Regularized Ising Formulation for Near-Optimal MIMO Detection using Quantum Inspired Solvers

Abhishek Kumar Singh[1,2,3], Kyle Jamieson[1], Peter L. McMahon[4] and Davide Venturelli[2,3]
[1]*Department of Computer Science, Princeton University*
[2]*USRA Research Institute for Advanced Computer Science*
[3]*Quantum AI Laboratory (QuAIL), NASA Ames Research Center*
[4]*School of Applied and Engineering Physics, Cornell University*

***Abstract*—Optimal MIMO detection is one of the most computationally challenging tasks in wireless systems. We show that the quantum-inspired computing approach based on Coherent Ising Machines (CIMs) is a promising candidate for performing near-optimal MIMO detection. We propose a novel regularized Ising formulation for MIMO detection that mitigates a common error floor issue in the direct approach adopted in the existing literature on MIMO detection using Quantum Annealing. We evaluate our methods using a simplified, quantum-inspired model and show that our methods can achieve a near-optimal performance for several Large MIMO systems, like 16×16, 20×20, and 24×24 MIMO with BPSK modulation.**

***Index Terms*—MIMO detection, Large MIMO, Quantum inspired solvers, Coherent Ising machines, Physics-inspired Ising machine-based computation.**

I. Introduction

Wireless technologies have recently undergone tremendous growth in terms of supporting more users and providing higher spectral efficiency, with the next generation of cellular networks planning to support massive machine-to-machine communication [1], large IoT networks [2], and unprecedented data rates [3]. The number of mobile users and data usage is rapidly increasing [4], and while data traffic has been predominantly downlink, the volume of uplink traffic is becoming ever higher [5] due to the emergence of interactive services and applications. The problem of optimal and efficient wireless signal detection in a multiple-input, multiple-output (MIMO) system is central to this rapid growth and has been a key interest of network designers for several decades. While the optimal Maximum Likelihood (ML) MIMO detector is well known, it attempts to solve an NP-Hard problem [6] exactly, and so its implementation is usually impractical and infeasible for real-world systems. These computational challenges have prompted network designers to seek optimized implementations such as the Sphere Decoder [7], or sub-optimal approximations with polynomial complexity like Minimum Mean Square Error decoder (MMSE) [8], successive interference cancellation (SIC) [9], Lattice-reduction based algorithms [10], and the Fixed Complexity Sphere Decoder [11]. However, even today, practical methods that achieve near-optimal performance for large MIMO systems are lacking [6].

In the Computer Architecture and Physics communities, the last decade has seen a rise of a novel class of analog computers that use the dynamics of a physical system to heuristically find solutions to optimization problems that are framed as instances of the *Ising model*, one of the most studied frameworks for magnetism in statistical mechanics. These solvers include Quantum Annealing [12], Coherent Ising Machines [13] (optical systems involving quantum parallel search and quantum filtering), and several classical solvers inspired by quantum systems [14], [15]. These already show promise as practical computational structures for addressing some NP-hard problems arising in practical applications but recent starting work [16], [17] on the MIMO detection with Quantum Annealing, leveraging a straightforward mapping of ML-MIMO decoding problem to the Ising model, experiences an error floor in the bit error rate (BER) versus the signal-to-noise ratio (SNR) characteristics, *i.e.*, the BER does not reduce when the SNR increases. In this paper, we observe that this error floor is present in the regime of practical deployment for MIMO detection. More specifically, in the regime relevant for real systems (uncoded BER of $10^{-3} - 10^{-6}$), even if we dismiss the limitations of non-idealized physics-based Ising solvers, depending on the SNR, there are many interesting scenarios in which they would not serve as good MIMO detectors in practical systems if the known Ising formulation of the ML-MIMO problem is used.

Hence we propose a novel regularized Ising formulation of the ML-MIMO problem, using a low-complexity approximation (see Fig. 1 for an overview of our approach). This new formulation leads us to propose Regularised Ising MIMO (RI-MIMO) algorithm, that can in principle, provide near-optimal MIMO detection on Ising machines. The rest of the paper is organized as follows. Section II provides a survey of the existing state-of-the-art to solve ML-MIMO. Section III describes the MIMO system model and the reduction of the ML-MIMO problem to an Ising optimization problem. Section IV is a primer on Coherent Ising Machines. Section V describes our novel Ising formulation and the proposed algorithm. Section VI contains the evaluation of BER of our method in various scenarios using a simplified Coherent Ising Machine model. We show that our techniques mitigate the error floor problem and achieve near-optimal performance for 16×16, 20×20, and 24×24 MIMO with BPSK modulation. We perform extensive empirical experimentation for parameter tuning. Finally, we conclude in Section VII and discuss the

