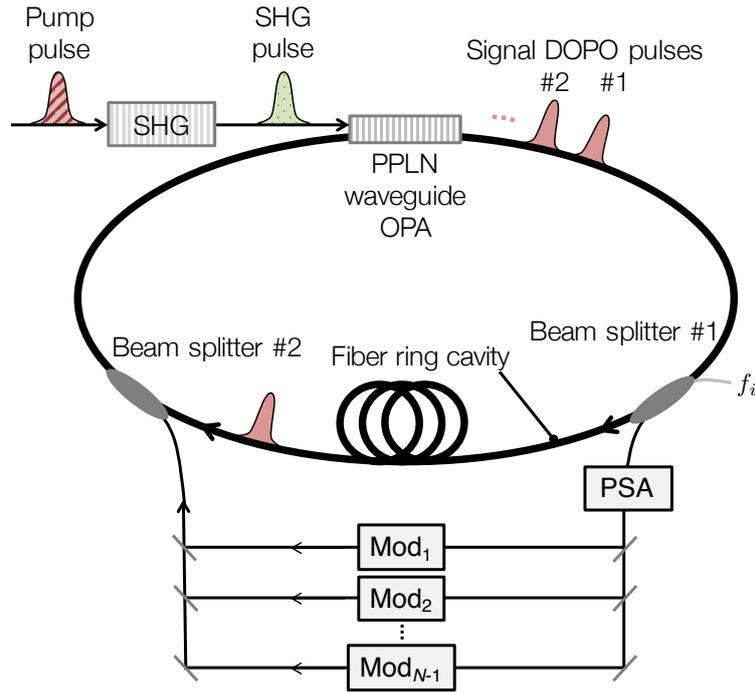


Figure 4.2. A coherent Ising machine implemented on the time-division multiplexed DOPO with mutual coupling implemented by optical delay lines (ODL-CIM). A part of each pulse is picked off from the main cavity by the output coupler followed by an optical phase sensitive amplifier (PSA) which amplifies the in-phase amplitude \tilde{c}_i of each DOPO pulse. The feedback pulses, which are produced by combining the outputs from $n-1$ intensity and phase modulators, are injected back to the main cavity by the injection coupler.

where

$$\tilde{c}_i = \frac{c_i^{\text{out}}}{\sqrt{T_{\text{mes}}}} = c_i \sqrt{\frac{1 - T_{\text{mes}}}{T_{\text{mes}}}} \frac{f_i}{A_s}$$

includes the noise invaded from the beam splitter 1. Here, (4.19a1,2) is the gain part, (4.19b1,2) is the loss in BS1 of transmittance T_{mes} , where f_j is a vacuum noise invaded from the open port of the beamsplitter and (4.19c1,2) is the injection part with BS2 of the transmittance T_{inj} after feedback pulse amplitudes ($\sum_i J_{ij} \tilde{c}_i$) are calculated in the FPGA.

This formulation is the same in the optical delay line coupled CIM (ODL-CIM) as depicted in Figure 4.2. As far as the current experimental region of parameters, these approximation and simplification are valid to predict the performance evaluation described in later chapters.

4.4 FPGA Design for Feedback Circuit

A field-programmable gate array (FPGA) is a configurable circuit which can be designed (“programmed”) by a software. It has hundreds of thousands of programmable blocks

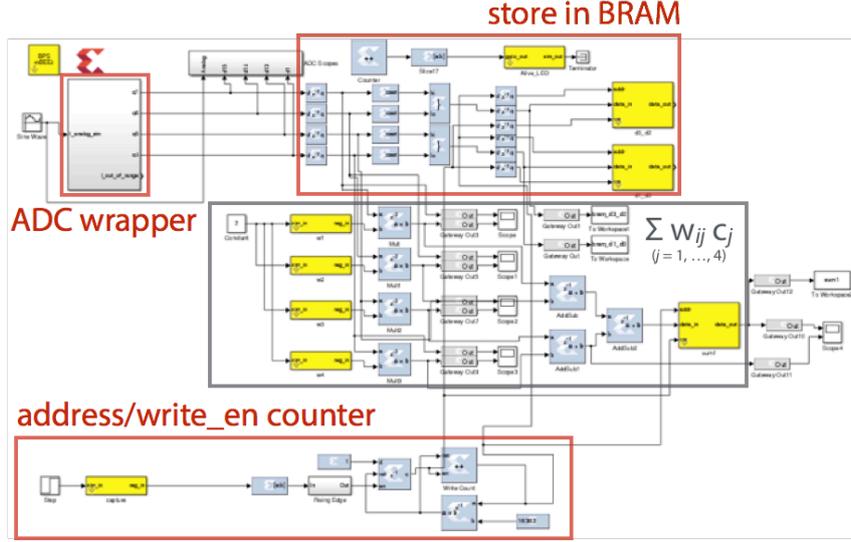


Figure 4.3. The FPGA design environment in GUI by BEEcube platform studio on Simulink and Xilinx ISE. The example circuit design show the part of the $n = 4$ feedback circuit.

called “logic cells”, which can perform the logical or arithmetical operation. To achieve the high throughput of the feedback calculation in CIM, we adopted feedback circuit design using FPGA. The fully managed clock was also the advantage to synchronize with the optical cavity, which cannot be done with GPU. It can be programmed with hardware description languages such as VHDL or Verilog, and there are several graphical user interface (GUI) programming environments as shown in Figure 4.3.

The main functionality of the FPGA in the system is to calculate the feedback part, i.e.,

$$\sum_{j=1}^n J_{ij} \tilde{c}_j, \quad (4.19e)$$

where J is the given coefficient matrix, \tilde{c} is the measured value of in-phase component. The sign and absolute value of the summation modulate the phase and intensity of the feedback pulse, respectively. The input stream is the measured value from the homodyne detection, whose sampling rate is equal to the pulse repetition frequency of the main ring cavity. The clock frequency of FPGA (few hundreds of MHz) is much slower than the pulse repetition rate in the current experimental setup (~ 1 GHz) so that the input

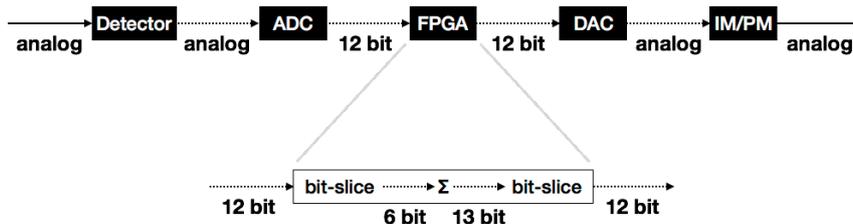


Figure 4.4. The example of the bit-slicing implementation in the feedback circuit. The input/output bit length of the FPGA is dependent on the spec of ADC/DAC. An FMC101 card is attached to the miniBEE configuration. The increase in the bit-length in the summation part is upper bounded by $\log_2(k)$, where $k < n = |V|$ is the vertex degree when we are solving unweighted graphs.

pulse stream should be downclocked. In our system, the input signals are 4-way or 8-way parallelized, i.e., ADC clock is downclocked to 1/4 or 1/8 and used as FPGA clock. In order to stabilize the system, the clock fed into the ADC (hence into the FPGA) and the pump pulse should be the same source.

In the FPGA circuit, the input amplitudes are sliced to low bits and the matrix-vector product is implemented in parallel. Of course, the feedback only calculated from a very low number of bits ends in the wrong results. On the other hand, the full precision detected in the detector is not practically suitable. The bit precision and circuit scale in the trade-off, which is discussed in the Appendix F. Then the sliced input amplitudes are used for the calculation. The example is presented in Figure 4.4.

After the summation, the bit-length of the feedback values is increased $\lceil \log_2(n) + \log_2(w - 1) \rceil$ from the input sliced bits. In the case of the complete graph of $n = 100$ and if the weight is 1 bit $w_{ij} = \pm 1$, the output bit length from sliced input 6 bit amplitudes swells up to $6 + \lceil \log(99) + \log(1) \rceil = 6 + \lceil 6.6 \rceil = 13$. Finally, this again sliced into the DAC bit length.

Chapter 5

Numerical Simulation

It is difficult to predict the performance of CIM analytically. Hence, numerical simulations are performed on the stochastic differential equation appeared in the previous chapter. The computational performance in terms of the MAX-CUT solver is evaluated

- [75] Y. Haribara, S. Utsunomiya, K. Kawarabayashi, and Y. Yamamoto. Quantum computing with non-classical light: A coherent ising machine for maximum-cut problems. In J. C. Lindon, G. E. Tranter, and D. W. Koppenaal, editors, *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, pages 824–831. Academic Press, Oxford, 2017,

which led to the experimental success in the next two chapters.

5.1 Implementation of Numerical Analysis

Based on the formulation in Section 4.3, numerical simulations are performed. From the simulation, DOPO in-phase component Equations (4.19a1), (4.19b1) and (4.19c1) are plotted since the quadrature components s_i ($i \in \{1, \dots, n\}$) are considerably small. For the numerical integration, Euler-Maruyama method was used to simulate the SDEs. The codes are implemented with C++11 compiled by GCC4.9. Most of the computation in this section is done with Linux server which has two 6-core Intel Xeon X5650 processors @ 2.67 GHz Westmere architecture, 94 GB DDR3 RAM and Ubuntu14.04 is installed.

Typical simulation parameters are summarized in Table 5.1. Most of them are fixed and automatically determined when the physical configuration is set, but two of them are tunable, namely, pump rate p and coupling constant ξ .

5.1.1 Parameter Optimization

To run a CIM with the best performance, the experimental parameters, especially for the pump rate p and the coupling constant ξ should be optimized. Note that, empirically, for the random graph whose degree distribution is a binomial distribution, the coupling constant is optimal around $\xi_{ij} = \xi w_{ij} / \sqrt{\langle k \rangle}$. For the complete graph, the coupling constant is $\xi_{ij} = \xi w_{ij} / n$ (it is natural in statistical mechanics since the energy diverges when $n \rightarrow \infty$).

Table 5.1. Parameter settings for numerical simulations. Basically, tunable parameters at the computation are the pump rate p and coupling strength ξ .

Parameter	Value (default)
Pump rate	$p > 0$ (1.1)
Coupling strength	$\xi < 0$ (-0.1)
Signal amplitude	$A_s = 70, 400, 10^7$ (70)
Signal photon loss	$\gamma_s = 1$
Pump photon loss	$\gamma_p = 100$
Parametric coupling constant	$\kappa = 0.1$
Number of cavity round trips	$n_{\text{RT}} > 0$ (200)
Readout coupler transmittance	$T_{\text{mes}} = 0.1$
Injection coupler transmittance	$T_{\text{mes}} = 0.01$
Time step size	$dt = 0.01$
Single round trip time steps	$t_{\text{RT}} = 10$
Precision	64-bit double

5.2 Single Run

Time Evolution

Figure 5.1 shows the simulation result for a single run. The CIM solves the MAX-CUT problem on the K_4 graph (in the Figure 2.1) and found the ground state (2-by-2 subsets of vertices of $\text{CUT} = 4$). In the magnified plot of Figure 5.1, we can see the change of the amplitudes in a single unit of time (circulation of a cavity). In the simulation for the single short window of time, A, B, C correspond to gain, loss, coupling expressed in Equations (4.19a1), (4.19b1) and (4.19c1), respectively.

Frustration

Figure 5.2 shows the frustrating case, where the initial spin configuration is wrong but finally converges to the ground state. In Figure 5.2, there is a turning point around $t = 140$ where three DOPOs $i = 1, 2, 3$ go to the final decision of the spin directions. The situation is clearer when we plot the DOPO amplitudes onto the graph as in Figure 5.2 (c-f). The notable point is that the right cluster of $i \in \{1, 2, 3\}$ is optimized in the latter half in the simulation. It is contributed by the nonlinear saturation in (4.19a1).

Relatively Large Instances

Larger instances are simulated. Figures 5.3 and 5.4 are the simulation results for MAX-CUT benchmark instances on sparse graphs (G-set) [148]. When the graph becomes larger, the injection coupling from the other OPOs conceal the binalized amplitudes. These figures show that MF-CIM can perform Ising energy minimization even for larger graphs as $n \leq 100$.

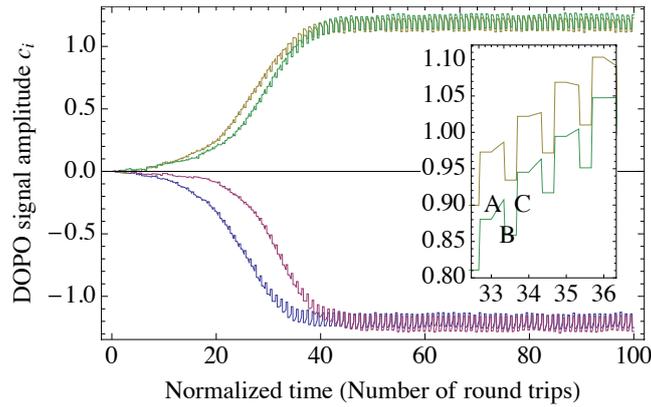


Figure 5.1. The time evolution of the signal DOPO in-phase amplitude c_i when solving the complete graph of order $n = 4$ presented in the Figure 2.1. In the magnified figure, A, B, and C indicate nonlinear saturable gain, out-coupler loss, and feedback injection, respectively.

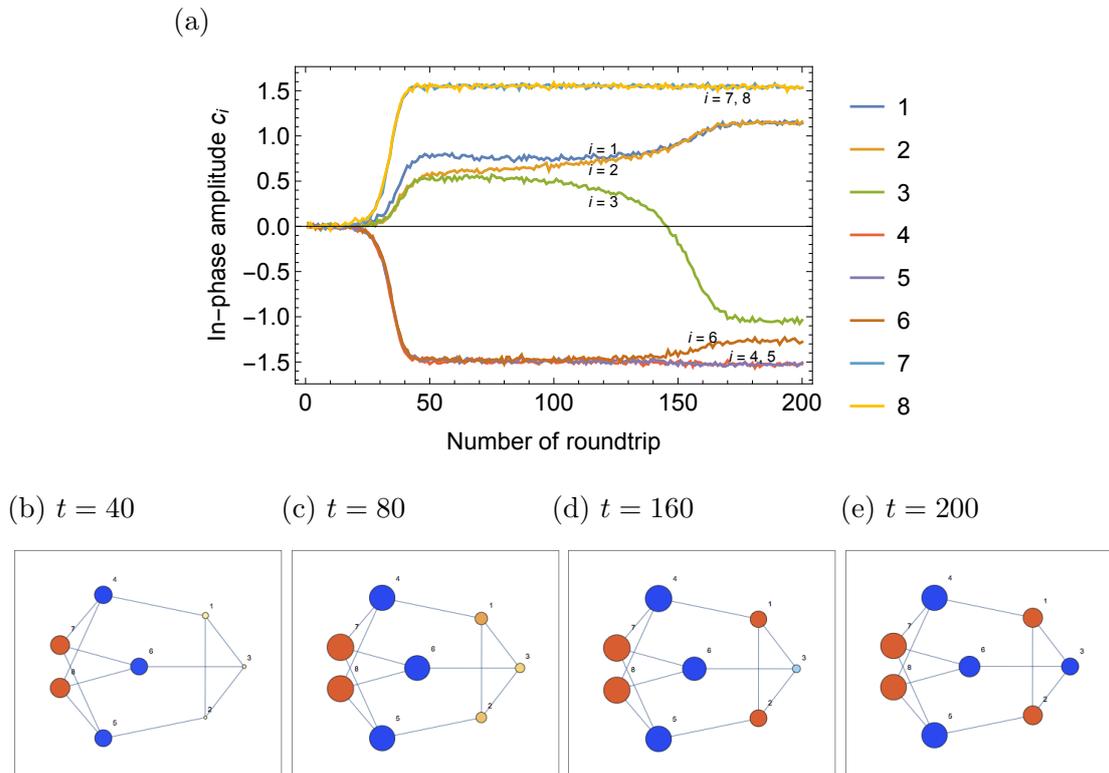


Figure 5.2. A cubic graph of order $n = 8$. (a) Time evolution of the DOPO in-phase amplitudes as a function of the number of cavity round trips. Bottom 4 figures show the amplitude (vertex size) and the sign of $+/-$ (as color red/blue) for each DOPO. (b) Initially the left larger bipartite cluster (indices $4 \leq i \leq 8$) grows. (c) The smaller cluster begins to grow. At this point, the spin states are ambiguously but moderately shifting to positive pushed by the adjacent vertices in the left cluster. (d) The vertex $i = 1$ went to negative to reflect the consistency in the right cluster. (e) CIM converged to the ground state.

