Convergence Analysis of Opto-Electronic Oscillator based Coherent Ising Machines

Sayantan Pramanik[1,2], Sourav Chatterjee[1], Harshkumar Oza[3*]
[1]Corporate Incubation, TCS Research & Innvovation, TATA Consultancy Services
[2]Robert Bosch Centre for Cyber-Physical Systems, Indian Institute of Science
[3]Department of Instrumentation and Applied Physics, Indian Institute of Science
sayantan.pramanik@tcs.com, sourav.chat@tcs.com, ozah@iisc.ac.in

Abstract—**Ising machines are purported to be better at solving large-scale combinatorial optimisation problems better than conventional von Neumann computers. However, these Ising machines are widely believed to be heuristics, whose promise is observed empirically rather than obtained theoretically. We bridge this gap by considering an opto-electronic oscillator based coherent Ising machine, and providing the first analytical proof that under reasonable assumptions, the OEO-CIM is not a heuristic approach. We find and prove bounds on its performance in terms of the expected difference between the objective value at the final iteration and the optimal one, and on the number of iterations required by it. In the process, we emphasise on some of its limitations such as the inability to handle asymmetric coupling between spins, and the absence of external magnetic field applied on them (both of which are necessary in many optimisation problems), along with some issues in its convergence. We overcome these limitations by proposing suitable adjustments and prove that the improved architecture is guaranteed to converge to the optimum of the relaxed objective function.**

Index Terms—**coherent Ising machine, stochastic approximation, convergence analysis, noisy gradient-descent**

I. INTRODUCTION

Recent developments in photonics-based analog computing have triggered a global exigency towards realizing new computational paradigms [1]–[3] that can potentially outperform conventional digital computers in executing challenging computational tasks, such as solving combinatorial optimisation problems that are known to be NP-hard [3]–[5]. Many NP-hard combinatorial optimisation problems can be efficiently mapped to a ground-state-search problem of the Ising model [6] – which is basically a mathematical abstraction of magnetic systems describing the behavior of competitively interacting spins or angular momenta of fundamental particles [7], [8]. Such an Ising model of coupled artificial spins can be physically realized using various systems ranging from Josephson junction [9], trapped ions [10], to optical states [11]. Among these possible realizations, the optics-based approach involving coherent Ising machines (CIMs) is advantageous because of operability at room temperature, the requirement of cost-effective off-the-shelf components, miniaturisation possibilities over integrated photonic circuits, and the scope to implement dense and flexible coupling topologies [12], [13]. Furthermore, being intrinsically gain-dissipative

systems, CIMs are naturally prone towards avoiding local energy minima and thus achieving optimal solutions [13], [14]. Given a very short-lived evolution to the ground state or optimal solution, the Ising machines promise significant speed up over conventional (digital) computers in solving hard optimisation problems [13], [15], efficient solution of which is central to several industry-verticals [3], [13].

CIMs belong to the non-von Neumann computing architecture. They compute very differently from conventional (digital) computers, digital annealers, quantum annealers, or even gate-based quantum computers. In that light, they exhibit a few distinct advantages over many other computing approaches. This includes higher speed and efficiency over classical computing, for instance: a few hundred THz bandwidth in optical compared to that of a few GHz in classical computing; along with inherent parallel operations and in-memory computation at lower energy costs. Next, as opposed to the alleged effectiveness of the quantum approximate optimisation algorithm (QAOA) [16] in efficiently solving NP-hard optimisation problems against commercial solvers executed on digital computers, the benchmarking results obtained with CIM are more promising and convincing [17]. Furthermore, unlike the gate and annealing based model of quantum computation, the operation of CIMs is not limited by factors such as scalability, high noise, and cryogenic cooling [17]. Lastly, in absence of noise the anneal time can be significantly slower for quantum annealing based approaches which is not the case for CIMs – being optically driven it is not subjected to thermal noise. Finally, digital annealers can be potentially outperformed by CIMs, since unlike CIMs they are practically limited by digital CMOS hardware-constraints.

Quite a few methods of experimentally realizing CIMs have been reported in literature. They all commonly use optical setups along with programmable fast feedback electronics to implement Ising spin networks with bistable coherent optical states [13]. This enables them to naturally approach the optimal states of the programmed Hamiltonian. These methods of realizing Ising spin systems with a network of coupled optical states can be broadly classified into three types: (a) CIM based on mutually coupled lasers [18] (b) degenerate optical parametric oscillator (DOPO) [19], [20] based CIMs [11], and (c) opto-electronic oscillator (OEO) based CIM [13], [21]. Moreover, the DOPO-based CIM method can be further