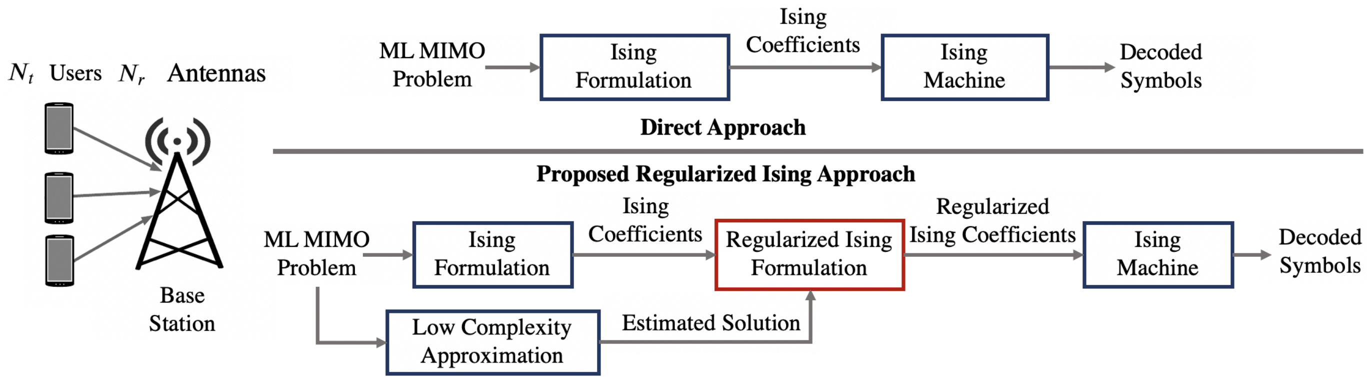

Fig. 1: Uplink Maximum Likelihood MIMO detection (ML-MIMO) using Ising Machines, illustrating the differences between the proposed regularised Ising approach and the direct application of the Ising formulation.

possible future directions for our work in Section VIII.

In this paper, we present the fundamental design ideas and evaluation for the proposed method, and further analysis along with several significant advancements can be found in the extended version of this work [18]: which proposes a new tree search algorithm (TRIM), extends the evaluation to include higher modulations, massive MIMO, finite precision limitations, and spectral efficiency evaluation. It also provides a detailed discussion on taking our work from simulation to an actual hardware implementation of the CIM that can meet the processing constraints of a typical LTE system.

II. Related Work

The Maximum Likelihood MIMO detection (ML-MIMO) has been a key problem of interest for wireless systems for several decades. The Sphere Decoder [7] (SD) is a reference algorithm to solve this problem, which performs an optimized, pruned tree search. Its average computational complexity is still exponential [19], so its deployment is practically infeasible for MIMO systems with a large number of users due to the strict timing requirements of state-of-the-art wireless systems like 5G NR. The Fixed Complexity Sphere Decoder [11] (FSD) is a polynomial-time approximation to the sphere decoder that aggressively prunes the search tree. In practice, wireless network designers resort to simpler methods like linear detectors (MMSE), which perform channel inversion, or successive interference cancellation (SIC) based techniques [9] that focus on decoding each user sequentially while canceling inter-user interference. Given the practical importance of MMSE and SIC, many techniques have been put forward to advance their performance, including Lattice Reduction (LR), which involves pre-processing the channel to produce a reduced lattice basis [10]. In [20] authors explore Gibbs Sampling for MIMO detection. In [21], the authors use the L2 norm of the solution to regularize the fixed complexity sphere decoder to deal with a rank deficient channel in an OFDM/SDMA uplink, and in [22], the authors use the MMSE estimate to determine the search radius during sphere decoding. In [23], authors explore L1 and L2 regularisation to improve the performance of lattice sphere decoding. In [24], authors propose a dead-zone penalty and infinity-norm-based regularisation to improve the performance of the MMSE detector. Regularised Lattice Decoding, like the MMSE-regularised lattice decoding, penalizes deviations from origin to mitigate the out-of-bound symbol events in lattice reduction-based MIMO detection [25]. In [25], authors propose a Lagrangian Dual relaxation for ML MIMO detection and generalize the regularised lattice decoding techniques. Regularisation techniques are also widely used in Machine Learning to prevent over-fitting [26]. Many of these works can achieve near-optimal performance for small systems. However, as the number of users and number of antennas at the base station increase, they require an exponential increase in computation time (to maintain near-optimal behavior), or their performance becomes progressively worse.