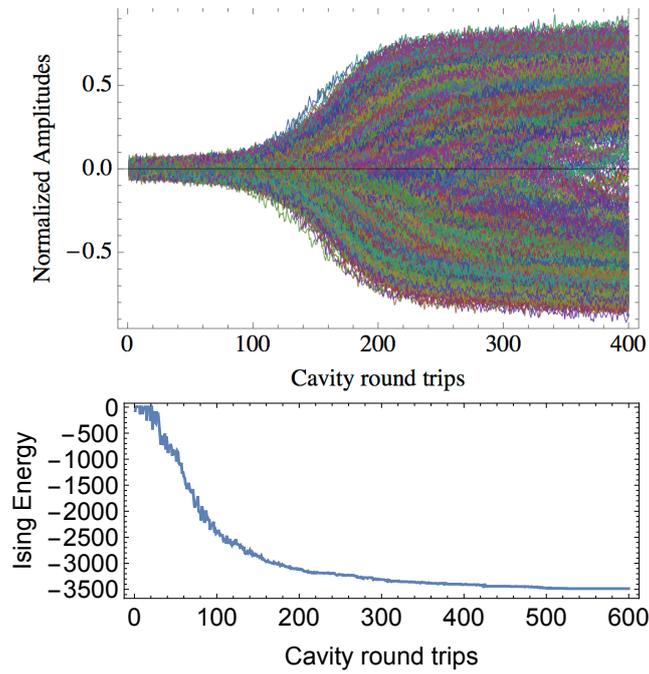


Figure 5.3. Simulation result for a sparse random graph g_{22} ($n = 2000, m = 19990$ and 1% density). The pump rate is gradually increased from $p = 0.9$ to 1.3 linearly, with the fixed coupling constant of $\xi = -0.02/\sqrt{\langle k \rangle} = 0.004$. The CUT results are 13081 (best), 12997.99 (mean) and 12909 (worst) in 100 trials. Note that $12993 \text{ (GW)} \leq 13351 \text{ (SA)} \leq 13359 \text{ (BLS)} \leq \text{OPT} \leq 14136 \text{ (SDP-UB)}$.

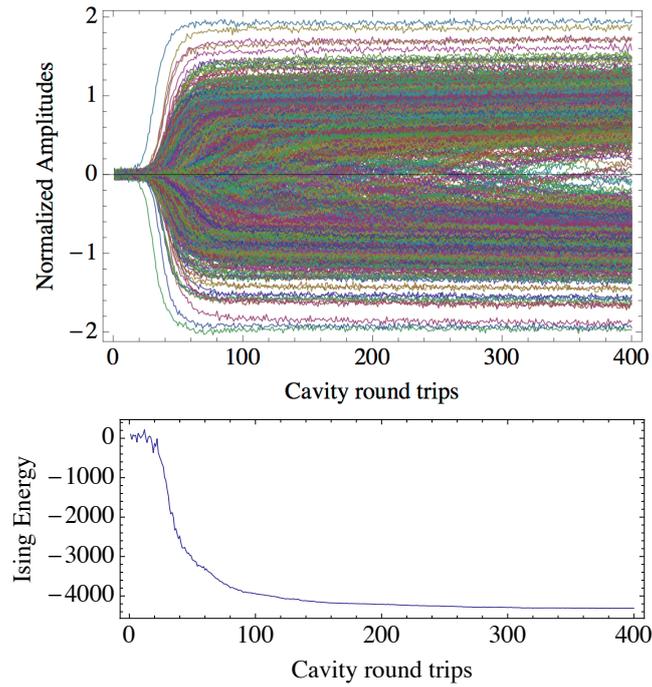


Figure 5.4. Simulation result for a ± 1 -weighted scale-free graph $g38$ ($n = 2000, m = 11779$ and 0.59% density). Several DOPOs have larger amplitudes than others since the corresponding vertices have large degree (hub). The gradual pumping with $p : 0.9 \mapsto 1.3$ and $\xi = -0.05/\sqrt{\langle k \rangle} = 0.0042$. The results are 2273 (best), 2228.69 (mean) in 100 trials, which is . Note that $2200 \text{ (GW)} \leq 2379 \text{ (SA)} \leq 2408 \text{ (BLS)} \leq \text{OPT} \leq 2877 \text{ (SDP-UB)}$.

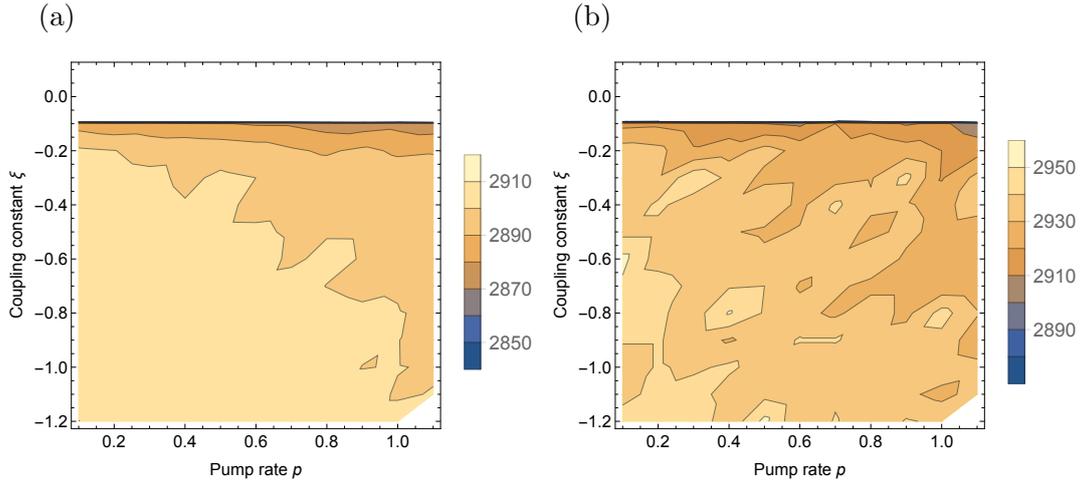


Figure 5.5. Parameter sweep in (p, ξ) space with G-set instance g14 (scale-free, $n = 800, m = 4694$). Brighter color is better with larger cut values. (a) Best case in 100 trials. (b) Mean in 100 trials.

5.3 Performances on Benchmark Instances

5.3.1 Parameter Sweep

When we solve larger problems, the choice of parameters (p, ξ) makes a significant effect in output values. While it is basically problem dependent, here we show a case study on a benchmark instance. Figures 5.6 and 5.5 show the lower pump rate is better, while the coupling constant should be large in sparse instances and small in dense complete graph instances.

5.3.2 Performance on G-set

The performance of a CIM with DOPO network was tested against the NP-hard MAX-CUT problems on sparse graphs (G-set) [148]. These test instances were randomly constructed using a machine-independent random graph generator written by G. Rinaldi, with the number of vertices ranging from 800 to 20000, edge density from 0.02% to 6%, and topology from random, scale-free, to toroidal.

The output cut values of running the CIM, SA, and GW for some of the G-set graphs are summarized in Table 5.2. The results for CIM are obtained in 50 ms, which correspond to the performance of an experimental system after 5000 DOPO cavity round trips. The best result and ensemble average value for 100 trials are shown. Here, the parameters are set to be $p = 1.6$, $\xi = -0.06$, and the coupling constant $\xi_{ij} = \xi w_{ij} / \sqrt{\langle k \rangle}$ is normalized by the square root of the average degree $\langle k \rangle$. The hysteretic optimization method, in which the swinging and decaying Zeeman term that flips the signal amplitude (spin) back and forth, is implemented four times after 10 ms initial free evolution [149] hence the larger pump rate is used for the system stability (see Reference [150] for further analysis). Each

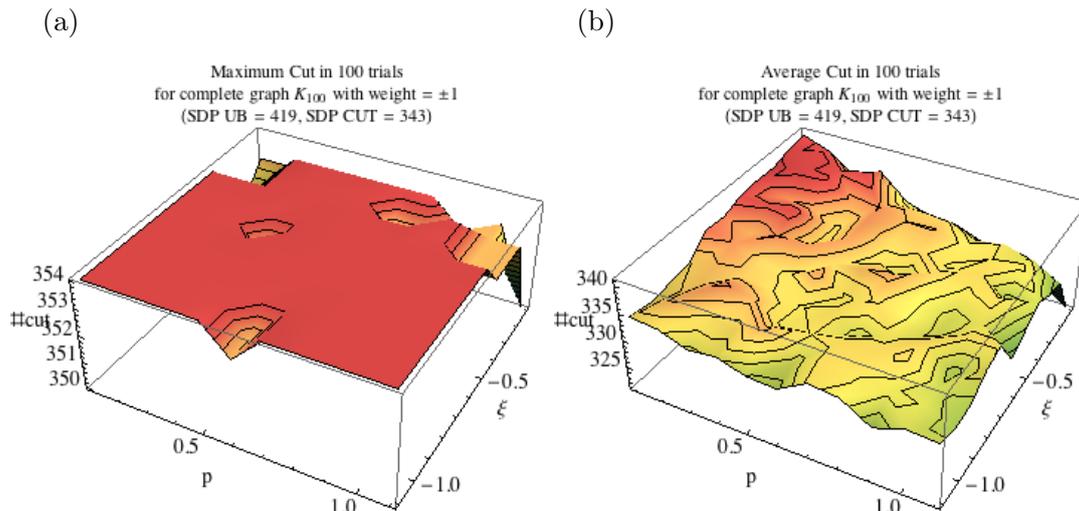


Figure 5.6. Parameter sweep in (p, ξ) space with dense instance K_{100} . (a) Best case in 100 trials. (b) Mean in 100 trials.

hysteretic optimization takes 10 ms so that the total search takes 50 ms. The result of SA is also obtained in 50 ms for each graph. For GW, the computation time ranged between 2.3 s and 1.1×10^5 s, depending on n . The best outputs of the CIM were $1.62 \pm 0.58\%$ better than GW but $0.38 \pm 0.40\%$ worse than SA, and CIM found better cut against GW except for a toroidal graph (g50) and a disconnected random graph (g70). Table 5.3 summarizes the above results and performance of other solvers with respect to each graph topology.

Scaling on G-set

Computation time for the random graphs in G-set instances is also studied. Here the subset of graphs in which MAX-CUT problems can be solved in polynomial time (i.e., planar graphs [99, 100], weakly bipartite graphs [98], positive weighted graphs without a long odd cycle [102], and graphs with integer edge weight bounded by n and fixed genus [103]) are excluded (see Section 2.3.4). The execution time of CIM is evaluated under the machine spec described in Section 5.1 with $p = 0.2$, $\xi = -0.06$, and $\xi_{ij} = \xi w_{ij} / \sqrt{\langle k \rangle}$. Again, the computation time of SA and CIM is the actual time (without graph file I/O) to obtain the same accuracy of solution as GW. Figure 5.7 shows the computation time as functions of the problem size n . The computational cost of interior point method dominates the GW algorithm. (Note that G-set contains graphs with both positive and negative edge weights so that we must use the slowest interior point method.) Then the computation time is almost constant for both SA and CIM. The computation time of SA with constant Monte Carlo sweep is expected to scale $O(n \langle k \rangle) \sim O(1)$ (here for the random graphs in G-set, $\langle k \rangle \sim O(n^{-1.09})$). The computation time of CIM here is governed by a turn-on delay time of the DOPO network to reach a steady state oscillation condition, which is constant for varying values of n as mentioned above [151].

Table 5.2. Performance of the coherent Ising machine, simulated annealing and Goemans-Williamson SDP algorithm in solving the MAX-CUT problems on sparse graphs (G-set). U_{SDP} is the optimal solution to the semidefinite relaxation of the MAX-CUT problem, and C_{GW} is the best solution obtained by n projections after SDP. C_{best} is the best known solutions already published. C_{SA} and $\langle C_{\text{SA}} \rangle$ are the best and average values obtained by SA in 100 trials of 50 ms. C_{CIM} and $\langle C_{\text{CIM}} \rangle$ are the best and average values in CIM in 100 runs of 50 ms (= 5000 DOPO cavity round trips), respectively. To make comparisons with each other, every cut value C generated from the CIM, SA, or GW algorithm is normalized according to $(C + E_{\text{neg}})/(U_{\text{SDP}} + E_{\text{neg}})$, where $E_{\text{neg}} \geq 0$ is the number of negative edges. In the bottom of this table, the average values of all 71 G-set graphs are shown.