---

sub-classified into two architectures, depending upon their ways of obtaining the effective spin-spin interaction between the realized Ising spins: $\sigma_{i\,(j)} \in \{\pm 1\}$, as specified by the coupling matrix $\boldsymbol{J}$ and the linear (Zeeman) term $\boldsymbol{h}$ representing the inhomogeneous magnetic field strength in the Ising model Hamiltonian [22]:

$$H = \frac{1}{2}\sum_{i,j}^{n} \boldsymbol{J}_{ij}\sigma_i\sigma_j + \sum_{i}^{n} \boldsymbol{h}_i\sigma_i, \qquad (1)$$

where i (or j) represents each spin index or each vertex of an (un)directed spin graph. Operationally, Eq. (1) denotes the cost (energy, or objective value) of a Quadratic Unconstrained Binary Optimisation (QUBO) problem [21], and the σ_is represent the binary decision variables. Among these two architectures, first is the optical delay line (ODL) architecture where a part of the time-delayed signal is redirected back into the fibre loop [23], [24]; while the other is the measurement-feedback (MFB) architecture where the feedback signal is computed electronically based on the measurement performed by a homodyne detector [4], [5], [15]. Recently, even quite a few quantum models for CIMs, based on DOPO and MFB architectures, have been theoretically developed and evaluated [12], [22], [25], [26]. Although more popular, the nonlinear DOPO-based coherent optical state generation process is neither quite scalable nor very resource efficient as it requires high power laser systems and temperature-regulated nonlinear elements, resulting into large setup footprints. On the contrary, the OEO-based CIM architecture involving self-feedback, which was first proposed in [13], offers a higher stability at a much lower form-factor. Due to its cost-effectiveness and complex nonlinear dynamics, this architecture has been lately gaining a lot of attention [21], [27]. Given that these advantages are instrumental for various applications including cryptography, microwave generation, and optical neuronal computing [13], henceforth, for our rigorous convergence analyses we will focus upon this OEO-based CIM architecture.

II. OEO-BASED COHERENT ISING MACHINES

An OEO-based CIM architecture fundamentally relies upon the idea that large and regulated artificial spin networks can be created by exploiting the rich bifurcation structure of OEOs. Building upon this central idea, in this work, we have adapted the compact CIM setup, that was first demonstrated by Bohm et al. in 2019 [13], as per our requirement for the convergence analysis. A conceptual schematic of this setup is presented in Fig. 1. As shown in the figure, a continuous wave laser beam is subjected to nonlinear opto-electronic feedback system via amplitude modulation followed by a direct photodetection of the optical signal. This enables the mapping of OEO network to a network of Ising spins. More specifically, the amplitude modulation of the input laser beam is internally performed by a Mach-Zehnder (electro-optic amplitude) modulator (MZM), set to operate in the linear regime by imparting a constant bias of $\pi/4$ [28]. The polarization controller contains a combination of two quarter- and one half-wave plates, and thus converts an arbitrarily polarized beam into the desired input polarization
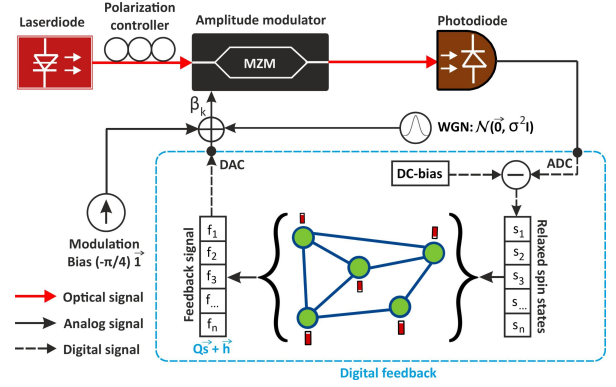

Fig. 1: A conceptual schematic of our OEO-based CIM setup (adapted from [13]). This diagram includes the modifications proposed in Sec. VIII such as the use of step-sizes β_k that are dependant on the iteration-number k, and the modulation of the Gaussian white noise with the step-size.

for the amplitude modulator. In a nutshell, this CIM setup generates the artificial spins in a feedback induced pitchfork bifurcation and encodes them in the intensity of coherent states. Due to its cost-effectiveness, the OEO-based CIM was named "a poor man's CIM" by its first demonstrators in [13]. In this article, we have referred to its customized design as Economical Coherent Ising Machines and henceforth abbreviated it as ECIM. In the ECIM, each binary variable σ_i is relaxed [29] to s_i such that the latter takes continuous values in $[-\frac{1}{2}, \frac{1}{2}]$. The objective function, under this relaxation, takes the form:

$$H(\boldsymbol{s}) = \frac{1}{2}\boldsymbol{s}^{\top}\boldsymbol{J}\boldsymbol{s}, \qquad (2)$$

which is written in the vector-form. Accordingly, the relaxed optimisation problem is given by: $\min_{\boldsymbol{s}\in[-1/2,1/2]^n} H(\boldsymbol{s})$. Once the optimal relaxed decision variables $\boldsymbol{s}^*$ have been obtained, they are "rounded off" to the discrete domain using an element-wise sign function: $\boldsymbol{\sigma}^* = \text{sgn}(\boldsymbol{s}^*)$.