The application of Quantum Annealing (QA) and Ising machines to MIMO detection is starting to be investigated in the last couple of years [27], [28] and has shown promising results. The QuAMax MIMO detector [16] leverages quantum annealing for MIMO detection. A classical-quantum hybrid approach to QA-based ML-MIMO was proposed in [29]. Quantum Annealing has shown promising results for other tough computational problems in wireless systems like Vector Perturbation Precoding [30]. In [31], authors explore the use of Oscillator-Ising machine for MIMO detection in Massive MIMO systems; however, they target 16×64 MIMO with QPSK modulation, which is not a very challenging problem and even MMSE detector is near optimal. In [17], authors explore the use of Parallel Tempering for Ising-based MIMO detection (ParaMax), improving the performance of QuAMax; however, both QuAMax and ParaMax suffer from the aforementioned bit error floor, which is addressed in our work.

III. MIMO Systems, Ising problems and Maximum Likelihood Detection

In this section, we will describe the MIMO system model, the MIMO Maximum Likelihood Detection (ML-MIMO) problem, and the transformation between the ML-MIMO problem and its equivalent Ising problem.

Consider the UL transmission in a MIMO system with N_r antennas at the base station (BS) and N_t users, each with a single antenna. $\mathbf{x_o} = \{x_1, x_2, ...x_{N_t}\}^T$ is the transmit vector where x_i is the symbol transmitted by user i. $\mathbf{y} = \{y_1, y_2, ...y_{N_r}\}^T$ is the received vector where y_j is the signal received by antenna j. Each x_i is a complex number drawn from a fixed constellation Ω. The channel between user j and receive antenna i is expressed as a complex number h_{ij} that represents the channel's attenuation and phase shift of the transmitted signal x_j. Let $\mathbf{H}$ denote the complex-valued channel matrix,

$$\mathbf{y} = \mathbf{H}\mathbf{x_o} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ denotes Additive White Gaussian Noise (AWGN). With AWGN, the optimal receiver is the Maximum Likelihood receiver [7] which is given by

$$\hat{\mathbf{x}}_{\mathbf{ML}} = \arg \min_{\mathbf{x} \in \Omega^{N_t}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \tag{2}$$

An Ising optimization problem [32] is quadratic unconstrained optimization problem over N spin variables:

$$\begin{aligned} &\arg \min_{s_1, s_2, ...s_N} -\sum_{i=1}^{N} h_i s_i - \sum_{i \neq j} J_{ij} s_i s_j \\ &= \arg \min_{\mathbf{s} \in \{-1,1\}^N} -\mathbf{h}^T \mathbf{s} - \mathbf{s}^T \mathbf{J} \mathbf{s}, \end{aligned} \tag{3}$$

where each spin variable $s_i \in \{-1, 1\}$, or in its vector form (RHS) $\mathbf{s} = \{s_1, s_2, ...s_N\}$, where all diagonal entries of the matrix $\mathbf{J}$ are zeros.

The minimization problem expressed in (2) can be equivalently converted into Ising form by expressing $\mathbf{x}$ using spin variables. The first step is derive a real valued equivalent of (1), which is obtained by the following transformation [6],

$$\tilde{\mathbf{H}} = \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}, \tag{4}$$

$$\tilde{\mathbf{y}} = \begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix}, \ \tilde{\mathbf{x}} = \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}, \tag{5}$$

where $\Re(.)$ and $\Im(.)$ represent the real and imaginary part.

The ML receiver described in (2) has the same expression under the transformation and the optimization variable $\tilde{\mathbf{x}}$ is real valued. Let us say $\mathbf{x}$ has $\frac{N}{2}$ elements drawn from a square M-QAM constellation, then each element of the optimization variable $\tilde{\mathbf{x}}$ takes integral values in the range $\Omega_r = \{-\sqrt{M}+1, -\sqrt{M}+3, ...\sqrt{M}-1\}$. The number of bits needed to express Ω_r are given by $r_b = \lceil \log_2(\sqrt{M}) \rceil$. let $\mathbf{s}$ be an $N * r_b \times 1$ spin vector such than each element of $\mathbf{s}$ can take values $\{-1, 1\}$. Then, element j of $\tilde{\mathbf{x}}$ can be represented using r_b spin variables $\{s_j, s_{j+N} ... s_{j+(r_b-1)N}\}$,

$$\tilde{x}_j = \sum_{i=1}^{r_b} 2^{r_b - i}(s_{j+(i-1)N} + 1) - (\sqrt{M} - 1) \tag{6}$$

We define the transform matrix

$$\mathbf{T} = \begin{bmatrix} 2^{r_b-1}\mathbb{I}_N & 2^{r_b-2}\mathbb{I}_N \ ... \ \mathbb{I}_N \end{bmatrix}, \tag{7}$$

then $\tilde{\mathbf{x}}$ can be expressed as,

$$\tilde{\mathbf{x}} = \mathbf{T}(\mathbf{s} + \bar{\mathbb{1}}_{N*r_b}) - (\sqrt{M} - 1)\bar{\mathbb{1}}_N \tag{8}$$