Instance	$ V $	$ E $	U_{SDP}	C_{best}	C_{GW}	C_{SA}	$\langle C_{\text{SA}} \rangle$	C_{CIM}	$\langle C_{\text{CIM}} \rangle$
g1	800	19176	12083	0.9620	0.9457	0.9620	0.9597	0.9614	0.9570
g2	800	19176	12089	0.9612	0.9437	0.9605	0.9588	0.9602	0.9562
g3	800	19176	12084	0.9618	0.9451	0.9618	0.9598	0.9609	0.9570
g4	800	19176	12111	0.9616	0.9451	0.9614	0.9600	0.9614	0.9568
g5	800	19176	12099	0.9613	0.9458	0.9612	0.9596	0.9601	0.9570
g6	800	19176	2656	0.9607	0.9448	0.9606	0.9592	0.9601	0.9559
g7	800	19176	2489	0.9603	0.9446	0.9599	0.9582	0.9587	0.9556
g8	800	19176	2506	0.9589	0.9443	0.9587	0.9569	0.9579	0.9547
g9	800	19176	2528	0.9610	0.9452	0.9607	0.9589	0.9598	0.9564
g10	800	19176	2485	0.9601	0.9444	0.9599	0.9580	0.9591	0.9554
g11	800	1600	629	0.9540	0.9327	0.9526	0.9478	0.9455	0.9370
g12	800	1600	623	0.9530	0.9333	0.9502	0.9465	0.9460	0.9370
g13	800	1600	647	0.9546	0.9336	0.9518	0.9464	0.9462	0.9393
g14	800	4694	3191	0.9602	0.9336	0.9580	0.9544	0.9514	0.9472
g15	800	4661	3171	0.9618	0.9398	0.9609	0.9549	0.9540	0.9481
g16	800	4672	3175	0.9613	0.9364	0.9587	0.9548	0.9528	0.9477
g17	800	4667	3171	0.9609	0.9376	0.9584	0.9546	0.9530	0.9468
g18	800	4694	1166	0.9500	0.9282	0.9492	0.9439	0.9434	0.9372
g19	800	4661	1082	0.9493	0.9279	0.9478	0.9430	0.9424	0.9362
g20	800	4672	1111	0.9510	0.9354	0.9510	0.9444	0.9461	0.9379
g21	800	4667	1104	0.9502	0.9286	0.9502	0.9436	0.9473	0.9362
g22	2000	19990	14136	0.9450	0.9191	0.9445	0.9409	0.9405	0.9361
g23	2000	19990	14145	0.9441	0.9188	0.9426	0.9402	0.9392	0.9358
g24	2000	19990	14140	0.9432	0.9186	0.9422	0.9399	0.9398	0.9357
g25	2000	19990	14144	0.9432	0.9183	0.9422	0.9397	0.9401	0.9357
g26	2000	19990	14132	0.9431	0.9174	0.9422	0.9399	0.9391	0.9356
g27	2000	19990	4141	0.9435	0.9174	0.9422	0.9400	0.9390	0.9356
g28	2000	19990	4100	0.9433	0.9182	0.9425	0.9401	0.9399	0.9356
g29	2000	19990	4208	0.9433	0.9167	0.9415	0.9394	0.9385	0.9353
g30	2000	19990	4215	0.9433	0.9189	0.9425	0.9402	0.9396	0.9354
g31	2000	19990	4116	0.9430	0.9181	0.9424	0.9398	0.9392	0.9355
g32	2000	4000	1567	0.9559	0.9272	0.9508	0.9478	0.9424	0.9384
g33	2000	4000	1544	0.9545	0.9275	0.9500	0.9469	0.9438	0.9384
g34	2000	4000	1546	0.9546	0.9277	0.9507	0.9474	0.9440	0.9386

Instance	$ V $	$ E $	U_{SDP}	C_{best}	C_{GW}	C_{SA}	$\langle C_{\text{SA}} \rangle$	C_{CIM}	$\langle C_{\text{CIM}} \rangle$
g35	2000	11778	8014	0.9588	0.9292	0.9551	0.9523	0.9471	0.9438
g36	2000	11766	8005	0.9592	0.9282	0.9557	0.9525	0.9475	0.9441
g37	2000	11785	8018	0.9590	0.9310	0.9555	0.9522	0.9482	0.9434
g38	2000	11779	8014	0.9592	0.9291	0.9550	0.9524	0.9471	0.9440
g39	2000	11778	2877	0.9464	0.9226	0.9431	0.9399	0.9364	0.9318
g40	2000	11766	2864	0.9473	0.9225	0.9442	0.9399	0.9380	0.9317
g41	2000	11785	2867	0.9473	0.9230	0.9440	0.9397	0.9372	0.9311
g42	2000	11779	2946	0.9470	0.9230	0.9448	0.9402	0.9365	0.9314
g43	1000	9990	7032	0.9471	0.9292	0.9471	0.9439	0.9458	0.9396
g44	1000	9990	7027	0.9464	0.9251	0.9462	0.9438	0.9444	0.9393
g45	1000	9990	7024	0.9473	0.9245	0.9468	0.9446	0.9448	0.9396
g46	1000	9990	7029	0.9459	0.9223	0.9455	0.9429	0.9424	0.9390
g47	1000	9990	7036	0.9461	0.9261	0.9460	0.9432	0.9423	0.9386
g48	3000	6000	6000	1.0000	1.0000	1.0000	0.9919	1.0000	0.9747
g49	3000	6000	6000	1.0000	1.0000	1.0000	0.9861	1.0000	0.9791
g50	3000	6000	5988	0.9820	0.9820	0.9760	0.9707	0.9770	0.9686
g51	1000	5909	4006	0.9606	0.9333	0.9583	0.9544	0.9506	0.9468
g52	1000	5916	4009	0.9606	0.9327	0.9576	0.9546	0.9519	0.9471
g53	1000	5914	4009	0.9603	0.9346	0.9574	0.9545	0.9524	0.9469
g54	1000	5916	4006	0.9616	0.9381	0.9578	0.9547	0.9511	0.9466
g55	5000	12498	11039	0.9325	0.9006	0.9264	0.9215	0.9193	0.9160
g56	5000	12498	4760	0.9322	0.9008	0.9245	0.9213	0.9204	0.9155
g57	5000	10000	3885	0.9561	0.9237	0.9496	0.9473	0.9419	0.9384
g58	5000	29570	20136	0.9566	0.9239	0.9475	0.9448	0.9434	0.9411
g59	5000	29570	7312	0.9440	0.9148	0.9376	0.9356	0.9308	0.9288
g60	7000	17148	15222	0.9313	0.8989	0.9231	0.9201	0.9191	0.9152
g61	7000	17148	6828	0.9317	0.8991	0.9227	0.9201	0.9185	0.9149
g62	7000	14000	5431	0.9552	0.9228	0.9481	0.9463	0.9412	0.9381
g63	7000	41459	28244	0.9559	0.9230	0.9413	0.9381	0.9428	0.9407
g64	7000	41459	10466	0.9440	0.9143	0.9347	0.9324	0.9320	0.9299
g65	8000	16000	6206	0.9548	0.9217	0.9472	0.9456	0.9401	0.9374
g66	9000	18000	7077	0.9555	0.9220	0.9472	0.9455	0.9401	0.9374
g67	10000	20000	7744	0.9554	0.9215	0.9480	0.9459	0.9411	0.9388
g70	10000	9999	9863	0.9674	0.9633	0.9523	0.9479	0.9515	0.9482
g72	10000	20000	7809	0.9550	0.9215	0.9466	0.9448	0.9395	0.9376
g77	14000	28000	11046	0.9557	0.9205	0.9449	0.9419	0.9407	0.9382
g81	20000	40000	15656	0.9551	0.9195	0.9187	0.9125	0.9393	0.9376
Mean				0.9540	0.9303	0.9502	0.9469	0.9464	0.9415
Worst				0.9313	0.8989	0.9187	0.9125	0.9185	0.9149

Table 5.3. The performance is summarized according to the topology. SDP is the GW algorithm solving SDP until primal-dual relative gap of 10^{-3} , followed by n rounding. MIQP is obtained by the IBM CPLEX MIQP solver, which uses the branch-and-bound method with QP relaxation. CIM and SA (with hysteretic optimization) show the best cut in 100 trials (same as Table 5.2). BLS is from Reference [115].

	SDP	MIQP	CIM	SA	BLS
Time (s)	$2 - 10^5$	$12 - 10^5$	0.05	0.05	$1 - 10^4$
Random	0.9273	0.9116	0.9448	0.9471	0.9491
Scale-free	0.9288	0.9269	0.9451	0.9510	0.9547
Toroidal	0.9375	0.9618	0.9511	0.9548	0.9616
All	0.9303	0.9349	0.9464	0.9502	0.9540

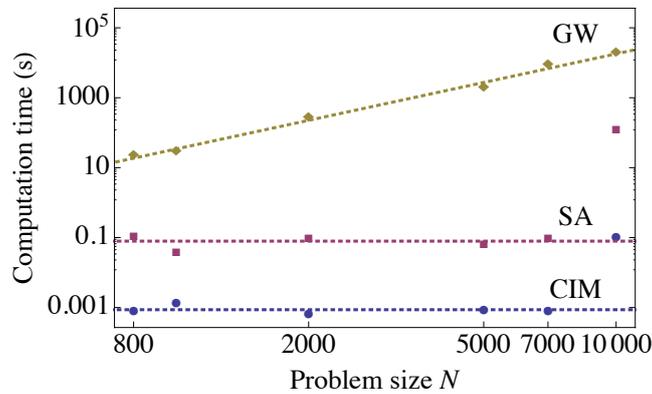


Figure 5.7. The computation time scaling on the G-set instances. The codes are not optimized so the performance is focused on the relative scaling. It does not indicate the relative speedup on CIM compared from SA as far as the instance is sparse. Note that the edge density is decreasing as the n increases.

5.3.3 Computation Time Scaling on Complete Graph

Computational experiments were conducted against fully connected complete graphs with the number of vertices ranging from 40 to 20000 and the edges are randomly weighted ± 1 . Since there is no guarantee that the CIM or SA can efficiently find the ground states (optimal solution) for MAX-CUT problems, the GW solution was used as the mark of sufficient accuracy (note that the actual ground states are generally not known for these problem sizes). The CIM and SA then competed for the computation time to reach that accuracy. The definition of computation time of CIM and SA is the same as the previous subsection. The time and temperature scheduling parameters of the SA were set as follows: Inverse temperature increased with the logarithmic function. The number of spin flipping was optimized to be 10^l times for some $l \in \mathbb{N}$, which requires the minimum computation time to achieve the same accuracy as with the GW.

Figure 5.8 (a) shows the computation time versus problem size (number of vertices). The computation time is defined as the actual time to solve a given MAX-CUT problem in complete graph for GW; as the CPU time to reach the same accuracy achieved by GW for SA, SG3, and BLS; and as the time estimated by the (number of round trips) \times (cavity round trip time) to obtain the same accuracy as GW for CIM. The preparation time needed to input J_{ij} into the computing system, i.e., the graph I/O time, is not included. The three types of time complexity are shown. The time complexity $O(n^{3.5})$ for the GW is dominated by the interior-point method in the general Goemans-Williamson algorithm. The SA scales as $O(m) = O(n^2)$, which indicates that it requires the number of spin flips to be proportional to n to achieve optimal performance. Each spin flip costs a computation time proportional to the degree k_i , where k_i is equal to $n - 1$ for all $i \in V$ in the case of complete graphs. Thus, the computation time scales as $O(n\langle k \rangle) = O(n^2)$ for the SA in the complete graphs¹. SG3 scales as $O(m) = O(n^2)$ in Figure 5.8, but the values for $n = 40, 160, 800$ are not shown because it did not reach the accuracy reached by the GW solution. BLS exhibits competitive performance against SA and also scales as $O(n^2)$. The CIM exhibits a problem-size independent computation time of less than 10^{-3} s in the case of $n \leq 20000$. The computation time of CIM is determined by the turn-on delay time of the DOPO network oscillation, which in turn depends on the round trip time and the pump rate [151].

5.4 Summary

In this chapter, the computational performance of CIM was estimated by numerical simulations. The relative performance is studied in both sparse (Section 5.3.2) and dense (Section 5.3.3) graph instances. Results imply that the application for dense graph is necessary to see a visible performance gain.

¹Note that there exists order- n Monte Carlo which scales linear to n on arbitrary topology. But practically, it requires a concealed factor and no speedup can be observed, which is discussed in the Appendix D.

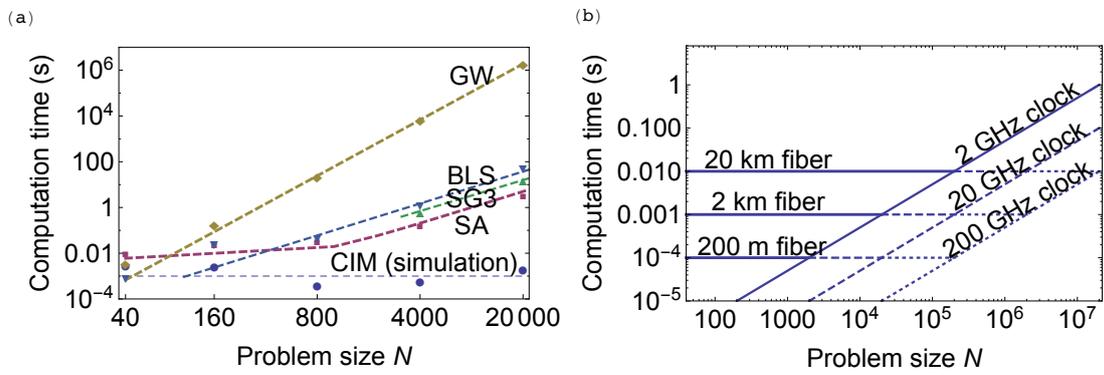


Figure 5.8. (a) Computation time scaling of GW, SA, BLS, SG3, and CIM, where the fiber cavity length is fixed to be 2 km. A sufficient FPGA resource is assumed for the feedback circuit. (b) Computation time scaling of CIM with variable cavity length (circulation frequency) and pulse repetition frequency (pulse interval).

Chapter 6

Experimental Results and Validation

In this chapter, results from a real experimental implementation is presented to validate the mathematical model and simulation results

- [76] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.

6.1 Setup: Machine with $n = 100$

In this section, the results from experimental MF-CIM in Stanford University are presented. The comparison with one in NTT, which will be appeared in the next Chapter 7 is summarized in Table 6.1. Basically the system is used with below threshold constant pump rate between $0.7 \leq p \leq 0.9$, which will go above threshold effectively with mutual injection. Here we show more detailed simulation assuming the specific experimental parameters: pump rate of $p = 0.88$ and the coupling coefficient of $\xi(t) = -0.001t$ to predict the performance correctly (we chose the best coefficient for the coupling, with optimizing the slope).

Table 6.1. Experimental configurations of two MF-CIMs.