As mentioned in [13], the iterative equations of the ECIM, again in vector formats, are as follows:

$$\boldsymbol{f}^{(k)} = \alpha\boldsymbol{s}^{(k)} - \beta\boldsymbol{J}\boldsymbol{s}^{(k)} \qquad (3)$$

$$\boldsymbol{s}^{(k+1)} = \cos^2\left(\boldsymbol{f}^{(k)} - \frac{\pi}{4}\mathbf{1} + \boldsymbol{\zeta}^{(k)}\right) - \frac{1}{2}\mathbf{1}, \qquad (4)$$

where, k denotes the iteration-number, $\boldsymbol{f}^{(k)}$ is referred to as the feedback vector, and $\boldsymbol{\zeta}^{(k)}$ is a white Gaussian noise whose importance is discussed in detail in Sec. VI. It is noteworthy that the non-linear, trigonometric transfer function used here ensures that the decision variables never leave the feasible region as the iterations progress. Finally, we briefly allude to the fact that the ECIM caters only to problems where the coupling matrix $\boldsymbol{J}$ is symmetric, and the linear term $\boldsymbol{h}^{\top}\boldsymbol{s}$ is absent. This has been emphasised and remedied in Sec. VII, where unlike [30], the linear term has been included without banking on ancillary spins. Fig. 1 portrays a schematic of the ECIM along with the modifications proposed in this paper, such as using a β_k which varies with iterations and modulating

the noise with β_k. These modifications are expounded on in sections VII and VIII.

III. MAIN CONTRIBUTIONS

Before proceeding further, we describe the key contributions made in this paper. The principal argument here is that, contrary to prior belief (as stated in [31]), the ECIM is not a heuristic, but a system whose performance on minimizing the relaxed objective function can be rigorously analysed. By drawing similarities between the former and noisy gradient-descent, we provide an alternate perspective to the ECIM. Further, we immaculately extend it by considering asymmetric coupling matrices and including external magnetic fields without relying on ancillary spins [30]. While theoretically analysing the ECIM's performance, we mathematically justify some hithertofore empirical observations such as the effect of the coupling term β and the role of noise. We also highlight that through the use of the ECIM in its present form, the expected value of the relaxed objective function converges in a region around the optimal value, find the dependence of the difference on the parameters, and calculate an upper bound on the number of iterations required to obtain a certain level of proximity to the optimal objective value. Finally, we address this issue by resorting to gradually-decreasing values of β which ensures convergence to the optimal value *with probability* 1, and also find the rate of convergence of the relaxed objective function.

IV. PRELIMINARY REMARKS AND ASSUMPTIONS

In our analysis of the ECIM's convergence, we assume that J has real eigenvalues, and foist the following conditions on the hyperparameters:

1) In Eq. (3), β is restricted to be a positive quantity, while the value of α is fixed to 1,
2) The noise $\boldsymbol{\zeta}^{(k)}$ is sampled from the normal distribution $\mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$, with $\boldsymbol{I}$ being the $n \times n$ identity matrix.

Furthermore, the following preliminary remarks and notations may help with the ensuing analysis, which will be followed throughout the rest of the paper:

1) $||\boldsymbol{v}||$ denotes the l_2 norm of the vector $\boldsymbol{v}$,
2) The number of relaxed decision variables is given by n, such that $\boldsymbol{s} \in [-1/2, 1/2]^n$,
3) Linear terms will henceforth be included into the objective function:
$$E(\boldsymbol{s}) = \frac{1}{2}\boldsymbol{s}^\top \boldsymbol{J} \boldsymbol{s} + \boldsymbol{h}^\top \boldsymbol{s}, \qquad (5)$$
and the optimisation problem accordingly becomes:
$$\min_{\boldsymbol{s} \in [-1/2, 1/2]^n} E(\boldsymbol{s}) \qquad (6)$$
4) We use the shorthand $E^* = \min_{\boldsymbol{s} \in [-1/2, 1/2]^n} E(\boldsymbol{s})$,
5) $\nabla E(\boldsymbol{s}) := \nabla_{\boldsymbol{s}} E(\boldsymbol{s}) = \frac{1}{2}(\boldsymbol{J} + \boldsymbol{J}^\top)\boldsymbol{s} + \boldsymbol{h} =: \boldsymbol{Q}\boldsymbol{s} + \boldsymbol{h}$ is used to represent the gradient of the objective function with respect to $\boldsymbol{s}$. We use c^2 to denote the upper-bound on the squared l_2 norm of the gradient for some $c < \infty$:
$$||\nabla E(\boldsymbol{s})||^2 \le c^2 \qquad (7)$$

Alternatively, it may be said that the function is Lipschitz-continuous.

6) It is easy to observe that the Hessian of $E(\boldsymbol{s})$ is $\boldsymbol{Q}$. Let λ_i, $i \in [n]$, denote the eigenvalues of $\boldsymbol{J}$, with $\bar{\lambda} := \max_i \lambda_i$. This implies that $E(\boldsymbol{s})$ is Lipschitz-smooth with a constant $\bar{\lambda}$,
7) We define $\mathcal{F}_k = \{\boldsymbol{s}^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\zeta}^{(1)}, \dots, \boldsymbol{\zeta}^{(k-1)}\}$ to be the increasing sequence of σ-fields that captures the history of the ECIM iteration process,
8) $\mathbb{E}[X|\mathcal{F}]$ denotes the expectation of the random variable X, conditioned on the σ-field $\mathcal{F}$; $\mathbb{E}[X|\mathcal{F}] = X$ if X is $\mathcal{F}$-measurable, and $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$ otherwise. Additionally, if the random variable Y is $\mathcal{F}$-measurable and X is not, then $\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X]$.