For BPSK and rectangular QAM constellations, the $\Re(\mathbf{x})$ and $\Im(\mathbf{x})$ in (5) have different range and (6) can be accordingly modified to construct the transform matrix T. We substitute (8) in the real valued maximum likelihood problem and simplify to obtain the ising formulation for ML receiver. Let $\mathbf{z} = \tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{T}\bar{\mathbb{1}}_{N*r_b} + (\sqrt{M} - 1)\tilde{\mathbf{H}}\bar{\mathbb{1}}_N$, then the Ising problem for ML-MIMO receiver is described by,

$$\mathbf{h} = 2 * \mathbf{z}^T \tilde{\mathbf{H}}\mathbf{T}, \quad \mathbf{J} = -\mathit{zeroDiag}(\mathbf{T}^T \tilde{\mathbf{H}}^T \tilde{\mathbf{H}}\mathbf{T}), \tag{9}$$

where *zeroDiag*$(\mathbf{W})$ sets the diagonal elements of matrix W to zero. We further scale the problem such that all the coefficients lie in $[-1, 1]$. The Ising solution can be converted to real valued ML solution using (8), which can be then converted to the complex valued solution for the original problem described in (2) by inverting the transform in (5).

IV. Coherent Ising Machines (CIM)

In simple terms, an Ising Machine can be described as a module that takes an Ising Problem (Eq. 3) as input and outputs a candidate solution, according to an unknown probability distribution that depends on a few parameters. We refer to a single, independent run on the Ising Machine as an "anneal", borrowing nomenclature from the simulated/quantum-annealing methods. After each run, the machine is reset. A common approach is to run several samples of a single Ising problem instance and then return the best-found solution (the "ground state") in the sample.

Coherent Ising Machines (CIMs), as originally conceived [14], implement the search for the ground state of an Ising problem by using an optical artificial spin network. Their baseline architecture encodes the Ising spins into a train of time-resolved, phase-coherent laser pulses traveling on an optical fiber loop, undergoing controlled interference between all pairs of wavepackets. The phase dynamics of the pulses is governed by the presence of a non-linear element in the form of a degenerate-optical-parametric-oscillator (DOPO). For the purpose of our work, this version of the CIM can be modeled using a system of stochastic differential equations [33]. In this work, we will use a model of the CIM that describes its continuous-time limit evolution neglecting quantum effects, which has been proven to be fitting the experiments of multiple devices and represents a baseline setup for more sophisticated embodiments. Following Ref. [13], the in-phase (c_i) and quadrature (q_i) components of each signal-variable (that describes the optical pulses) can be modeled using the following differential equations:

$$\begin{aligned} dc_i &= [(-1 + p - c_i^2 - q_i^2)c_i + C\sum_j J_{ij}c_j]dt \\ &+ \frac{1}{A_s}\sqrt{c_i^2 + q_i^2 + \frac{1}{2}}dW_1 \\ dq_i &= (-1 - p - c_i^2 - q_i^2)q_i dt + \frac{1}{A_s}\sqrt{c_i^2 + q_i^2 + \frac{1}{2}}dW_2 \end{aligned} \tag{10}$$

where the normalized pump rate (p) are CIM parameters that relates to the laser used in the machine and can be tuned easily. The constant C is typically fixed by design considerations (mostly by the power transmission coefficient and the laser saturation amplitude). J_{ij} is the Ising coupling coefficient from the j^{th} pulse to the i^{th} pulse, which is programmable. The stochasticity is introduced through dW_1 and dW_2, which are independent Gaussian-noise processes. The variable t is time (normalized with respect to the photon decay rate). An Ising problem with the spin-spin-coupling matrix $\mathbf{J}$ is encoded in the CIM by setting the optical couplings $\tilde{\zeta}_{ij} \propto J_{ij}$. The anneal consists of pumping energy into the system by gradually varying p. Heuristically this is implemented in a schedule at a speed that is some monotonic function of N. The solution to the Ising problem is read out at the end of the anneal by measuring the in-phase component of each DOPO c_i, and interpreting the sign of each as a spin value s_i, i.e. $s_i = \text{sign}(c_i)$. To enable the study of how an ideal CIM would perform on solving Ising instances related to the application at hand (MIMO detection), we implemented a software simulator of a CIM that integrates the differential equations described in (11), using double precision in MATLAB.

V. Design

In this section, we propose the RI-MIMO detector, based on our novel regularised Ising formulation of maximum-likelihood MIMO receiver, which mitigates the error floor problem and uses a single auxiliary spin variable to transform the Ising problem into a form compatible with CIMs.