	Stanford [76]	NTT [77]
Number of pulses n	100	2048
Round trip time	1.6 μ s	5 μ s
Pulse repetition frequency	100 MHz	1 GHz
Model of FPGA	Xilinx Virtex-6	2 \times Xilinx Virtex-7
ADC/DAC resolution	12/12 bit	12/14 bit
Operational scheme	Gradual coupling	Gradual pumping

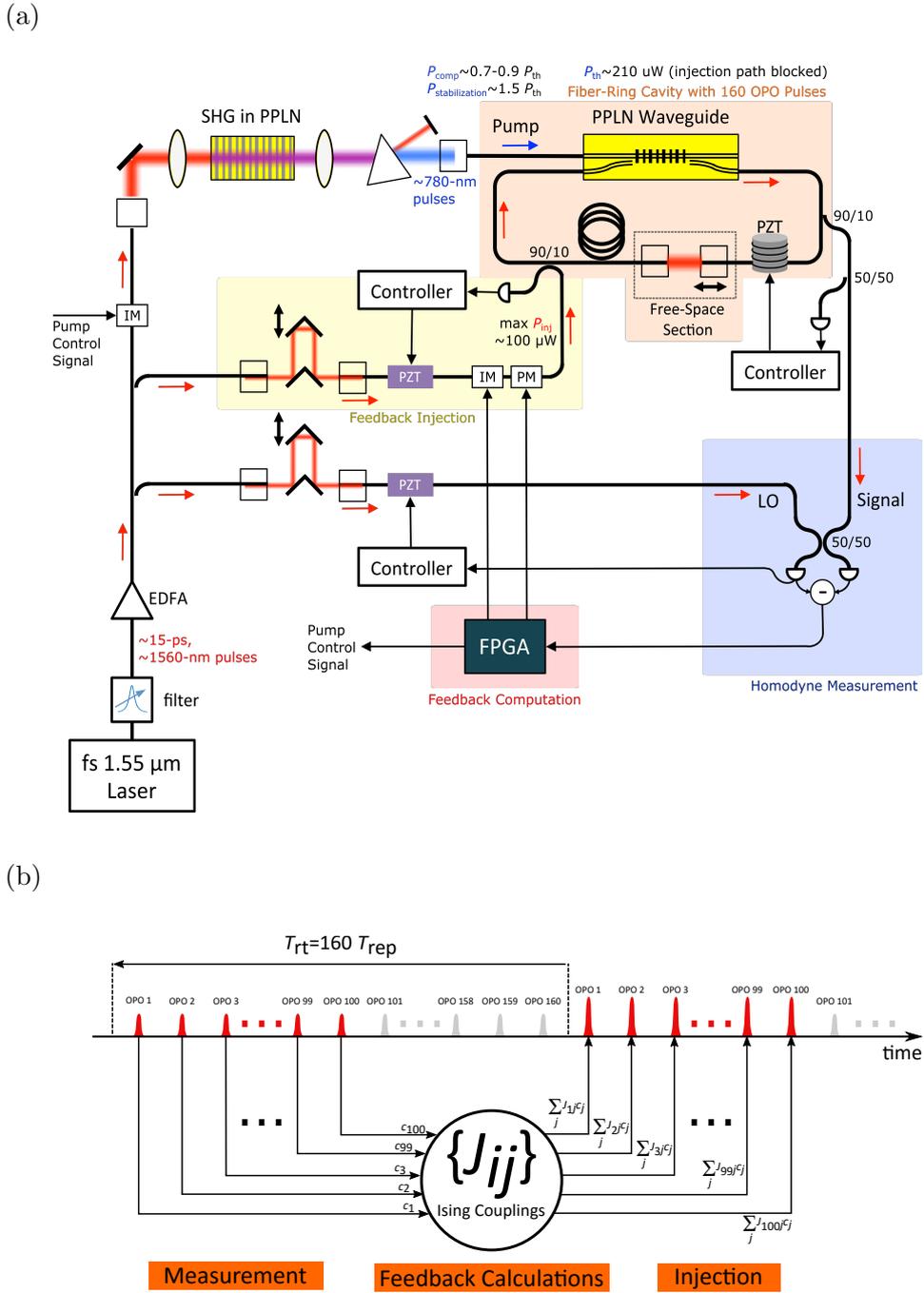


Figure 6.1. Experimental setup of machine 1. (a) The main fiber ring cavity (in the right upper part of Figure) contains $n = 160$ DOPO pulses, where 100 of them will be used as the signal DOPO. (b) The FPGA calculates the coupling term from the $n \times n$ matrix J and the measured in-phase signal $\tilde{c} \in \mathbb{R}^{100}$. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

The experimental system accommodates 160 pulses in a 330 m length ring cavity, 100 of which is used as the computation (signal pulses) and the rest is used to match the measurement-feedback circuit waiting time. Since the pulse repetition frequency is 100 MHz (interval of 10 ns), the cavity round trip time is $160 \times 10 \text{ ns} = 1.6 \mu\text{s}$. The coupling term is calculated on a Xilinx Virtex-6 SX475T FPGA. The FPGA clock is set to 200 MHz with an FMC 101 board containing 12-bit ADC and 12-bit DAC both clocked at 800 MHz.

6.2 Results

In this section, experimental results are shown. Here, the optimal solutions shown here are obtained BiqMac solver [152].

6.2.1 Single Run

Möbius ladder graph M_{16}

First, we checked with the simple cubic graph instance of $n = 16$ which has the frustration, namely Möbius ladder graph M_{16} defined in Def. 2.7. The results of both simulation and experiment are shown in Figure 6.2. The middle figures (b,c) show the DOPO signal amplitudes measured by the photodetector. It captured that the time scale of the simulation and experiment matches well, where both results start to bifurcate around 60-70 round trips. Figures 6.2 (d,e) shows the system could find the optimal value (CUT = 22) in less than 100 round trips in both cases. Note that the M_n has the frustration when $n \in 4\mathbb{Z}$, because of the cycle and diagonal connections conflict, resulting the ground states have two 2 spins clusters which are depicted in Figure 6.2 (a).

Erdős-Rényi random graph of order 100

Next, we checked the performance on a larger instance with the random connection, namely Erdős-Rényi random graph of order $n = 100$ and the connecting probability $p = 0.1$ ($m = 495$) plotted in Figure 6.3 (a). Again, the middle Figures (b,c) show the measured signals from simulation and experiment, which show the simulation-experiment correspondence. The bottom Figures (b,c) show the system could find the optimal CUT value in both cases. Note that the signal amplitudes are not uniform due to injection, which is remarkable in such a large graph.

6.2.2 Energy Distribution

The energy distribution in statistical trials are investigated. Figure 6.4 shows the results for 3 cubic graphs of order $n = 16$. Among three instances, the Möbius ladder (Figure 6.4 (a)) was the highest success, where all trials were successful in 100 runs. Even in the trials when CIM could not find the ground state, right two figures show that it could find near-optimal states, namely first or second excited states in these instances.

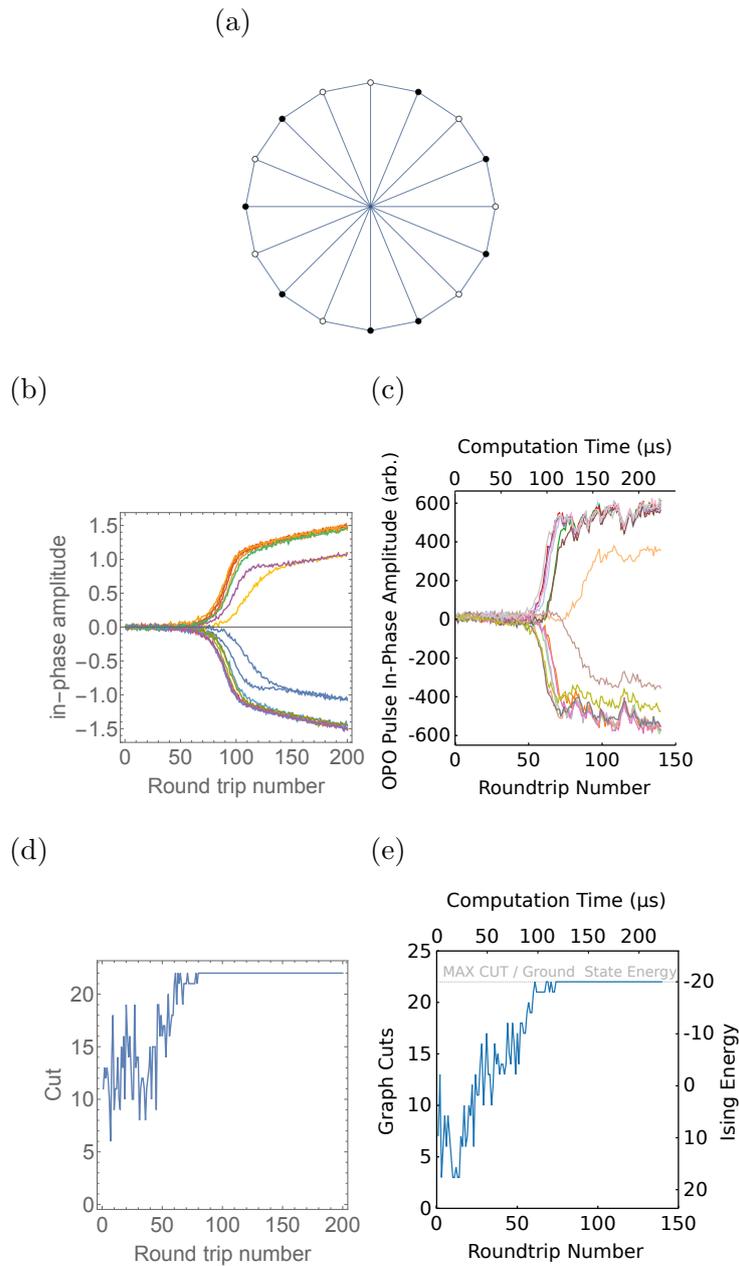


Figure 6.2. The simulation and experimental results when solving the Möbius ladder graph M_{16} . (a) The ground state configuration of the M_{16} . (b,c) The DOPO amplitude as a function of the number of cavity round trip time obtained from the simulation and experiment, respectively. (d,e) The corresponding cut values as a function of the number of cavity round trip time obtained from the simulation and experiment, respectively. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

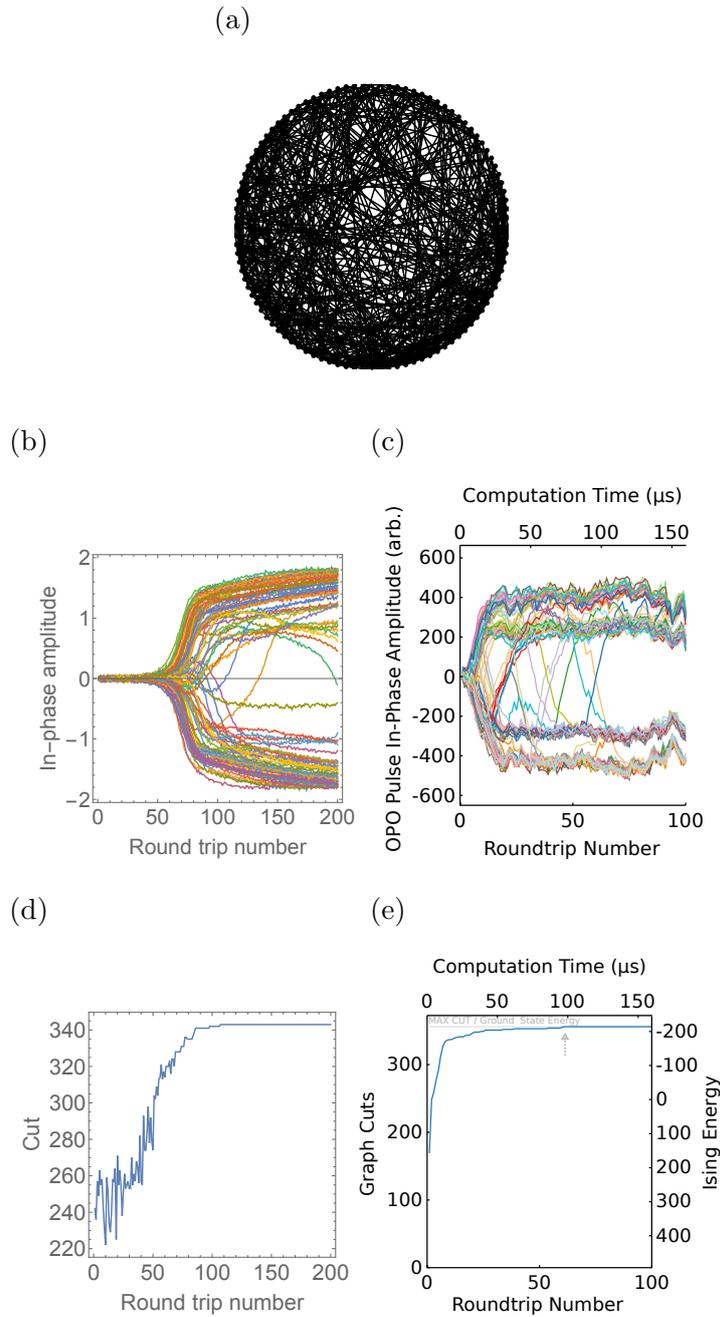


Figure 6.3. The simulation and experimental results when solving the Erdős-Rényi random graph of order 100. (a) Graph plot. (b,c) The DOPO amplitude as a function of the number of cavity round trip time obtained from the simulation and experiment, respectively. (d,e) The corresponding cut values as a function of the number of cavity round trip time obtained from the simulation and experiment, respectively. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

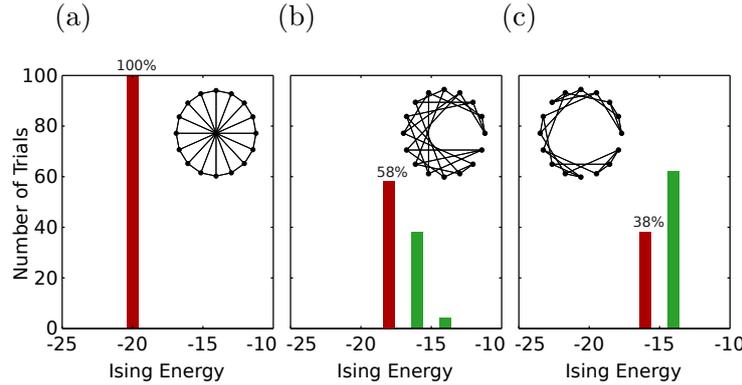


Figure 6.4. Energy distribution when solving the cubic graphs of order $n = 16$. Red bars indicate the ground states with success probability shown above. (a) CIM found the optimal CUT on the M_{16} 100 times in 100 trials. (b,c) The success probability of finding the ground state is not 1 but CIM found low-energy states in rest of the trials. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

Remember the single DOPO steady-state distribution becomes Boltzmann distribution under several assumptions in Section 3.6.4. There are several studies which state the steady state of CIM become a Boltzmann distribution in simple cases [18, 153].

6.2.3 Statistics with Cubic Graph

The results from a several instances is insufficient. Next we investigated the performance of MF-CIM as the success rate statistics for many trials of many problem instances.

Möbius ladder

We performed simulation and experiments with the Möbius ladder graph of $4 \leq n \leq 100$ (see Figure 6.5). As explained above and in Figure 6.2 (e), the graphs have two types of ground state configurations, depending on its number of vertices is multiples of 4 ($n \in 4\mathbb{Z}$) or not ($n \in 4\mathbb{Z}+2$). Somehow the bipartite graphs without frustration ($n \in 4\mathbb{Z}+2$) is more difficult for CIM. The consistency can be checked to see the 100% success for the small-sized instances and the 80% success around $n = 32$ especially for the $n \in 4\mathbb{Z}$ instances.