Finally, we assume that the objective function obeys the PL inequality [32] with some $\mu > 0$, i.e.,
$$\frac{1}{2}||\nabla E(\boldsymbol{s})||^2 \ge \mu(E(\boldsymbol{s}) - E^*), \qquad (8)$$

This assumption imposes invexity on the objective function, which is a less stringent condition than convexity [32]. It is noteworthy that equations (7) and (8) do not contradict each other, unlike the argument in [33], since in our case:
$$||\boldsymbol{s}_1 - \boldsymbol{s}_2||^2 \le n, \ \forall \boldsymbol{s}_1, \boldsymbol{s}_2 \in [-1/2, 1/2]^n \qquad (9)$$

V. COHERENT ISING MACHINE AND GRADIENT-DESCENT

With the details mentioned in Section IV, we begin by establishing an equivalence between Eq. (4) and first-order, noisy gradient-descent [34], the latter of which is given by[1]:
$$\boldsymbol{s}^{(k+1)} = \boldsymbol{s}^{(k)} - \gamma_k \nabla E(\boldsymbol{s}^{(k)}) + \boldsymbol{\zeta}^{(k)}, \qquad (10)$$
where, $\gamma_k > 0$ is the step-size at the kth iteration.

On the other hand, we expand Eq. (4) about $\frac{\pi}{4}\mathbf{1}$ using Taylor's series up to the first order:
$$\boldsymbol{s}^{(k+1)} \approx \cos^2\left(\frac{\pi}{4}\mathbf{1}\right) - \left(\nabla_{\boldsymbol{\theta}} \cos^2 \boldsymbol{\theta}\right)_{\boldsymbol{\theta}=\frac{\pi}{4}\mathbf{1}}(\boldsymbol{f}^{(k)} + \boldsymbol{\zeta}^{(k)}) - \frac{1}{2}\mathbf{1}$$
$$= \boldsymbol{f}^{(k)} + \boldsymbol{\zeta}^{(k)} = \alpha\boldsymbol{s}^{(k)} - \beta\boldsymbol{J}\boldsymbol{s}^{(k)} + \boldsymbol{\zeta}^{(k)}$$
$$(11)$$
where $o(||\boldsymbol{f}^{(k)} + \boldsymbol{\zeta}^{(k)}||)$ terms have been ignored. It is immediately apparent that equations (10) and (11) are equivalent when $\boldsymbol{J} = \boldsymbol{J}^\top$, $\boldsymbol{h} = \mathbf{0}$, $\alpha = 1$, and $\gamma_k = \beta$. Thus, the Ising machine may be interpreted as noisy gradient-descent with β as the step-size. Further, keeping this perspective in mind, it is natural to replace the symmetric couple matrix with an asymmetric one, and to include the external magnetic field term in the ECIM, resulting in Eq. (3) to be modified to:
$$\boldsymbol{f}^{(k)} = \alpha\boldsymbol{s}^{(k)} - \beta\nabla E(\boldsymbol{s}^{(k)}) = \alpha\boldsymbol{s}^{(k)} - \beta(\boldsymbol{Q}\boldsymbol{s}^{(k)} + \boldsymbol{h}), \ (12)$$
which is the same equation as arrived in [21], for a symmetric coupling matrix, through the use of an ancillary spin [30]. Establishing a relationship between noisy gradient-descent and the ECIM leaves the latter susceptible to analysis using stochastic approximation methods, which we begin in Sec. VII.

---

[1]strictly speaking, the update-equation for gradient methods with errors is $\boldsymbol{s}^{(k+1)} = \boldsymbol{s}^{(k)} - \gamma_k(\nabla E(\boldsymbol{s}^{(k)}) + \boldsymbol{\zeta}^{(k)})$. We will eventually modify the ECIM's equations within this section such that they correspond exactly with the above equation.

VI. NECESSITY OF NOISE

Before delving into the convergence-analysis of the ECIM, we take a short detour to discuss the importance of noise and the role it plays in different problem-scenarios. It has been argued in [13] that noise helps the ECIM escape unstable fixed points of the iteration scheme in Eq. (4), which in the absence of an external magnetic field $\boldsymbol{h}$, is given by $\boldsymbol{s} = \boldsymbol{0}$. In the presence of linear terms and absence of noise, the fixed points are such that $\boldsymbol{s}^{(k+1)} = \boldsymbol{s}^{(k)}$. From Eq. (12) we see that these correspond to the stationary points of $E(\boldsymbol{s})$ where $\nabla E(\boldsymbol{s}) = \boldsymbol{0}$, which in the interior of the feasible region may include global/local minima, maxima, and saddle points. The exact nature and number of such stationary points depends on the matrix $\boldsymbol{Q}$. For various types of $\boldsymbol{Q}$, we briefly discuss on the type and number of stationary points and the significance of noise in each case. It must, nonetheless, be noted that these stationary points may fall within, without, or even on the boundary of the feasible region.

1) First, we consider $\boldsymbol{Q}$ to be a positive definite matrix. In this case, the relaxed objective function is strictly convex and admits a single stationary point which is the global minimum. If $\boldsymbol{Q}$ is positive semidefinite, then there exist multiple global minima. The presence of noise appears to be inconsequential for such convex problems.

2) Conversely, if $\boldsymbol{Q}$ is negative (semi)definite, then the relaxed objective function has (multiple) maxima, and the solution is known to lie on the boundary of the feasible region. In this case, noise helps the ECIM escape from the maxima in these concave problems.