A. RI-MIMO: Regularized Ising-MIMO

The key idea is to add a regularisation term based on a low complexity estimate of the solution, which, as we will see later, will improve the BER performance. The maximum likelihood MIMO receiver is given by (2). Let us say that we have a polynomial-time estimate (obtained by algorithms like MMSE or ZF) $\mathbf{x}_P$. Let $\mathbf{s}_P$ be the spin vector corresponding to $\mathbf{x}_P$ obtained from (8). We add to the Ising form a penalty term for deviations from the poly-time estimate, which would penalize non-optimal solutions in low noise scenarios, to obtain the following:

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \min_{\mathbf{s}\in\{-1,1\}^N} -\mathbf{h}^T\mathbf{s} - \mathbf{s}^T\mathbf{J}\mathbf{s} + r(\rho, M, N_t)\|\mathbf{s}-\mathbf{s}_P\|^2 \\ &= \arg \min_{\mathbf{s}\in\{-1,1\}^N} -(\mathbf{h} + 2r(\rho, M, N_t)\mathbf{s}_P)^T\mathbf{s} - \mathbf{s}^T\mathbf{J}\mathbf{s}, \end{aligned} \quad (11)$$

where $r(\rho, M, N_t)$ is a regularization parameter dependent of the SNR, modulation and number of users. This style of regularisation falls in the class of generalized Tikhonov regularisation. We will look at the choice of $r(\rho, M, N_t)$ in Section VI-B. The *RI-MIMO-N_a* algorithm is as follows:

- Convert the ML-MIMO detection problem into the Ising form as described in Section III.
- Add the regularisation term as described by (11).
- Perform N_a anneals using an Ising machine.
- Select the best solution from the candidate solutions generated by the Ising machine and the MMSE solution.

B. Solution of an Ising problem having access only to programmable quadratic couplings

Not all CIMs are designed to solve Ising problems containing a bias term ($\mathbf{h}^T\mathbf{s}$ in (3)). In order to solve a general Ising problem using a CIM that does not natively support bias terms (although some, such as the implementation used in [14], do), we introduce an auxiliary spin variable s_a and solve the following Ising problem:

$$\arg \min_{s_a, s_1, s_2, \ldots s_N} -\sum_{i=1}^{N} h_i s_i s_a - \sum_{i \neq j} J_{ij} s_i s_j, \quad (12)$$

which contains no bias terms and can be solved using an Ising machine that doesn't support bias terms. (12) has two degenerate solutions: $[\{\hat{s}_i\}_{i=1}^N, \hat{s}_a = 1]$ and $[\{-\hat{s}_i\}_{i=1}^N, \hat{s}_a = -1]$. Note that $\{\hat{s}_i\}_{i=1}^N$ is the solution for the original Ising problem in (3). Hence we can obtain the solution to the original Ising problem from the solutions of the auxiliary Ising problem.

VI. Evaluation

In this section, we will perform an extensive evaluation of RI-MIMO in various scenarios. We will simulate uplink wireless MIMO transmission between N_t users with one transmit antenna each and a base station with N_r receive antenna. We assume Rayleigh fading channel between them and additive white Gaussian noise (AWGN) at the receiver. We further assume, for simplicity, that the channel is known at the receiver and all users use the same modulation scheme. The BER is calculated as the mean BER of the N_t independent data streams transmitted by N_t users.

A. BER Performance

We start comparing the optimal decoder (the Sphere Decoder) and the linear MMSE decoder against *RI-MIMO* and the unregularized *ML-MIMO* using as a test case BPSK 16×16. This case will represent a baseline for our benchmarks and their sophistication. Note that a trivial way to remove the error floor is to take the better solution out of those generated by MMSE and *CIM-ML-M_a*. We see from Fig 2 that *RI-MIMO* provides much better BER than *CIM-ML*, mitigating the error floor problem associated with it. We note that if we run concurrently MMSE and *ML-MIMO* for each instance and we select the best of both results (*CIM-ML+MMSE*), we are still less performant than *RI-MIMO*. We can see that *RI-MIMO*-64 achieves near-optimal performance for 16×16 and 20×20 MIMO BPSK modulation, achieving a BER of 10^{-4} at slightly ($\leq$ 2 dB) higher SNR. For 24×24 MIMO, we see that the performance gap between RI-*RI-MIMO*-64 and Sphere Decoder increases, and a higher number of anneals are required to bridge the gap. Note that increasing the number of anneals (N_a) will improve the performance further; however, it comes at the cost of increased implementation/computational complexity. We try to strike a balance between these two aspects with our choice of $N_a = 64$. A detailed analysis of the appropriate choice of N_a and implementation aspects (to meet the LTE requirements) of our algorithms can be found in the extended version of this work [18].