All Cubic Graphs of $n = 16$

To see the performance in a batch of instances, we created all the cubic graphs of the order $n = 16$. In conformity with the experiment, DOPO amplitudes in the simulation are initialized by biased Zeeman term in 30 round trips and then gradual coupling with

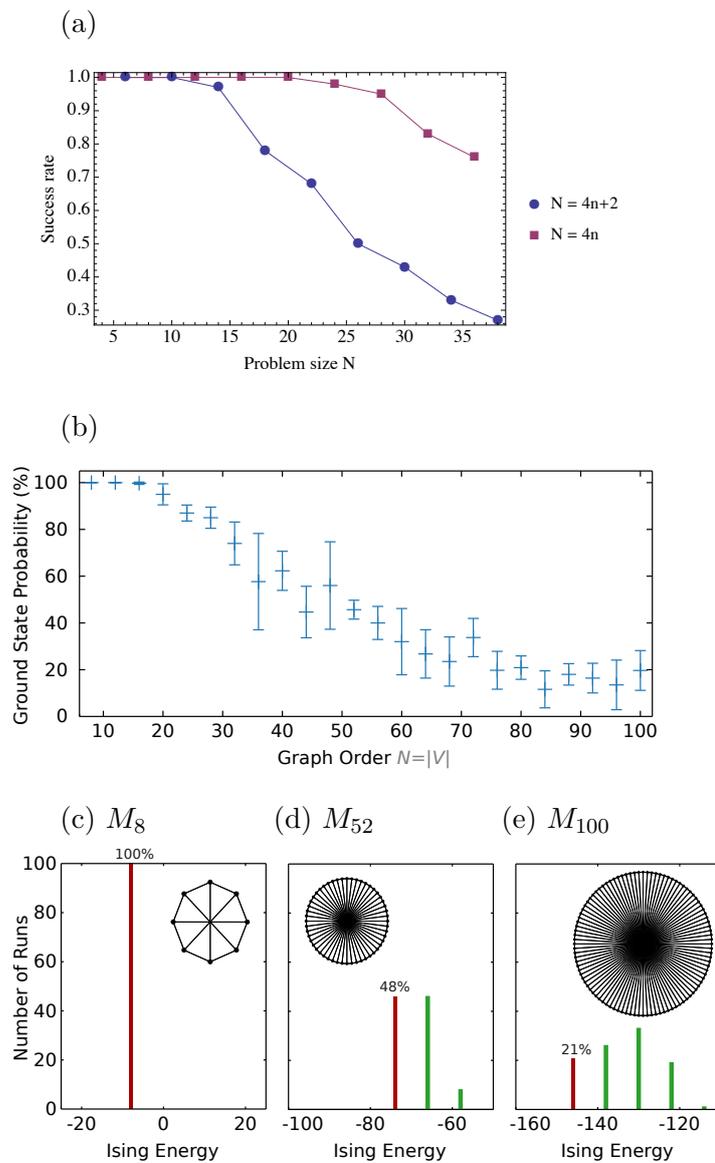


Figure 6.5. Success probability scaling to find the optimal CUT in Möbius ladder graphs M_n of $8 \leq n \leq 100$. (a) simulation and (b) experimental results for $n \in 4\mathbb{Z}$ case. Error bar is standard deviation of multiple sessions of 100 trials. (c-e) Experimental energy distributions in 100 trials are shown for each graph. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

$d\xi/dt = -0.001$ is operated in 300 round trips under the fixed pump rate of $p = 0.88$. Results are summarized in Figure 6.6. As seen in Figure 6.6 (a), so far the success rate is better in the simulation than the experiment, especially for the easy (success rate is around 1) instances. The correlation between the simulation and experiment on each instance is described in Figure 6.6 (b), which should be compared with the self-correlation between the different sessions of trials in Figure 6.6 (c). The correlation coefficient is $\rho = 0.80$. The simulation model in the Section 4.3 well describes the experimental situation.

The reason of the worse performance can be inferred to be due to the experimental instability, which is partly shown in Figure 6.7. The moderately slow wave in Figure means the possibility of the system performance depending on the time. Note that the multimode simulation discussed in the Appendix B can improve the success rate but it can only be applied to the free-space system not in the current fiber-based implementation.

6.2.4 Finite Size Scaling with Random Graph

To Graph order n

In Figure 6.8, the two aspects of scaling are shown. Here, 10 random cubic graphs are generated for each n . Figures 6.8 (a) and (b) show the success probability and computation time to find α times the ground state, where α is swept from 0.9 to 1.0. Computation time is defined as $T_{\text{comp}} \times \lceil \log(1 - 0.99) / \log(1 - p_{\text{success}}) \rceil$, which is the $\underset{k}{\text{argmin}}[(1 - p_{\text{success}})^k < (1 - 0.99)]$.

For the ground states, the success probability decays and computation time scales exponentially as n increase. It is MAX-CUT is NP-hard and no exponential speed-up is observed. It improves as the solution accuracy becomes slack. Around $\alpha = 0.94$, it almost become flat in both Figures 6.8 (a) and (b).

To Edge density

Above results are obtained from sparse, mostly cubic, graphs. Since measurement-feedback CIM is designed to implement arbitrary graph topology thanks to FPGA, we checked the performance with variable density of graphs. Results are shown in Figure 6.9. The experimental success probability drops in the complete graphs. This is due to the insufficient pump power, due to the high sensitivity to the coupling constant. Current experimental setup is operated with gradual coupling under fixed ($p = 0.88$) pump rate. So the performance improvement with higher pump rate is studied in the simulation, which is shown in the bottom. It demonstrates the high pump rate will improve the performance and as far as we need CUT better than 0.99 of the ground states, the success probability can be 1.

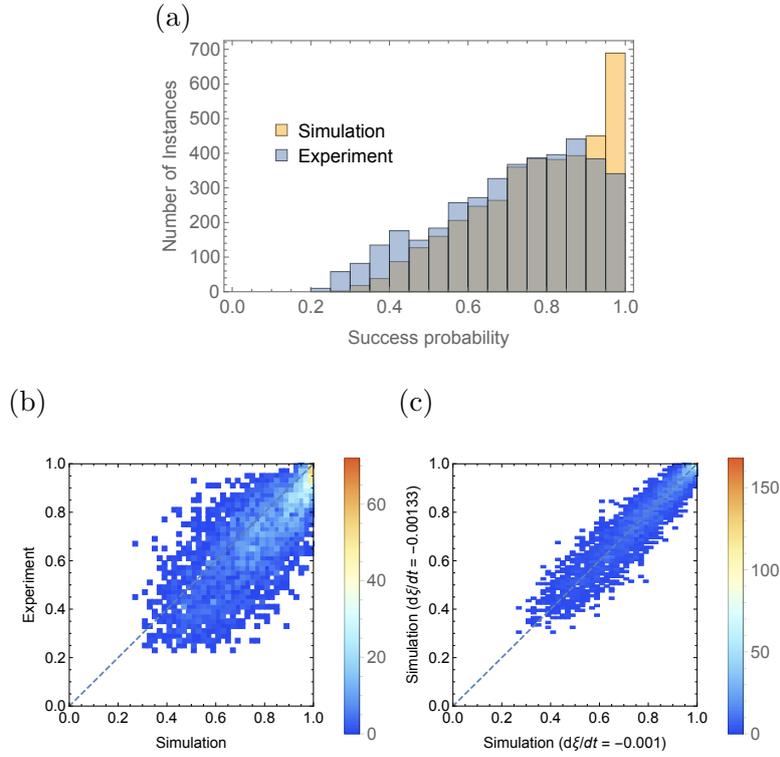


Figure 6.6. The comparison between the experiment and simulation on all the cubic graph instances of order $n = 16$. (a,b) Histogram and correlation histogram of the success probability, respectively. In conformity with the experiment, DOPO amplitudes in the simulation are initialized by biased Zeeman term and then gradual coupling with $d\xi/dt = -0.001$ is operated in 300 round trips under the fixed pump rate of $p = 0.88$. The histogram shows the simulation result is better around the successful region (nearly 100% success instances). The Pearson correlation coefficient for these data is $\rho = 0.80$ (c) Correlations between two independent sets of 100 trials with the different parameters $d\xi/dt \in \{-0.001, -0.00133\}$. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

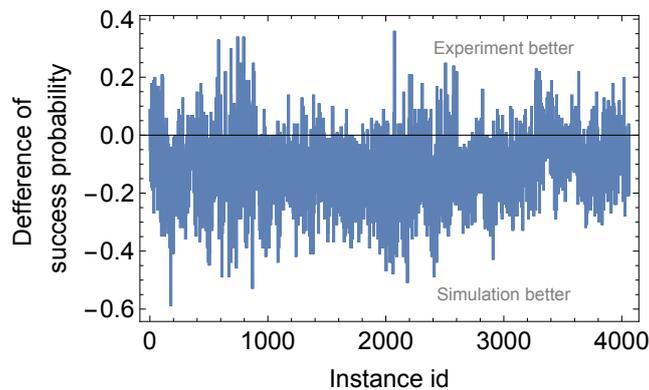


Figure 6.7. Difference in the success rate between simulation and experiments. The low success rate in the experimental system can be considered to originate in the experimental instability.

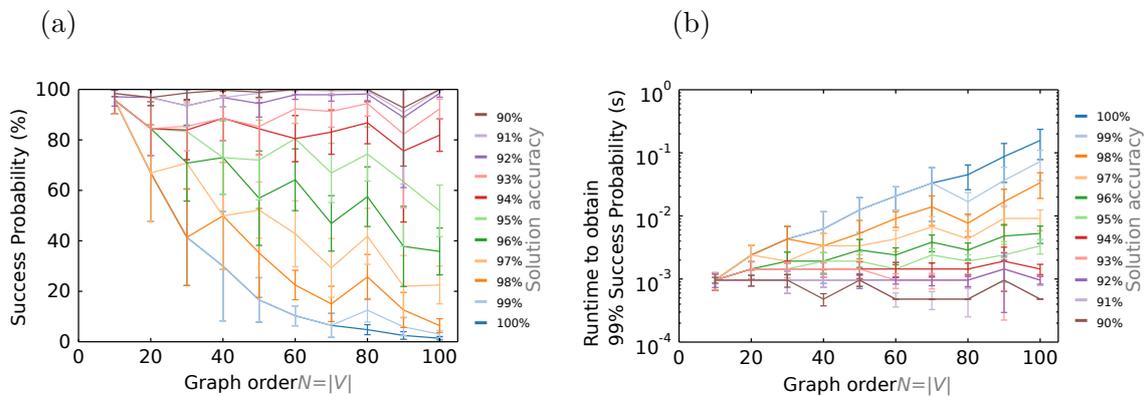
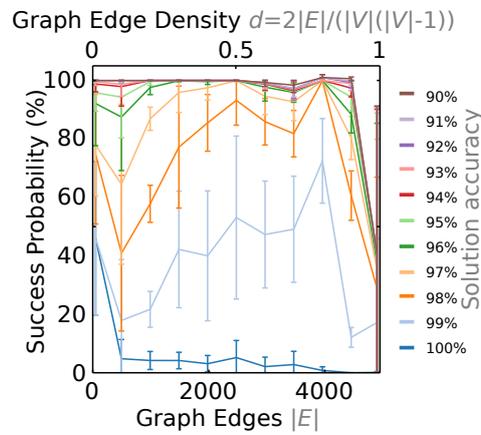


Figure 6.8. (a) Success probability and (b) computation time scaling evaluated on the random cubic graphs. Each line of color means the performance is measured for different residuals for the ground state, e.g., blue is the performance for ground state and red is to the 94% of the ground state. Error bars indicate the standard deviation, which is affected by the differences in problem difficulty. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

(a) Experiment ($p = 0.88$)

Simulations

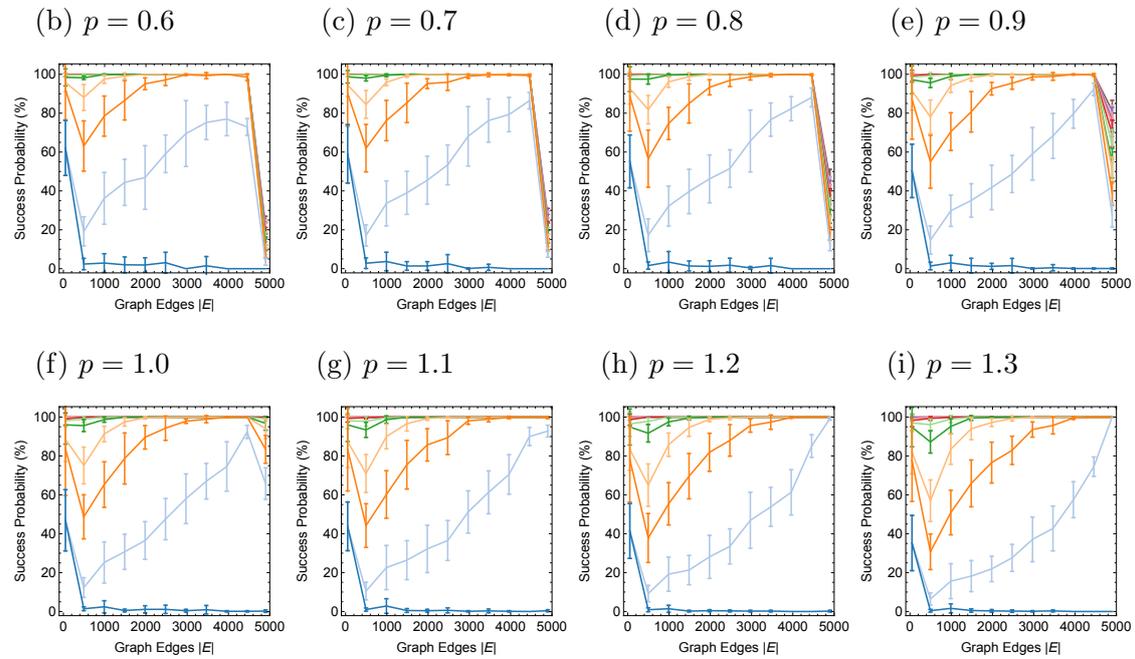


Figure 6.9. The scaling of the success probability to find the ground state for fixed order ($n = 100$) random graphs in (a) .experiment and (b-i) simulations. It is well corresponding with the experimental parameter ($p = 0.88$) and simulation (around $0.8 \leq p \leq 0.9$) especially for the success rate in the dense case. As the pump rate increases in the simulation, success rate around the dense region dramatically increase, while the sparse instances drop. From [P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.]. Reprinted with permission from AAAS.

6.3 Summary

In this chapter, we checked the correspondence between the theoretical model presented in Chapter 4 and the experimental data in relatively small-sized instances ($n \leq 100$). Results are consistent with three points of view.

- The order of the time constant and amplitude evolution are confirmed with several single trials (Section 6.2.1).
- Computational performance in terms of scaling property and CUT histogram (Section 6.2.3).
- Parameter dependence (Section 6.2.4).

In the next chapter, the experimental model will be compared with existing algorithms.

Chapter 7

Performance Evaluation with Larger Problems

Chapter 5 implied CIM performs relatively better on larger problems. In this chapter, the computational performance is evaluated on $n \geq 2000$ MAX-CUT benchmark instances against two methods: approximation algorithm of GW by SDP relaxation and metaheuristic of SA

- [77] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue. A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312):603–606, 2016.

The experimental results again evaluated with various implementations of neural network based heuristics

- [78] Y. Haribara, H. Ishikawa, S. Utsunomiya, K. Aihara, and Y. Yamamoto. Performance evaluation of coherent Ising machines against classical neural networks. *Quantum Science and Technology*, 2(4):044002, 2017.