3) Finally, if $\boldsymbol{Q}$ is has a mix of both positive and negative eigenvalues, then the problem has only saddle points, which the noise helps escape from. Here, too, the solution lies on the boundary.

VII. CONVERGENCE ANALYSIS

We embark on the analysis of the economic CIM by first setting $\alpha = 1$ and employing Taylor's series again, but this time expanding the objective function $E(\boldsymbol{s}^{(k+1)})$ about $\boldsymbol{s}^{(k)}$ instead, to get:

$$
\begin{aligned}
E(\boldsymbol{s}^{(k+1)}) &= E(\boldsymbol{s}^{(k)} - \beta \nabla E(\boldsymbol{s}^{(k)}) + \boldsymbol{\zeta}^{(k)}) \\
&= E(\boldsymbol{s}^{(k)}) - (\beta \nabla E(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)})^\top \nabla E(\boldsymbol{s}^{(k)}) + \\
&\quad (\beta \nabla E(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)})^\top \boldsymbol{Q} (\beta \nabla E(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)}) \\
&\leq E(\boldsymbol{s}^{(k)}) - (\beta \nabla E(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)})^\top \nabla E(\boldsymbol{s}^{(k)}) + \\
&\quad \bar{\lambda} || \beta \nabla E(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)}||^2 \qquad (13)
\end{aligned}
$$

where the inequality above is a consequence of the fact that $\boldsymbol{v}^\top \boldsymbol{J} \boldsymbol{v} \leq \bar{\lambda} ||\boldsymbol{v}||^2$ for any $\boldsymbol{v} \in \mathbb{R}^n$. Next, we subtract E^* from both sides and take expectation conditioned on $\mathcal{F}_k$, and note the following facts:

1) E^*, $E(\boldsymbol{s}^{(k)})$ and $\nabla E(\boldsymbol{s}^{(k)})$ are $\mathcal{F}_k$-measurable, while $\boldsymbol{\zeta}^{(k)}$ is not,

2) Further, $\mathbb{E}[\boldsymbol{\zeta}^{(k)}|\mathcal{F}_k] = \mathbb{E}[\boldsymbol{\zeta}^{(k)}] = \boldsymbol{0}$, $\mathbb{E}[||\boldsymbol{\zeta}^{(k)}||^2|\mathcal{F}_k] = \mathbb{E}[||\boldsymbol{\zeta}^{(k)}||^2] = n\sigma^2$,

applying which, we arrive at:

$$
\begin{aligned}
\mathbb{E}[E(\boldsymbol{s}^{(k+1)}) - E^*|\mathcal{F}_k] &\leq [E(\boldsymbol{s}^{(k)}) - E^*] - \beta ||\nabla E(\boldsymbol{s}^{(k)})||^2 \\
&\quad + \bar{\lambda}(\beta^2 ||\nabla E(\boldsymbol{s}^{(k)})||^2 + n\sigma^2) \quad (14)
\end{aligned}
$$

which, on using equations (7) and (8), becomes:

$$
\begin{aligned}
\mathbb{E}[E(\boldsymbol{s}^{(k+1)}) - E^*|\mathcal{F}_k] &\leq (1 - 2\beta\mu)[E(\boldsymbol{s}^{(k)}) - E^*] \\
&\quad + \bar{\lambda}(\beta^2 c^2 + n\sigma^2) \quad (15)
\end{aligned}
$$

Finally, we take unconditional expectation on both sides of the equation above to obtain:

$$
\mathbb{E}[E(\boldsymbol{s}^{(k+1)}) - E^*] \leq (1 - 2\beta\mu)\mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] + \bar{\lambda}(\beta^2 c^2 + n\sigma^2) \tag{16}
$$

At this point, we highlight a minor problem with the convergence of the ECIM and hypothesise that $\mathbb{E}[E(\boldsymbol{s}^{(k)})]$ cannot get arbitrarily close to E^* with a suitably-chosen, constant step-size β.

We formalise our claim as:

$$
\liminf_{k \to \infty} \mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \leq \frac{\bar{\lambda}}{2\mu}\left(\beta c^2 + \frac{n\sigma^2}{\beta}\right) \tag{17}
$$

and set about proving it by contradiction, along the lines of that in [35]. To do so, we assume the above statement to be false, i.e., for some $\epsilon > 0$:

$$
\liminf_{k \to \infty} \mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \geq \frac{\bar{\lambda}}{2\mu}\left(\beta c^2 + \frac{n\sigma^2}{\beta}\right) + 2\epsilon \tag{18}
$$

Now suppose there exists a k_0 which is large enough such that for all $k \geq k_0$, we have:

$$
\mathbb{E}[E(\boldsymbol{s}^{(k)})] \geq \liminf_{k \to \infty} \mathbb{E}[E(\boldsymbol{s}^{(k)})] - \epsilon \tag{19}
$$

Adding equations (18) and (19):

$$
\mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \geq \frac{\bar{\lambda}}{2\mu}\left(\beta c^2 + \frac{n\sigma^2}{\beta}\right) + \epsilon \tag{20}
$$

Substituting this back into Eq. (16), we end up with the following recurrence inequality:

$$
\mathbb{E}[E(\boldsymbol{s}^{(k+1)}) - E^*] \leq \mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] - 2\beta\mu\epsilon \tag{21}
$$

We then unroll the above inequality for K steps to get:

$$
\mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \leq \mathbb{E}[E(s^{(k-K)}) - E^*] - 2\beta\mu\epsilon K \tag{22}
$$

But at this recall we note that E^* is the optimal objective value, and hence, $(E(\boldsymbol{s}^{(k)}) - E^*) \geq 0$ for any k. Further, the objective values are bounded, which means that the above recurrence cannot continue indefinitely. This leads to the desired contradiction, thus completing the proof of our claim.