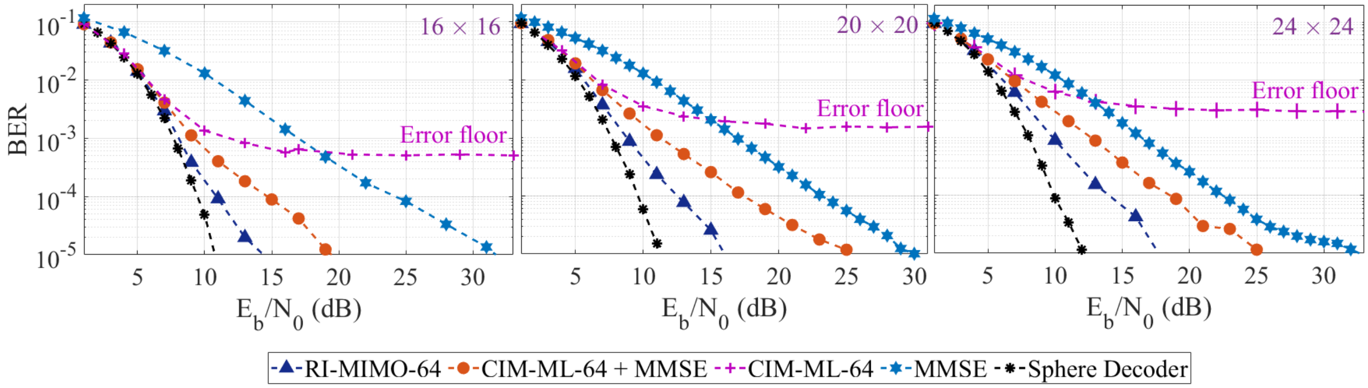

Fig. 2: Bit Error Rate (BER) Curves for (Left) 16×16, (Center) 20×20, (Right) 24×24 MIMO and BPSK modulation, illustrating the error floor problem and performance of all the tested solvers. The curves are computed over $\approx 25 \times 10^3$ MIMO instances (128 channel instances, 198 transmit vectors per channel).

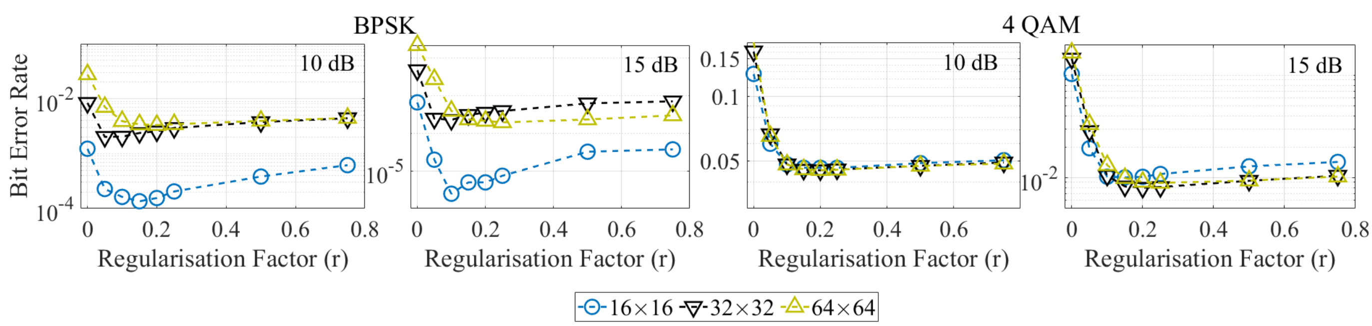

Fig. 3: Bit Error Rate at 10 dB and 15 dB SNR , illustrating the performance of RI-MIMO on Coherent based Ising Machines (CIM) for various value of regularisation factor with different MIMO sizes and modulation (using 64 anneals per instance).

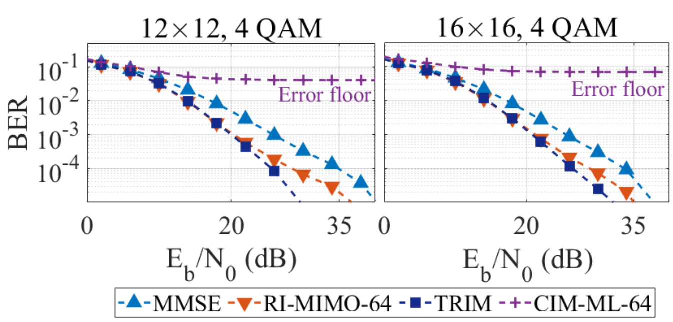

Fig. 4: Bit Error Rate Curves for 4-QAM modulation, illustrating the performance of RI-MIMO. The curves are computed over $\approx 25 \times 10^3$ MIMO instances (128 channel instances, 198 transmit vectors per channel).