7.1 Setup: Machine with $n = 2048$

The experimental configuration is slightly different from the previous Chapter 6 and one in this Chapter can handle larger problems [77]. The DOPO part consists of a 1 km optical fiber (round trip time of $5 \mu\text{s}$) with an externally pumped PPLN waveguide. The pulsed pump laser, at the 1 GHz repetition rate of 5000 times as the cavity circulation frequency, generates 5000 individual DOPO pulses in a single fiber ring cavity. A segment of them (2000 pulses) is used as the signal pulses for computation and the remaining portion (3000 pulses) is used for the cavity stabilization.

The feedback circuit stores the interaction strength for each pair of DOPO pulses. A portion of the optical pulse is picked-off by a beamsplitter and measured by balanced a balanced homodyne detector. The measured values of DOPO pulse amplitudes $(\tilde{c}_i)_{1 \leq i \leq n}$

are fed into an analog-digital converter (ADC), followed by two FPGAs of Xilinx Virtex-7 VX690T (693,120 logic cells, 3,600 DSP slices, a 52,920-Kb memory). Here, 1 GHz repetition rate of signal pulses is downclocked to 125 MHz (8 parallel) and the measured amplitudes \tilde{c}_i are sliced into the digital signals of 5 bits. Then 2 FPGAs sum up the coupling effect from the other vertices (in the given topology) $\sum_j J_{ij}\tilde{c}_j$ for the i th pulse. The feedback pulse train is modulated in intensity and phase by this output electrical signal after a digital-analog converter (DAC). The feedback pulse is injected to the signal DOPO pulse running through the main fiber ring cavity via the coupler 2.

The DOPO is operated near the oscillation threshold by crossing the pump rate from below to above the threshold in the case of gradual pumping [77]. In the beginning, the DOPO is biased at below the threshold in which all phase configuration is established so as a superposition state and the quantum parallel search is implemented [136]. Then, the external pump rate (or the feedback) strength is gradually increased, and once the whole system reaches the oscillation threshold, it selects a particular phase configuration which corresponds to the near-optimal solution of the original optimization problem.

7.2 Benchmark results against SA

7.2.1 CUT Histogram

We investigated the performance of CIM with relatively larger problems of order $n = 2000$. Instances are chosen from typical graph topologies: random, scale-free and complete. Former two graphs are in G-set and the last one is a complete graph, whose random weight is generated from uniform distribution of $\{-1, 1\}$. The instance property and the results are summarized in Table 7.1. Figure 7.1 shows the solution distribution. In both cases, the computation time is fixed to be 5 ms for CIM and 50 ms (in case of future speed-up) for SA.

7.3 Hardware Configurations for Comparison

Here classical neural network based optimization algorithms are chosen which are described in Section 2.4. They are implemented different hardware configurations, which will be used in the benchmark section. Note that all codes are implemented with C++¹.

7.3.1 CPU (for SA and HN)

SA and HN are iterative updating algorithms for discrete spins. We can achieve CPU implementation efficiently by SIMD bitwise operations in parallel². In this paper, we mainly used Intel Xeon E3-1225 v3 @ 3.2 GHz (Haswell architecture shipped in 2013). Note that the performance of SA is slightly improved from the previous paper, in which

¹We used Ubuntu 16.04.4 with GCC 5.4.0 (CPU) and CentOS 7.1.1503 with GCC 4.8.3 (PEZY-SC)

²The code is available here. <https://github.com/haribara/SA-complete-graph>

Table 7.1. Output CUT values. The computation time of CIM and SA are set to be 5 ms and 50 ms, respectively. Then the best (maximum) and mean CUT after 100 trials are listed. The corresponding histograms are shown in Figure 7.1.

	G22	G39	K_{2000}
topology	Random	Scale-free	Complete
n	2000	2000	2000
m	19990	11778	1999000
w	$\{0, 1\}$	$\{-1, 0, 1\}$	$\{-1, 1\}$
Best known	13359	2408	–
SA best	13336	2384	32781
SA mean	13298	2359	32314
CIM best	13313	2361	33191
CIM mean	13248	2328	32457
GW-SDP	12992	2200	29619

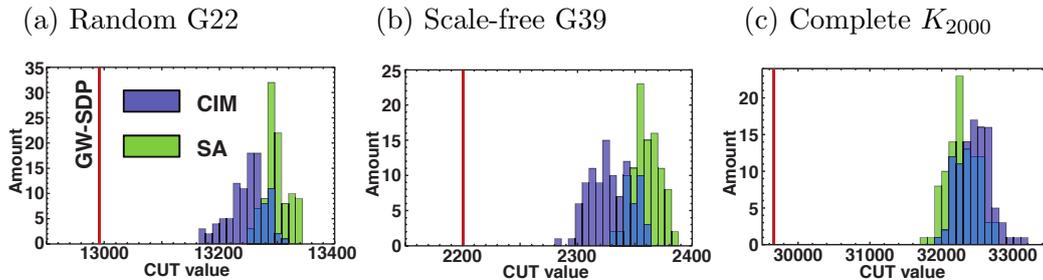


Figure 7.1. The cut statistics in 100 trials for typical topologies of graphs, random, scale-free, and complete in Table 7.1. The computation time of CIM and SA are set to be 5 ms and 50 ms, respectively. From [T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue. A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312):603–606, 2016.]. Reprinted with permission from AAAS.

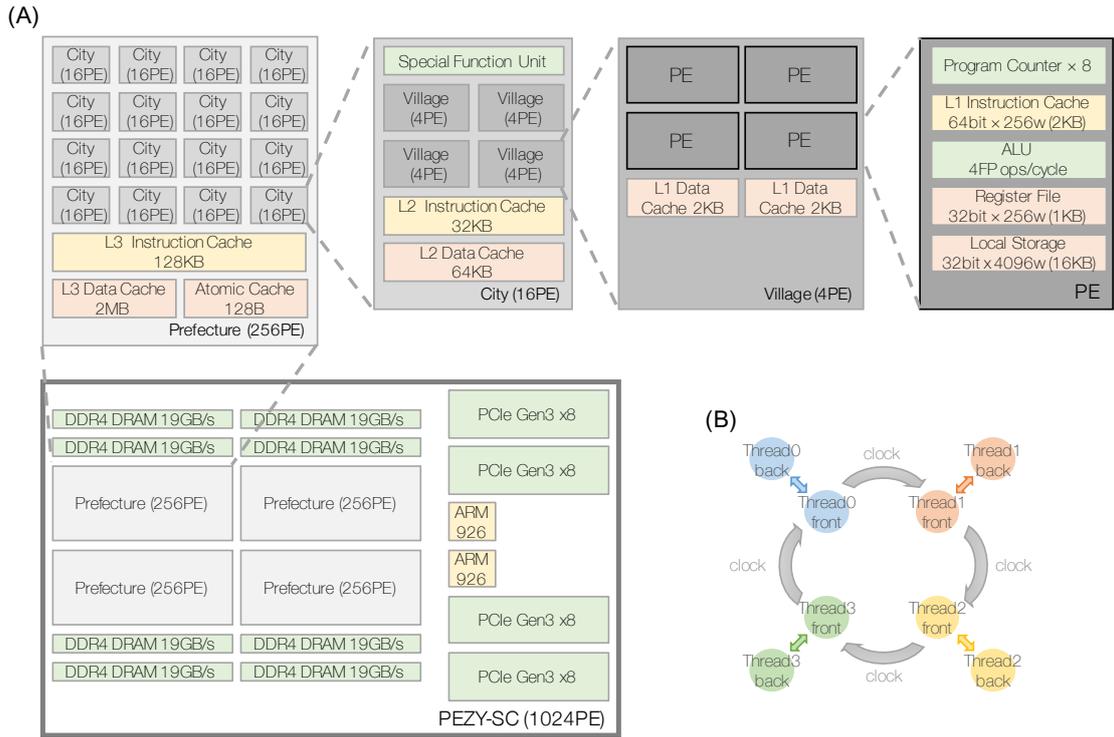


Figure 7.2. (A) The hierarchical architecture of a PEZY-SC many core processor. There are 1024 processing elements (PE) packed in a single chip. (B) Each PE core handles 8 threads independently.

SA is run on an older processor (Intel Xeon X5650 @ 2.67 GHz Westmere architecture shipped in 2010) [77]. 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 high enough as 98.8%.

7.3.2 MIMD Many-Core Processor (for HTNN)

Since HTNN is based on ordinary differential equations (a continuous-valued continuous-time system) and requires 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 Figure 7.2), which is set in Shoubu or Satsuki supercomputers at Riken (Japan). The peak performance of the processor is 3 TFlops. 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. The benchmark of the hardware itself is shown in the Appendix C.

7.4 Time To Target against Neural Network Based Heuristics

We compared the performance of HN, SA, HTNN, and CIM by solving the MAX-CUT problems on a dense graph. The particular problem instance is a complete graph, in which all pair of vertices are connected and edges are weighted by $\{+1, -1\}$ in uniform distribution. We used the identical instance for $n = 2000$ as in Reference [77] and generated a larger instance of $n = 20000$ in the same manner using rudy [154]. Figure 7.3 shows the performance on the complete graph, while the detailed computation time to target and the hardware configurations are summarized in Table 7.2.

We ran 100 different trials for the same problem instance (except for CIM experiment, which consists of 26 trials). Each solid line in Figure 7.3 indicates the ensemble average of all trials, while the lower and upper shaded lines indicate the best and worst case envelopes, respectively. Here, parameters for SA and HTNN are optimized to achieve the shortest computation time to the target which is obtained by the SDP relaxation algorithm [84]. The computation time to the SDP-produced target is shorter in the order of CIM, HN, SA, HTNN on the instance. The data from CIM in Figure 7.3 (a) are noisy due to experimental noise, but it can find better solutions than the target in all 26 trials. HN is faster than SA since HN can be regarded as a derandomized 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 Figure 7.3 (a)). It can be understood that HTNN performs much slower than HN/SA since it solves ODEs which deals with the analog variables. Note that HTNN achieves lower energy than SA in Figure 7.3 but the performance of SA heavily depends on temperature scheduling. We optimized to reach the target shorter but slower scheduling ends up lower energy generally.

The computation time to the SDP target is in the same sequence when the number of vertices increases to $n = 20000$. Here the cavity round trip time of CIM is assumed to be $10 \mu s$. Then the relative speed-up of CIM is raised to 2-3 orders of magnitude compared to other implementations.

7.5 Summary

In this chapter, the CIM performance was evaluated on relatively large instances.

- Time constant and parameter setting agrees with the numerical simulation.
- Experimental machine could achieve predicted performance in terms of computation time and solution quality.
- Also compared to SIMD and multi-thread operations on a CPU and many-core processor. CIM still show the relative speed-up.

Note that, the computational performance depends on experimental stability, which requires precise temperature control. When the system is unstable, we need to wait

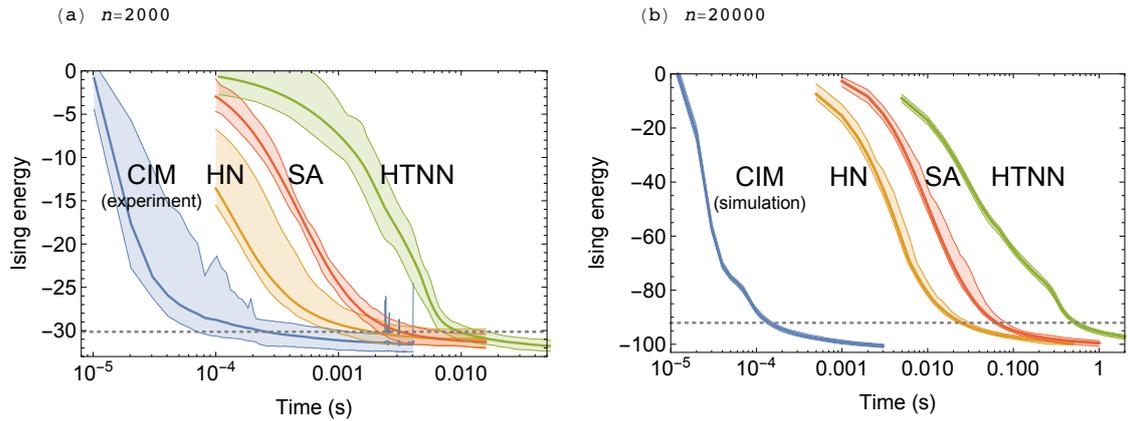


Figure 7.3. Energy descent when solving $\{+1, -1\}$ -weighted (a) $n = 2000$ and (b) $n = 20000$ complete graphs. Each thick line is the ensemble average of 100 trials (except for CIM experiment, which consists of 26 trials), while the lower and upper shaded error bars show the best and worst envelopes for each computational model. The gray dotted line is the target values $-60278/n$ and $-1841216/n$ for $n = 2000$ and 20000 , respectively, which are obtained by the SDP relaxation algorithm [84]. In the CIM simulation of $n = 20000$, the cavity round trip time is assumed to be $10 \mu\text{s}$.

the stabilization of cavity for DOPO oscillation. And also, in the results shown in this dissertation, input and output time of the problem is not included for both CIM and other implementations. There can be a communication bottleneck if we treat huge size of matrix.

Table 7.2. Time to target and hardware configurations. The best (shortest) time to reach the target value and the time to cross the ensemble averaged line are listed in the upper table. Note that the target value is obtained by the SDP relaxation algorithm which has performance guaranty to the 87% of the optimal value.

	$n = 2000$		$n = 20000$		Hardware
	Best (ms)	Avg. (ms)	Best (ms)	Avg. (ms)	
CIM	0.071	0.264	0.14	0.15	fiber DOPO+2FPGAs ($n = 2000$)
HN	0.924	1.84	23	26	CPU
SA	2.10	3.20	60	65	CPU
HTNN	7.04	9.67	500	540	PEZY-SC

Hardware	Model	Clock and Architecture	Release
FPGA	Xilinx Virtex-7 VX690T	125 MHz 693k logic cells	2010
CPU	Intel Xeon E3-1225 v3	3.2 GHz, Haswell	2013
MIMD many core	PEZY-SC	733 MHz, 1024 core	2014

Chapter 8

Conclusion

8.1 Experimental Validity and Computational Performance

In this dissertation, the measurement-feedback CIM with DOPO and FPGA is proposed in the Chapter 4. The working equation is formulated by the CSDE via truncated Wigner representation. The numerical simulations predicted the performance of CIM in Chapter 5. It can deal with the MAX-CUT on graphs of order few to several tens of thousands. It converges with several hundreds of round trips and can find the ground states in small problems and good approximate solutions in larger benchmark instances. Scaling up the problem, It showed the relative speedup on dense graphs against existing heuristics. In next Chapter 6, the numerical simulations are compared to the results from a real experimental machine. It confirmed that the CIM will find the ground states with instances up to $n = 100$ experimentally. Then, the computational performance is checked against variety of hardware implementations of metaheuristic algorithms in the Chapter 7. The final results indicate, the CIM was 6.9697 and 12.1212 times faster than HN and SA, respectively, in terms of the average computation time to the target solution given by GW-SDP on the randomly weighted complete graph of $n = 2000$. Many other results suggest the more experimental stability will contribute to the performance improvement and relative speedup will be possible when the size of the problem grows.