A keen observation of Eq. (17) confirms our initial speculation that one may not be able to select a suitable low β to get as close to the optimal objective value as required. If a low value of β is chosen, then the second term in the equation increases in magnitude, according us no benefit towards reducing the expected gap between $E(\boldsymbol{s}^{(k)})$ and E^*. This second term has an even greater effect and significance in larger problems with a larger number of decision variables n, indicating a

poor scaling of the ECIM in its present form. This helps explain the observation in Fig. 3(c) of [36] that the TTS ("Time-to-Solution", as defined in the same reference) first decreases with increasing β and then, counter-intuitively, starts increasing; and that there is a minute increase in TTS as σ^2 is increased. Granted, that one may opt to reduce the noise-variance to counter the effect of n/β, but this leaves the ECIM prone to get stuck at local minima, saddle points, or maxima, as discussed in Sec. VI.

VIII. PROPOSED MODIFICATIONS

To offset the difficulties in convergence of the ECIM, as discussed at the end of the last section, we first present a slight shift in perspective of the ECIM's update equations, which will offer no immediate benefit, but facilitate the modifications that will be communicated in the latter half of this section. The proposition is simple: that the white Gaussian noise be scaled with the step-size β such that equations (3) and (4) are modified to:

$$\boldsymbol{f}^{(k)} = \alpha\boldsymbol{s}^{(k)} - \beta_k\left(\nabla E(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)}\right), \qquad (23)$$

and

$$\boldsymbol{s}^{(k+1)} = \cos^2\left(\boldsymbol{f}^{(k)} - \frac{\pi}{4}\mathbf{1}\right) - \frac{1}{2}\mathbf{1}, \qquad (24)$$

respectively, where β_k is the step-size at iteration k. At face-value, nothing appears to have changed, but we posit that with gradually-decreasing step-sizes, the ECIM converges to the optimal objective value "almost surely".

However, before discussing about diminishing step-sizes, we repeat the analysis in Sec. VII with $\beta = \beta_k$ to arrive at the modified versions of equations (15) and (26):

$$\mathbb{E}[E(\boldsymbol{s}^{(k+1)}) - E^*|\mathcal{F}_k] \leq (1 - 2\beta_k\mu)[E(\boldsymbol{s}^{(k)}) - E^*] \\ + \bar{\lambda}\beta_k^2(c^2 + n\sigma^2), \quad (25)$$

and

$$\mathbb{E}[E(\boldsymbol{s}^{(k+1)}) - E^*] \leq (1 - 2\beta_k\mu)\mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \\ + \bar{\lambda}\beta_k^2(c^2 + n\sigma^2). \quad (26)$$

A. With a Constant Step-Size

With the modified equations presented above, using techniques similar to the ones employed previously, it may be shown that with constant step-sizes $\beta_k = \beta$:

$$\liminf_{k \to \infty} \mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \leq \frac{\bar{\lambda}}{2\mu}\beta(c^2 + n\sigma^2) \qquad (27)$$

It is important to note here that the step-size β may be controlled by the user to make $E(\boldsymbol{s}^{(k)})$ get arbitrarily close to E^*. Further, it is also possible to find an estimate of the number of iterations κ required such that:

$$\min_{0 \leq k \leq \kappa} \mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \leq \frac{\bar{\lambda}}{2\mu}\beta(c^2 + n\sigma^2) + \epsilon, \qquad (28)$$

where $\epsilon > 0$ and

$$\kappa \leq \left\lfloor \frac{E(\boldsymbol{s}^{(0)}) - E^*}{2\beta\mu\epsilon} \right\rfloor \qquad (29)$$

This is easily obtained by setting both k and K to κ in Eq. (22) such that:

$$0 \leq \mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*] \leq [E(\boldsymbol{s}^{(0)}) - E^*] - 2\beta\mu\epsilon\kappa \qquad (30)$$

We would like to reiterate that merely scaling the noise term with the step-size is equivalent to using noise with a lower variance and does not provide any additional benefits with constant step-sizes, but the variation in perspective allows us to inspect the ECIM with decreasing step-sizes, that we now foray into.

B. With Decreasing Step-Sizes

Here, we prove that it is possible to drive $E(\boldsymbol{s}^{(k)})$ to E^* with probability 1 with the help of constantly decreasing step-sizes. As is customary in machine learning literature, we consider step-sizes of the following form:

$$\sum_{k=0}^{\infty} \beta_k = \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \beta_k^2 < \infty \qquad (31)$$

One popular example of a sequence of step-sizes that follows the above criteria is $\beta_k = \beta_0/(k+1)^r$ where, $\beta_0 > 0$ and $r \in (0.5, 1]$.