B. Optimal regularisation factor for RI-MIMO and Higher Order Modulations

In this section, we will discuss the tuning of the value for the regularisation prefactor for *RI-MIMO* in (11), $r(\rho, M, N_t)$, relative to the magnitude of Ising coefficients of the original un-regularised problem. To maintain consistency of results, we normalize the Ising coefficients of the original problem to $[-1, 1]$. Starting from the 16×16 BPSK baseline MIMO system, we compute performance for various values.

In order to determine the impact of modulation(M), number of users (N_t) and SNR (ρ) on the optimal value of $r(\rho, M, N_t)$, we look at BER vs regularisation factor for various MIMO sizes and modulations, while keeping SNR fixed at 10 dB and 15 dB in Fig. 3. We note that the BER reduces dramatically from $r = 0$ (unregularized) to around $r = 0.1$, beyond which the sensitivity of BER to choice of r is not much. We note that the optimal value is around 0.15, which acts as a threshold: for larger r the BER performance is only slightly affected. Based on these observations, for practicality, in our benchmarks, we will be using $r(\rho, M, N_t) = 0.15$, irrespective of SNR, modulation, and the number of users. In a practical system, similar experiments can be used to construct a lookup table for the optimal value of r as a function of M, N_t and ρ.

Using the prescriptions above, Fig. 4 provides the BER performance of *RI-MIMO* for a 12×12 and 16×16 MIMO system with 4 QAM modulation. We see that the error floor problem appears even for higher modulation, and *RI-MIMO* successfully mitigates it. Note that the difference between *RI-MIMO* and MMSE reduces as the modulation order increases. More details on the performance of our methods for higher modulations, and further enhancements (to improve the BER for higher modulations) are available in the extended version of this work [18].

VII. Conclusion

In this paper, we explore the application of Coherent Ising Machines (CIM) for maximum likelihood detection for MIMO detection. We see that previous approaches used by MIMO detectors based on the Ising model suffer from an error floor problem and, unless many repetitions are allowed, does not have a satisfactory Bit Error Rate (BER) performance in practice. We propose a novel Regularized Ising approach and show that it mitigates the error floor problem. We demonstrate, using a CIM simulator, that our algorithm can outperform the previous Ising approach and have the potential to achieve near-optimal performance for large MIMO systems. By means of an extensive numerical evaluation, we see that the Regularized Ising approach has an impressive error performance when compared to the state-of-art. In conclusion, our results indicate that Coherent Ising Machines (using the proposed Regularized Ising approach) are a promising candidate for providing a superior alternative to the existing MIMO detection approach and achieving near-optimal performance for practical systems with a large number of users and antennas.

VIII. Future Work

Our evaluation is based on simulating a simplified CIM model; as a next step, we plan to evaluate our methods on more recent extensions to the CIM (e.g., variants incorporating amplitude-heterogeneity correction [34], [35]) and on an experimental CIM implementation [14], [36].

Acknowledgments

D.V. acknowledges support from NSF award CNS-1824470 and both D.V. and P.L.M. acknowledge support from NSF award CCF-1918549. A.K.S. and K.J. acknowledge support from NSF award CNS-1824357. A.K.S.'s internship has also been supported by the USRA Feynman Quantum Academy, as part of the NASA Academic Mission Services (NNA16BD14C) – funded under SAA2-403506. P.L.M. also acknowledges financial and technical support from NTT Research and membership in the CIFAR Quantum Information Science Program as an Azrieli Global Scholar.

References

[1] C. Kalalas and J. Alonso-Zarate, "Massive Connectivity in 5G and Beyond: Technical Enablers for the Energy and Automotive Verticals," *2020 2nd 6G Wireless Summit (6G SUMMIT)*.

[2] S. Narayanan, D. Tsolkas *et al.*, "NB-IoT: A Candidate Technology for Massive IoT in the 5G Era," in *23rd IEEE CAMAD*, 2018, pp. 1–6.

[3] Yinan Qi, M. Hunukumbure *et al.*, "Quantifying data rate and bandwidth requirements for immersive 5G experience," in *IEEE ICC 2016*.

[4] "IMT Traffic estimates for the years 2020 to 2030," *ITU-R Report M.2370-0, July 2015.*

[5] J. Oueis and E. C. Strinati, "Uplink traffic in future mobile networks: Pulling the alarm," in *Cognitive Radio Oriented Wireless Networks*. Springer International Publishing, 2016, pp. 583–593.

[6] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs," *IEEE Communications Surveys Tutorials*, 2015.