8.2 Possibility of Other Implementations

Our experimental success stimulated several groups and several similar implementations are reported. A group in Hewlett Packard Labs reported an all-optical system, where the coupling strength is controlled by heat oven on a chip [155]. They succeeded to implement small ($n = 4$) system on a chip. Technische Universität Dresden (TU Dresden) is working on an on-chip optical-electronic hybrid system [156]. Its experimental implementation is under examination, it is interesting to see if the optical-electrical synchronization can be seamlessly realized. A groups in MIT, Oxford, and UC Berkley are also investigating the simplified model with Gaussian approximation [157].

8.3 Practical Issues

As described in Introduction, the saturation of a single processor performance made a shift to alternative architectures. A problem-oriented Ising computer is one of the choices and the promising performance is implied. Note that the graph I/O time is excluded in the benchmark sections. Practically, the data transfer time will be not negligible and roughly $37 \mu\text{s}$ when the PCI-Express Gen3 (x8) is assumed.

The initial motivation to restrict the target to the Ising/MAX-CUT problems was because there exist polynomial time reduction from NP problems to them since they are NP-hard. But practically, this polynomial mapping cost can be critical when we search the approximate solutions to many other combinatorial optimization problems. This could be an overlooked disadvantage compared to the computational complexity theory and software implementation flexibility. For example, to represent n -city TSP in Ising model, we need n^2 spins to store a permutation as an $n \times n$ city-order matrix with two constraint terms: one-city-at-once and never-visit-again constraints [110]. A difficulty comes from the constraint must be satisfy to obtain feasible solutions, it originates from the new degree of freedom in parameter for redundant expression of the permutation. It means more highly application-specific hardware design is necessary, not just for Ising/MAX-CUT problems.

8.4 Discussion

8.4.1 Related Quantum Nature

The quantum nature of CIM is discussed from four points of view, while its contribution to the computational speedup is not known.

1. Quantum correlation (including entanglement) under the threshold [141].
2. Entanglement between inside/outside cavity by squeezing [144].
3. Skewed distribution by back action of measurement [143].

The first point, a quantum entanglement is numerically observed assuming optical coupled high Q (i.e., very low loss) cavity in Reference [141], though the contribution to the computational ability is not yet known. In the measurement-feedback system, the inter-pulse coupling is local operations and classical communication (LOCC), hence no entanglement exists among the different DOPO pulses in theory.

Regarding the second point, it is shown that the squeezing is essential to make the correlation in between the pulse in the main cavity and its counter part in the picked-off measurement path in Reference [144].

It is an open question whether the last point works as a computational resource. Since the computational effort of simulating the measurement-feedback model in positive P represented presented in Reference [143] (see Section 4.2.3) is expensive, the simulator is ported to GPUs and will be studied.

8.4.2 Further Discussion

We designed an optical-pulse-based scalable computer and studied the performance on NP-hard combinatorial optimization problems. Though the target is interesting in theoretical points of view, the real-world application is not directly solved by this. To extend applications go beyond the Ising/MAX-CUT, we present some of the examples in Appendix A.

It is still an open question: how quantum nature affects the computational performance, even if the exponential speedup is not expected. It should be studied with large problem instances compared to positive P , which is enabled by massive parallel implementation of simulation. We should care, as the multimode DOPO case in Appendix B tells, it does not necessarily mean that the physical computational error in small-sized system $n = 2, 4, 16$ directly improves the performance in larger instances. It is because the problem difficulty comes from the combinatorial structure.

It is becoming more important to focus on the application-/service-specific design of the hardware. Note that the benchmarks are done on the current generation of computer architecture, and the performance depends on the generation (see Appendix C). It is difficult to predict the future, but we are optimistic since the improvement in FPGA would be larger than that on CPU in coming few decades.

Not only the hardware for comparison, the possibility for improvement on algorithms or theory exist while we extensively studied the existing methods. A different types of Monte Carlo are considered in Appendix D, while it does not change the conclusion.

On the other hand, there is a possibility to gain the CIM performance by modifying the implementation. Appendix E sketches an idea to implement weighted graphs efficiently.

The reduced bit implementation is similar to recent architecture in neural network accelerators. In such system, the depth of bit slicing is critical. Several examples are shown in Appendix F.

8.5 Perspectives and Concluding Remarks

The evaluation of the computational performance cannot be done in a single point of view. It is difficult to determine how much abstraction level of problems we should consider. In this dissertation, we treated one of the most abstract model of optimization, Ising or MAX-CUT problems. Although, as mentioned above, more problem specific approach is necessary to achieve the performance gain.

In this dissertation, the benchmark was studied against approximation algorithms and heuristics on general-purpose processors such as CPU, GPU, and many-core processors. While it will be beyond the scope of this dissertation since there are many implementations, the relative performance against special-purpose processors such as Ising machines on FPGA or ASIC should be studied.

There is still a possibility of variants to match a good machine-problem fit, hopefully, the novel architecture would truly show the overwhelming performance. We hope this study on CIM can contribute to the enhancement of modern computer architectures.

Bibliography

- [1] J. L. Hennessy and D. A. Patterson. *Computer architecture: a quantitative approach*. Elsevier, 2011.
- [2] J. Von Neumann. First draft of a report on the edvac. *IEEE Annals of the History of Computing*, 15(4):27–75, 1993.
- [3] G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8):114–117, 1965.
- [4] W. J. Dally and B. P. Towles. *Principles and practices of interconnection networks*. Elsevier, 2004.
- [5] K. Rupp. 40 years of microprocessor trend data. <https://www.karlsruhp.net/2015/06/40-years-of-microprocessor-trend-data/>, (accessed: Nov. 8, 2017).
- [6] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 1. generalized born. *Journal of chemical theory and computation*, 8(5):1542–1555, 2012.
- [7] O. Kutter, R. Shams, and N. Navab. Visualization and GPU-accelerated simulation of medical ultrasound from CT images. *Computer Methods and Programs in Biomedicine*, 94(3):250–266, 2009.
- [8] L. Nyland, M. Harris, J. Prins, et al. Fast n-body simulation with cuda. *GPU gems*, 3(31):677–695, 2007.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS'12)*, pages 1097–1105, 2012.
- [10] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [11] N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, et al. In-datacenter performance analysis of a tensor processing unit. [arXiv:1704.04760](https://arxiv.org/abs/1704.04760) [cs.AR], 2017.

- [12] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. Mastering the game of go without human knowledge. *Nature*, 550:354–359, 2017.
- [13] A. Sebastian, T. Tuma, N. Papandreou, M. Le Gallo, L. Kull, T. Parnell, and E. Eleftheriou. Temporal correlation detection using computational phase-change memory. *Nature Communications*, 8(11115), 2017.
- [14] The standard performance evaluation corporation (spec) benchmarks. <https://www.spec.org/>, (accessed: Nov. 8, 2017).
- [15] F. Barahona, M. Grötschel, M. Jünger, and G. Reinelt. An application of combinatorial optimization to statistical physics and circuit layout design. *Operations Research*, 36(3):493–513, 1988.
- [16] D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- [17] B. H. S. Abeysekera, M. Matsui, Y. Asai, and M. Mizoguchi. Network controlled frequency channel and bandwidth allocation scheme for ieee 802.11 a/n/ac wireless lans: Ratop. In *Personal, Indoor, and Mobile Radio Communication (PIMRC), 2014 IEEE 25th Annual International Symposium on*, pages 1041–1045. IEEE, 2014.
- [18] H. Sakaguchi, K. Ogata, T. Isomura, S. Utsunomiya, Y. Yamamoto, and K. Aihara. Boltzmann sampling by degenerate optical parametric oscillator network for structure-based virtual screening. *Entropy*, 18(10):365, 2016.
- [19] M. Stolpe. Truss optimization with discrete design variables: a critical review. *Structural and Multidisciplinary Optimization*, 53(2):349–374, 2016.
- [20] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London mathematical society*, 2(1):230–265, 1937.
- [21] A. Schönhage. On the power of random access machines. In *International Colloquium on Automata, Languages, and Programming (ICALP’79)*, pages 520–529, 1979.
- [22] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O’Brien. Quantum computers. *Nature*, 464(7285):45–53, 2010.
- [23] J. Preskill. Quantum computing in the nisq era and beyond. [arXiv:1801.00862 \[quant-ph\]](https://arxiv.org/abs/1801.00862), 2018.
- [24] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 439(1907):553–558, 1992.

- [25] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science (FOCS'94)*, pages 124–134, 1994.
- [26] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC'96)*, pages 212–219, 1996.
- [27] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.
- [28] D. J. Bernstein, J. Buchmann, and E. Dahmen. *Post-quantum cryptography*. Springer, Berlin, Heidelberg, 2009.
- [29] M. Ajtai. Generating hard instances of lattice problems (extended abstract). In *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing (STOC'96)*, pages 99–108, 1996.
- [30] J. Hoffstein, J. Pipher, and J. H. Silverman. *NTRU: A ring-based public key cryptosystem*, pages 267–288. Springer, Berlin, Heidelberg, 1998.
- [31] N. Koblitz. Elliptic curve cryptosystems. *Mathematics of computation*, 48(177):203–209, 1987.
- [32] V. S. Miller. *Use of Elliptic Curves in Cryptography*, pages 417–426. Springer, Berlin, Heidelberg, 1986.
- [33] A. W. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009.
- [34] S. Aaronson. Read the fine print. *Nature Physics*, 11(4):291–293, 2015.
- [35] F. G. Brandao and K. Svore. Quantum speed-ups for semidefinite programming. [arXiv:1609.05537 \[quant-ph\]](https://arxiv.org/abs/1609.05537), 2016.
- [36] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549:195 – 202, 2017.
- [37] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508(7497):500–503, 2014.
- [38] Y. Nakamura, Y. A. Pashkin, and J. S. Tsai. Coherent control of macroscopic quantum states in a single-cooper-pair box. *Nature*, 398(6730):786–788, 1999.

- [39] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto. Layered architecture for quantum computing. *Physical Review X*, 2(3):031007, 2012.
- [40] J. R. McClean et al. Openfermion: The electronic structure package for quantum computers. [arXiv:1710.07629](https://arxiv.org/abs/1710.07629) [quant-ph], 2017.
- [41] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 1982.
- [42] I. M. Georgescu, S. Ashhab, and F. Nori. Quantum simulation. *Reviews of Modern Physics*, 86:153–185, 2014.
- [43] C. Gross and I. Bloch. Quantum simulations with ultracold atoms in optical lattices. *Science*, 357(6355):995–1001, 2017.
- [44] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, et al. Probing many-body dynamics on a 51-atom quantum simulator. [arXiv:1707.04344](https://arxiv.org/abs/1707.04344) [quant-ph], 2017.
- [45] P. A. Lee, N. Nagaosa, and X.-G. Wen. Doping a mott insulator: Physics of high-temperature superconductivity. *Reviews of Modern Physics*, 78:17–85, 2006.
- [46] S. F. Edwards and P. W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.
- [47] F. Barahona. On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.
- [48] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, and Y. Wang. The unconstrained binary quadratic programming problem: a survey. *Journal of Combinatorial Optimization*, 28(1):58–81, 2014.
- [49] H. Neven, V. S. Denchev, G. Rose, and W. G. Macready. Image recognition with an adiabatic quantum computer i. mapping to quadratic unconstrained binary optimization. [arXiv:0811.0416](https://arxiv.org/abs/0811.0416) [quant-ph], 2008.
- [50] B. Apolloni, C. Carvalho, and D. de Falco. Quantum stochastic optimization. *Stochastic Processes and their Applications*, 33(2):233–244, 1989.
- [51] T. Kadowaki and H. Nishimori. Quantum annealing in the transverse ising model. *Physical Review E*, 58(5):5355–5363, 1998.
- [52] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516):472–475, 2001.

- [53] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer. Defining and detecting quantum speedup. *Science*, 345(6195):420–424, 2014.
- [54] D. S. Steiger, T. F. Rønnow, and M. Troyer. Heavy tails in the distribution of time to solution for classical and quantum annealing. *Physical Review Letters*, 115(23):230501, 2015.
- [55] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar. Experimental signature of programmable quantum annealing. *Nature Communications*, 4(2067), 2013.
- [56] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani. How” quantum” is the D-Wave machine? [arXiv:1401.7087 \[quant-ph\]](https://arxiv.org/abs/1401.7087), 2014.
- [57] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. H. Amin, A. Y. Smirnov, M. Mohseni, and H. Neven. Computational multiqubit tunnelling in programmable quantum annealers. *Nature communications*, 7(10327), 2016.
- [58] R. Martoňák, G. E. Santoro, and E. Tosatti. Quantum annealing by the path-integral monte carlo method: The two-dimensional random ising model. *Physical Review B*, 66(9):094203, 2002.
- [59] M. Yamaoka, C. Yoshimura, M. Hayashi, T. Okuyama, H. Aoki, and H. Mizuno. A 20k-spin Ising chip to solve combinatorial optimization problems with CMOS annealing. *IEEE Journal of Solid-State Circuits*, 51(1):303–309, 2016.
- [60] C. Yoshimura, M. Hayashi, T. Okuyama, and M. Yamaoka. Fpga-based annealing processor for ising model. In *Fourth International Symposium on Computing and Networking (CANDAR’16)*, pages 436–442. IEEE, 2016.
- [61] S. Matsubara, H. Tamura, M. Takatsu, D. Yoo, B. Vatankhahghadim, H. Yamasaki, T. Miyazawa, S. Tsukamoto, Y. Watanabe, K. Takemoto, et al. Ising-model optimizer with parallel-trial bit-sieve engine. In *Proceedings of the 11th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS’17)*, pages 432–438, 2017.
- [62] Fujitsu laboratories develops new architecture that rivals quantum computers in utility. <http://www.fujitsu.com/global/about/resources/news/press-releases/2016/1020-02.html>, (accessed: Nov. 8, 2017).
- [63] Fujitsu technology facilitates application of combinatorial optimization methods to real-world problems. <http://www.fujitsu.com/global/about/resources/news/press-releases/2017/0920-03.html>, (accessed: Nov. 8, 2017).