With these step-sizes, we note that in Eq. (25), the terms $(\mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*])$, $2\beta_k\mu(\mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*])$ and $\bar{\lambda}\beta_k^2(c^2 + n\sigma^2)$ are non-negative and $\mathcal{F}_k$-measurable. Further, the latter sequence is summable, i.e.:

$$\bar{\lambda}(c^2 + n\sigma^2)\sum_{k=0}^{\infty} \beta_k^2 < \infty. \qquad (32)$$

This enables us to invoke the Supermartingale Convergence Theorem (SMCT, as mentioned in Proposition 2.2 of [37]), which guarantees that:

1) $E(\boldsymbol{s}^{(k)}) - E^*$ converges to some non-negative value almost surely, and

2)
$$\sum_{k=0}^{\infty} 2\beta_k\mu(E(\boldsymbol{s}^{(k)}) - E^*) < \infty \text{ w.p. } 1 \qquad (33)$$

Now we show that not only does $(E(\boldsymbol{s}^{(k)}) - E^*)$ converge to a non-negative value, but that it converges to 0 almost surely, essentially meaning that $E(\boldsymbol{s}^{(k)}) \to E^*$ a.s. To demonstrate this, we fall back on proof by contradiction again, and consider an $\epsilon > 0$ and a $k_0 > 0$ such that for all $k \geq k_0$,

$$(E(\boldsymbol{s}^{(k)}) - E^*) \geq \epsilon, \qquad (34)$$

which leads to:

$$\sum_{k=k_0}^{\infty} 2\beta_k\mu(E(\boldsymbol{s}^{(k)}) - E^*) \geq 2\mu\epsilon\sum_{k=k_0}^{\infty} \beta_k, \qquad (35)$$

the right hand side of which is ∞ as per the first part of Eq. (31). But this contradicts the consequence of SMCT shown in Eq. (33), which is known to be true. Thus, the assumption in Eq. (34) is false and $(E(\boldsymbol{s}^{(k)}) - E^*) \to 0$ almost surely, thus concluding our proof.

Further, it is also possible to get the rate of decrease of $\mathbb{E}[E(\boldsymbol{s}^{(k)}) - E^*]$ with diminishing step-sizes by a direct application of Lemma 2.1 in [38] on Eq. (26).

IX. Conclusion and Outlook

In this paper, we demonstrated that ECIMs are not heuristics, unlike they were previously believed to be. This was, however, done in a restricted setting where only a linear approximation of a single trigonometric transfer function was considered, and α was fixed to 1. We conjecture that the proofs presented here may be extended to the more holistic case that generalises these aforementioned constraints. Besides, the performance of the ECIM was theoretically analysed only for the instance where the binary decision variables were relaxed to take continuous values. As future work, the analysis may be carried out by considering a conjunction of Ising Machines and rounding-off methods (such as branch-and-bound [39]), the latter of which may be used to convert the decision variables back to the discrete domain. Overall, this paper paves the way for rigorous analysis of Ising Machines which may prove to be compelling alternatives towards solving QUBO problems.

X. Acknowledgement

References

[1] S. Abel, F. Horst, P. Stark *et al.*, "Silicon photonics integration technologies for future computing systems," in *2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)*. IEEE, 2019, pp. 1–3.

[2] Y. Yamamoto, T. Leleu, S. Ganguli, and H. Mabuchi, "Coherent ising machines — quantum optics and neural network perspectives," *Applied Physics Letters*, vol. 117, no. 16, 2020.

[3] G. Mourgias-Alexandris, H. Ballani, N. G. Berloff *et al.*, "Analog iterative machine (aim): using light to solve quadratic optimization problems with mixed variables," *arXiv preprint arXiv:2304.12594*, 2023.

[4] T. Honjo, T. Sonobe, K. Inaba *et al.*, "100,000-spin coherent ising machine," *Science advances*, vol. 7, no. 40, p. eabh0952, 2021.

[5] P. L. McMahon, A. Marandi, Y. Haribara *et al.*, "A fully programmable 100-spin coherent ising machine with all-to-all connections," *Science*, vol. 354, no. 6312, pp. 614–617, 2016.

[6] A. Lucas, "Ising formulations of many NP problems," *Frontiers in Physics*, vol. 2, 2014.

[7] W. Lenz, "Beitrag zum verständnis der magnetischen erscheinungen in festen körpern," *Zeit. fur Physik*, vol. 21, p. 613, 1920.

[8] E. Ising, "Beitrag zur theorie des ferromagnetismus," *Zeit. fur Physik*, vol. 31, p. 253, 1925.

[9] R. Barends, A. Shabani, L. Lamata *et al.*, "Digitized adiabatic quantum computing with a superconducting circuit," *Nature*, vol. 534, no. 7606, pp. 222–226, 2016.

[10] K. Kim, M.-S. Chang, S. Korenblit *et al.*, "Quantum simulation of frustrated ising spins with trapped ions," *Nature*, vol. 465, no. 7298, pp. 590–593, 2010.

[11] Z. Wang, A. Marandi, K. Wen *et al.*, "Coherent ising machine based on degenerate optical parametric oscillators," *Physical Review A*, vol. 88, no. 6, p. 063853, 2013.