[7] C. Hung and T. Sang, "A Sphere Decoding Algorithm for MIMO Channels," in *IEEE ISSPIT*, 2006, pp. 502–506.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[9] C. Navarro i Manchon, L. Deneire *et al.*, "On the Design of a MIMO-SIC Receiver for LTE Downlink," in *2008 IEEE 68th VTC*.

[10] B. Gestner, W. Zhang, X. Ma, and D. V. Anderson, "Lattice Reduction for MIMO Detection: From Theoretical Analysis to Hardware Realization," *IEEE Transactions on Circuits and Systems I: Regular Papers*.

[11] L. G. Barbero and J. S. Thompson, "Fixing the Complexity of the Sphere Decoder for MIMO Detection," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2131–2142, 2008.

[12] D. de Falco and D. Tamascelli, "An introduction to quantum annealing," *RAIRO - Theoretical Informatics and Applications*, vol. 45, 07 2011.

[13] Y. Haribara, S. Utsunomiya, and Y. Yamamoto, "Computational Principle and Performance Evaluation of Coherent Ising Machine Based on Degenerate Optical Parametric Oscillator Network," *Entropy*, 2016.

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

[15] 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, 2019.

[16] M. Kim, D. Venturelli, and K. Jamieson, "Leveraging Quantum Annealing for Large MIMO Processing in Centralized Radio Access Networks," *31st ACM SIGCOMM*, 2019.

[17] M. Kim, S. Mandra *et al.*, "Physics-Inspired Heuristics for Soft MIMO Detection in 5G New Radio and Beyond," in *27th ACM MobiCom*, 2021.

[18] 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*, pp. 1–1, 2022.

[19] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Transactions on Signal Processing*, 2005.

[20] A. M. Mussi, "Multiple restarts mixed gibbs sampling detector for large-scale antenna systems," *IET Signal Processing*, May 2019.

[21] S. V. and A. K., "A signal power adaptive regularized sphere decoding algorithm based multiuser detection technique for underdetermined OFDM/SDMA uplink system," *International Journal of Communication Systems*, vol. 32, 11 2018.

[22] I. Kanaras *et al.*, "A combined MMSE-ML detection for a spectrally efficient non orthogonal FDM signal," in *2008 5th IEEE International Conference on Broadband Communications, Networks and Systems*.

[23] M. Albreem and M. F. Mohd Salleh, "Regularized Lattice Sphere Decoding for Block Data Transmission Systems," *Wireless Personal Communications*, 01 2015.

[24] M. Alghoniemy, "Regularized MIMO decoders," *Journal of Communications Software and Systems*, vol. 5, pp. 149–153, 12 2010.

[25] J. Pan, W.-K. Ma, and J. Jaldén, "MIMO Detection by Lagrangian Dual Maximum-Likelihood Relaxation: Reinterpreting Regularized Lattice Decoding," *IEEE Transactions on Signal Processing*, 2014.

[26] J. Kukačka, V. Golkov, and D. Cremers, "Regularization for deep learning: A taxonomy," *CoRR*, vol. abs/1710.10686, 2017.

[27] M. Kim, S. Kasi *et al.*, "Heuristic quantum optimization for 6g wireless communications," *IEEE Network*, vol. 35, no. 4, pp. 8–15, 2021.

[28] J. C. De Luna Ducoing and K. Nikitopoulos, "Quantum Annealing for Next-Generation MU-MIMO Detection: Evaluation and Challenges."

[29] M. Kim, D. Venturelli, and K. Jamieson, "Towards Hybrid Classical-Quantum Computation Structures in Wirelessly-Networked Systems," *Nineteenth ACM Workshop on Hot Topics in Networks*, 2020.

[30] S. Kasi, A. K. Singh *et al.*, "Quantum Annealing for Large MIMO Downlink Vector Perturbation Precoding," in *IEEE ICC*, 2021.

[31] J. Roychowdhury, J. Wabnig, K. P. Srinath *et al.*, "Performance of Oscillator Ising Machines on Realistic MU-MIMO Decoding Problems, 22 September 2021, PREPRINT (Version 1)."

[32] Z. Bian, F. A. Chudak, W. Macready, and G. Rose, "The Ising model : teaching an old problem new tricks," 2010. [Online]. Available: https://www.dwavesys.com/media/vbklsvbh/weightedmaxsat_v2.pdf

[33] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, "Coherent Ising machine based on degenerate optical parametric oscillators," *Physical Review A*, vol. 88, no. 6, p. 063853, 2013.

[34] T. Leleu, Y. Yamamoto, P. L. McMahon, and K. Aihara, "Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity," *Physical review letters*, vol. 122, no. 4, p. 040607, 2019.

[35] T. Leleu, F. Khoyratee *et al.*, "Chaotic Amplitude Control for Neuromorphic Ising Machine in Silico," *arXiv preprint arXiv:2009.04084*, 2020.

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