- [64] H. Goto. Bifurcation-based adiabatic quantum computation with a nonlinear oscillator network. *Scientific Reports*, 6(21686), 2016.
- [65] R. Dridi and H. Alghassi. Prime factorization using quantum annealing and computational algebraic geometry. *Scientific Reports*, 7(43048), 2017.
- [66] A. Mott, J. Job, J.-R. Vlimant, D. Lidar, and M. Spiropulu. Solving a higgs optimization problem with quantum annealing for machine learning. *Nature*, 550(7676):375–379, 2017.
- [67] F. Neukart, G. Compostella, C. Seidel, D. von Dollen, S. Yarkoni, and B. Parney. Traffic flow optimization using a quantum annealer. *arXiv:1708.01625 [quant-ph]*, 2017.
- [68] S. Utsunomiya, K. Takata, and Y. Yamamoto. Mapping of Ising models onto injection-locked laser systems. *Optics express*, 19(19):18091–18108, 2011.
- [69] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto. Coherent Ising machine based on degenerate optical parametric oscillators. *Physical Review A*, 88(6):063853, 2013.
- [70] A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto. Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. *Nature Photonics*, 8(12):937–942, 2014.
- [71] K. Takata, A. Marandi, R. Hamerly, Y. Haribara, D. Maruo, S. Tamate, H. Sakaguchi, S. Utsunomiya, and Y. Yamamoto. A 16-bit coherent Ising machine for one-dimensional ring and cubic graph problems. *Scientific Reports*, 6(34089), 2016.
- [72] T. Inagaki, K. Inaba, R. Hamerly, K. Inoue, Y. Yamamoto, and H. Takesue. Large-scale Ising spin network based on degenerate optical parametric oscillators. *Nature Photonics*, 10:415–419, 2016.
- [73] Y. Haribara, S. Utsunomiya, and Y. Yamamoto. Computational principle and performance evaluation of coherent Ising machine based on degenerate optical parametric oscillator network. *Entropy*, 18(4):151, 2016.
- [74] Y. Haribara, S. Utsunomiya, and Y. Yamamoto. *A Coherent Ising Machine for MAX-CUT Problems: Performance Evaluation against Semidefinite Programming and Simulated Annealing*, pages 251–262. Springer Japan, Tokyo, 2016.
- [75] Y. Haribara, S. Utsunomiya, K. Kawarabayashi, and Y. Yamamoto. Quantum computing with non-classical light: A coherent Ising machine for maximum-cut problems. In J. C. Lindon, G. E. Tranter, and D. W. Koppenaal, editors, *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, pages 824–831. Academic Press, Oxford, 2017.

- [76] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto. A fully-programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, 2016.
- [77] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue. A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312):603–606, 2016.
- [78] Y. Haribara, H. Ishikawa, S. Utsunomiya, K. Aihara, and Y. Yamamoto. Performance evaluation of coherent Ising machines against classical neural networks. *Quantum Science and Technology*, 2(4):044002, 2017.
- [79] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [80] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>, 1999.
- [81] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, Berlin, Heidelberg, 1972.
- [82] C. H. Papadimitriou and K. Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1998.
- [83] M. R. Garey and D. S. Johnson. *Computers and intractability: A Guide to the Theory of NP-Completeness*, volume 29. W. H. Freeman, New York, 2002.
- [84] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
- [85] S. Arora and S. Kale. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing (STOC'07)*, pages 227–236. ACM, 2007.
- [86] P. Klein and H.-I. Lu. Efficient approximation algorithms for semidefinite programs arising from max cut and coloring. In *Proceedings of the twenty-eighth annual ACM Symposium on Theory of computing (STOC'96)*, pages 338–347. ACM, 1996.
- [87] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, 1995.

- [88] M. Yamashita, K. Fujisawa, M. Fukuda, K. Kobayashi, K. Nakata, and M. Nakata. Latest developments in the sdpa family for solving large-scale sdps. In *Handbook on semidefinite, conic and polynomial optimization*, pages 687–713. Springer, Berlin, Heidelberg, 2012.
- [89] L. Grippo, L. Palagi, M. Piacentini, V. Piccialli, and G. Rinaldi. Speedp: an algorithm to compute sdp bounds for very large max-cut instances. *Mathematical programming*, 136(2):353–373, 2012.
- [90] Y. Ye. Computational optimization laboratory. <https://web.stanford.edu/~yyye/Col.html>, (accessed: Nov. 8, 2017).
- [91] S. J. Benson, Y. Ye, and X. Zhang. Solving large-scale sparse semidefinite programs for combinatorial optimization. *SIAM Journal on Optimization*, 10(2):443–461, 2000.
- [92] S. Sahni and T. Gonzalez. P-complete approximation problems. *Journal of the ACM (JACM)*, 23(3):555–565, 1976.
- [93] S. Kahruman, E. Kolotoglu, S. Butenko, and I. V. Hicks. On greedy construction heuristics for the max-cut problem. *International Journal of Computational Science and Engineering*, 3(3):211–218, 2007.
- [94] L. Trevisan. Max cut and the smallest eigenvalue. *SIAM Journal on Computing*, 41(6):1769–1786, 2012.
- [95] S. Kale and C. Seshadhri. Combinatorial approximation algorithms for maxcut using random walks. [arXiv:1008.3938](https://arxiv.org/abs/1008.3938) [cs.DS], 2010.
- [96] C. Kuratowski. Sur le problème des courbes gauches en topologie. *Fundamenta mathematicae*, 15(1):271–283, 1930.
- [97] N. Robertson and P. D. Seymour. Graph minors VIII. a Kuratowski theorem for general surfaces. *Journal of Combinatorial Theory, series B*, 48(2):255–288, 1990.
- [98] M. Grötschel and W. R. Pulleyblank. Weakly bipartite graphs and the max-cut problem. *Operations research letters*, 1(1):23–27, 1981.
- [99] G. Orlova and Y. G. Dorfman. Finding maximum cut in a graph. *Engineering Cybernetics*, 10(3):502–506, 1972.
- [100] F. Hadlock. Finding a maximum cut of a planar graph in polynomial time. *SIAM Journal on Computing*, 4(3):221–225, 1975.
- [101] F. Barahona. The max-cut problem on graphs not contractible to K_5 . *Operations Research Letters*, 2(3):107–111, 1983.

- [102] M. Grötschel, L. Lovász, and A. Schrijver. Polynomial algorithms for perfect graphs. *North-Holland mathematics studies*, 88:325–356, 1984.
- [103] A. Galluccio, M. Loeb, and J. Vondrák. Optimization via enumeration: a new algorithm for the max cut problem. *Mathematical Programming*, 90(2):273–290, 2001.
- [104] R. Hamerly, K. Inaba, T. Inagaki, H. Takesue, Y. Yamamoto, and H. Mabuchi. Topological defect formation in 1D and 2D spin chains realized by network of optical parametric oscillators. *International Journal of Modern Physics B*, 30(25):1630014, 2016.
- [105] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [106] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.
- [107] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.
- [108] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.
- [109] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, 1984.
- [110] J. J. Hopfield and D. W. Tank. Computing with neural circuits – a model. *Science*, 233(4764):625–633, 1986.
- [111] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- [112] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- [113] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

- [114] B. Hajek. Cooling schedules for optimal annealing. *Mathematics of operations research*, 13(2):311–329, 1988.
- [115] U. Benlic and J.-K. Hao. Breakout local search for the max-cut problem. *Engineering Applications of Artificial Intelligence*, 26(3):1162–1173, 2013.
- [116] K. F. Pál. The ground state energy of the edwards-anderson Ising spin glass with a hybrid genetic algorithm. *Physica A: Statistical Mechanics and its Applications*, 223(3):283–292, 1996.
- [117] K. Hukushima and Y. Iba. Population annealing and its application to a spin glass. *AIP Conference Proceedings*, 690(1):200–206, 2003.
- [118] W. Wang, J. Machta, and H. G. Katzgraber. Population annealing: Theory and application in spin glasses. *Physical Review E*, 92:063307, Dec 2015.
- [119] S. Dash. A note on qubo instances defined on chimera graphs. [arXiv:1306.1202](https://arxiv.org/abs/1306.1202) [math.OC], 2013.
- [120] T. Ikuta, H. Imai, and Y. Yano. Solving max-cut benchmark by optimization solvers (in japanese). *RIMS Kôkyûroku*, 1941(9):49–56, 2015.
- [121] I. Dinur. The pcg theorem by gap amplification. *Journal of the ACM (JACM)*, 54(3):12, 2007.
- [122] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM (JACM)*, 45(1):70–122, 1998.
- [123] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998.
- [124] J. Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.
- [125] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing (STOC'02)*, pages 767–775, 2002.
- [126] H. J. Carmichael. *Statistical Methods in Quantum Optics 2: Non-Classical Fields*. Springer, Berlin, Heidelberg, 2008.
- [127] H. J. Carmichael. *Statistical Methods in Quantum Optics 1: Master Equations and Fokker-Planck Equations*. Springer, Berlin, Heidelberg, 2002.
- [128] R. L. Byer. Parametric oscillators and nonlinear materials. *Nonlinear Optics*, 4:47–160, 1977.

- [129] T. Kennedy and D. Walls. Squeezed quantum fluctuations and macroscopic quantum coherence. *Physical Review A*, 37(1):152, 1988.
- [130] M. D. Reid and P. D. Drummond. Quantum correlations of phase in nondegenerate parametric oscillation. *Physical Review Letters*, 60(26):2731, 1988.
- [131] J. Jing, S. Feng, R. Bloomer, and O. Pfister. Experimental continuous-variable entanglement from a phase-difference-locked optical parametric oscillator. *Physical Review A*, 74(4):041804, 2006.
- [132] A. Marandi, N. C. Leindecker, K. L. Vodopyanov, and R. L. Byer. All-optical quantum random bit generation from intrinsically binary phase of parametric oscillators. *Optics express*, 20(17):19322–19330, 2012.
- [133] M. Orszag. *Quantum optics: including noise reduction, trapped ions, quantum trajectories, and decoherence*. Springer, Berlin, Heidelberg, 2016.
- [134] P. D. Drummond and C. W. Gardiner. Generalised p-representations in quantum optics. *Journal of Physics A: Mathematical and General*, 13(7):2353–2368, 1980.
- [135] P. D. Drummond, K. J. McNeil, and D. F. Walls. Non-equilibrium transitions in sub/second harmonic generation. *Optica Acta: International Journal of Optics*, 28(2):211–225, 1981.
- [136] D. Maruo, S. Utsunomiya, and Y. Yamamoto. Truncated wigner theory of coherent ising machines based on degenerate optical parametric oscillator network. *Physica Scripta*, 91(8):083010, 2016.
- [137] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer, Berlin, Heidelberg, 3rd edition, 2004.
- [138] F. Aluffi-Pentini, V. Parisi, and F. Zirilli. Global optimization and stochastic differential equations. *Journal of optimization theory and applications*, 47(1):1–16, 1985.
- [139] H. Risken. Fokker-planck equation. In *The Fokker-Planck Equation*, pages 63–95. Springer, Berlin, Heidelberg, 1996.
- [140] G. P. Agrawal. *Nonlinear fiber optics*. Academic Press, Cambridge, Massachusetts, 5th edition, 2012.
- [141] K. Takata, A. Marandi, and Y. Yamamoto. Quantum correlation in degenerate optical parametric oscillators with mutual injections. *Physical Review A*, 92(4):043821, 2015.
- [142] R. J. Glauber. Coherent and incoherent states of the radiation field. *Physical Review*, 131(6):2766–2788, 1963.

- [143] T. Shoji, K. Aihara, and Y. Yamamoto. A quantum model for coherent Ising machine: Stochastic differential equations with replicator dynamics. [arXiv:1706.02034](https://arxiv.org/abs/1706.02034) [quant-ph], 2017.
- [144] A. Yamamura, K. Aihara, and Y. Yamamoto. A quantum model for coherent Ising machines: Discrete-time measurement feedback formulation. [arXiv:1706.02454](https://arxiv.org/abs/1706.02454) [quant-ph], 2017.
- [145] H. Wiseman and G. Milburn. Quantum theory of optical feedback via homodyne detection. *Physical Review Letters*, 70(5):548, 1993.
- [146] S. P. Lalley and T. Sellke. A conditional limit theorem for the frontier of a branching brownian motion. *The Annals of Probability*, 15(3):1052–1061, 1987.
- [147] S. Burer, R. D. C. Monteiro, and Y. Zhang. Rank-two relaxation heuristics for max-cut and other binary quadratic programs. *SIAM Journal on Optimization*, 12(2):503–521, 2002.
- [148] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10(3):673–696, 2000.
- [149] G. Zarand, F. Pazmandi, K. Pál, and G. Zimányi. Using hysteresis for optimization. *Physical Review Letters*, 89(15):150201, 2002.
- [150] T. Leleu, Y. Yamamoto, S. Utsunomiya, and K. Aihara. Combinatorial optimization using dynamical phase transitions in driven-dissipative systems. *Physical Review E*, 95:022118, Feb 2017.
- [151] K. Takata, S. Utsunomiya, and Y. Yamamoto. Transient time of an Ising machine based on injection-locked laser network. *New Journal of Physics*, 14(1):013052, 2012.
- [152] F. Rendl, G. Rinaldi, and A. Wiegele. Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. *Mathematical Programming*, 121(2):307–335, 2010.
- [153] T. Aonishi, K. Mimura, S. Utsunomiya, M. Okada, and Y. Yamamoto. Statistical mechanics of coherent ising machine – the case of ferromagnetic and finite-loading hopfield models –. *Journal of the Physical Society of Japan*, 86(10):104002, 2017.
- [154] G. Rinaldi. Rudy. <https://www.tu-chemnitz.de/~helmberg/rudy.tar.gz>, (accessed: Nov. 8, 2017).
- [155] D. Kielpinski, R. Bose, J. Pelc, T. V. Vaerenbergh, G. Mendoza, N. Tezak, and R. G. Beausoleil. Information processing with large-scale optical integrated circuits. In *Proceedings of the International Conference on Rebooting Computing (ICRC'16)*, pages 1–4, 2016.

- [156] J. Kambiz. Private communication, 2017.
- [157] W. R. Clements, J. J. Renema, Y. H. Wen, H. M. Chrzanowski, W. S. Kolthammer, and I. A. Walmsley. Gaussian optical Ising machines. [arXiv:1706.06488](https://arxiv.org/abs/1706.06488) [quant-ph], 2017.
- [158] S. Rajan. Sizing, shape, and topology design optimization of trusses using genetic algorithm. *Journal of Structural Engineering*, 121(10):1480–1487, 1995.
- [159] V. Betz and J. Rose. Vpr: a new packing, placement and routing tool for fpga research. In *Proceedings of the 7th International Workshop on Field-Programmable Logic and Applications (FPL'97)*, pages 213–222, 1997.
- [160] D. Lewis, G. Chiu, J. Chromczak, D. Galloway, B. Gamsa, V. Manohararajah, I. Milton, T. Vanderhoek, and J. Van Dyken. The stratixTM10 highly pipelined fpga architecture. In *Proceedings of the 2016 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA'16)*, pages 159–168, 2016.
- [161] S. Santra, G. Quiroz, G. V. Steeg, and D. A. Lidar. Max 2-SAT with up to 108 qubits. *New Journal of Physics*, 16(4):045006, 2014.
- [162] A. Lucas. Ising formulations of many np problems. *Frontiers in Physics*, 2:5, 2014.
- [163] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik. Finding low-energy conformations of lattice protein models by quantum annealing. *Scientific Reports*, 2(571), 2012.
- [164] G. Patera, N. Treps, C. Fabre, and G. J. De Valcarcel. Quantum theory of synchronously pumped type i optical parametric oscillators: characterization of the squeezed supermodes. *European Physical Journal D*, 56(1):123–140, 2010.
- [165] T. Sonobe and K. Kawarabayashi. Private communication, 2016.
- [166] K. Fukui and S. Todo. Order-n cluster monte carlo method for spin systems with long-range interactions. *Journal of Computational Physics*, 228(7):2629–2642, 2009.
- [167] D. A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing (STOC'04)*, pages 81–90, 2004.
- [168] S. H. Schmieta. 7th dimacs implementation challenge. <http://dimacs.rutgers.edu/Challenges/Seventh/Instances/>, (accessed: Nov. 8, 2017).
- [169] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10)*, pages 807–814, USA, 2010. Omnipress.