[12] Y. Yamamoto, K. Aihara, T. Leleu *et al.*, "Coherent ising machines — optical neural networks operating at the quantum limit," *npj Quantum Information*, vol. 3, no. 1, p. 49, 2017.

[13] F. Böhm, G. Verschaffelt, and G. Van der Sande, "A poor man's coherent ising machine based on opto-electronic feedback systems for solving optimization problems," *Nature communications*, vol. 10, no. 1, p. 3538, 2019.

[14] T. Leleu, Y. Yamamoto, S. Utsunomiya *et al.*, "Combinatorial optimization using dynamical phase transitions in driven-dissipative systems," *Physical Review E*, vol. 95, no. 2, p. 022118, 2017.

[15] T. Inagaki, Y. Haribara, K. Igarashi *et al.*, "A coherent ising machine for 2000-node optimization problems," *Science*, vol. 354, no. 6312, pp. 603–606, 2016.

[16] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014.

[17] Y. Yamamoto, "Optical neural network operating at the quantum limit-coherent ising/xy/recurrent neural network machines," in *2018 Photonics in Switching and Computing (PSC)*. IEEE, 2018, pp. 1–4.

[18] S. Utsunomiya, N. Namekata, K. Takata *et al.*, "Binary phase oscillation of two mutually coupled semiconductor lasers," *Optics express*, vol. 23, no. 5, pp. 6029–6040, 2015.

[19] P. D. Drummond, K. J. McNeil, and D. F. Walls, "Non-equilibrium transitions in sub/second harmonic generation i. semiclassical theory," *Optica Acta: International Journal of Optics*, vol. 27, no. 3, pp. 321–335, 1980.

[20] ——, "Non-equilibrium transitions in sub/second harmonic generation ii. quantum theory," *Optica Acta: International Journal of Optics*, vol. 28, no. 2, pp. 211–225, 1981.

[21] A. Prabhakar, P. Shah, U. Gautham *et al.*, "Optimization with photonic wave-based annealers," *Philosophical Transactions of the Royal Society A*, vol. 381, no. 2241, p. 20210409, 2023.

[22] S. Kiesewetter and P. D. Drummond, "Coherent ising machine with quantum feedback: The total and conditional master equation methods," *Physical Review A*, vol. 106, no. 2, p. 022409, 2022.

[23] A. Marandi, Z. Wang, K. Takata *et al.*, "Network of time-multiplexed optical parametric oscillators as a coherent ising machine," *Nature Photonics*, vol. 8, no. 12, pp. 937–942, 2014.

[24] K. Takata, A. Marandi, R. Hamerly *et al.*, "A 16-bit coherent ising machine for one-dimensional ring and cubic graph problems," *Scientific reports*, vol. 6, no. 1, p. 34089, 2016.

[25] T. Shoji, K. Aihara, and Y. Yamamoto, "Quantum model for coherent ising machines: Stochastic differential equations with replicator dynamics," *Physical Review A*, vol. 96, no. 5, p. 053833, 2017.

[26] A. Yamamura, K. Aihara, and Y. Yamamoto, "Quantum model for coherent ising machines: Discrete-time measurement feedback formulation," *Physical Review A*, vol. 96, no. 5, p. 053834, 2017.

[27] G. Umasankar, P. S. Shah, N. Chandrachoodan *et al.*, "Benchmarking the poor man's ising machine," in *2021 IEEE Photonics Society Summer Topicals Meeting Series (SUM)*. IEEE, 2021, pp. 1–2.

[28] T. A. Maldonado, "Electro-optic modulators," *Handbook of optics*, vol. 2, pp. 13–11, 1995.

[29] L. Lovász, "On the ratio of optimal integral and fractional covers," *Discrete Mathematics*, vol. 13, no. 4, pp. 383–390, 1975. [Online]. Available: https://doi.org/10.1016/0012-365x(75)90058-8

[30] A. K. Singh, K. Jamieson, P. L. McMahon, and D. Venturelli, "Ising machines' dynamics and regularization for near-optimal mimo detection," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 11 080–11 094, 2022.

[31] N. Mohseni, P. L. McMahon, and T. Byrnes, "Ising machines as hardware solvers of combinatorial optimization problems," 2022.

[32] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition," 2020.

[33] L. M. Nguyen, P. H. Nguyen, M. van Dijk *et al.*, "Sgd and hogwild! convergence without the bounded gradients assumption," 2018.

[34] D. P. Bertsekas and J. N. Tsitsiklis, "Gradient convergence in gradient methods with errors," *SIAM J. Optim.*, vol. 10, pp. 627–642, 1999.

[35] A. Nedić and D. Bertsekas, *Convergence Rate of Incremental Subgradient Algorithms*. Boston, MA: Springer US, 2001, pp. 223–264.

[36] F. Böhm, T. V. Vaerenbergh, G. Verschaffelt *et al.*, "Order-of-magnitude differences in computational performance of analog ising machines induced by the choice of nonlinearity," *Communications Physics*, vol. 4, no. 1, jul 2021.

[37] D. P. Bertsekas, "Incremental gradient, subgradient, and proximal methods for convex optimization: A survey," 2017.

[38] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.

[39] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, p. 497, Jul. 1960. [Online]. Available: https://doi.org/10.2307/1910129