

Exploring Embeddings for MIMO Channel Decoding on Quantum Annealers

Ádám Marosits^{1,2}, Zsolt Tabi^{1,3}, Zsófia Kallus¹, Péter Vaderna¹, István Gódor¹, and Zoltán Zimborás^{4,2}

Abstract—Quantum Annealing provides a heuristic method leveraging quantum mechanics for solving Quadratic Unconstrained Binary Optimization problems. Existing Quantum Annealing processing units are readily available via cloud platform access for a wide range of use cases. In particular, a novel device, the D-Wave Advantage has been recently released. In this paper, we study the applicability of Maximum Likelihood (ML) Channel Decoder problems for MIMO scenarios in centralized RAN. The main challenge for exact optimization of ML decoders with ever-increasing demand for higher data rates is the exponential increase of the solution space with problem sizes. Since current 5G solutions can only use approximate methodologies, Kim et al. [1] leveraged Quantum Annealing for large MIMO problems with phase shift keying and quadrature amplitude modulation scenarios. Here, we extend upon their work and present embedding limits for both more complex modulation and higher receiver / transmitter numbers using the Pegasus P16 topology of the D-Wave Advantage system.

Index Terms—Quantum Computing, Quantum Annealing, NP-hard optimization, Graph embedding, Telecommunication, Massive-MIMO

I. INTRODUCTION

QUANTUM Computers use the unique information processing possibilities offered by quantum mechanics to solve complex problems [2], [3]. At the current level of technological maturity, universal large-scale Quantum Computers are still many years away. However, today's Noisy Intermediate-Scale Quantum (NISQ) devices already offer experimental platforms, and Quantum Annealers [4]–[6] play a prominent role, as they enable running optimization algorithms with few hundreds or even few thousands of qubits, although with considerable noise present in the system. In this paper, we study the embedding problem for topologies of state-of-the-art Quantum Annealers for the telecommunication problem of decoding wireless physical channel transmission by Large and Massive multiple input multiple output (MIMO) [7] antenna arrays.

Due to the ever-increasing demand for higher data rates, capacity and throughput, the application of MIMO antenna

arrays is indispensable to support multiple users near a wireless access point or base station at the same time [8]. As the number of antennas in MIMO setups increases, the complexity of encoding and decoding of signals requires an increasing computing power. [9]

The maximum likelihood (ML) MIMO decoding technique is – in theory – capable of high throughput, but is rarely used in practice as it requires exponential computing complexity in the number of antennas [10]. Kim et al. [1] examined the idea of quantum computation leveraged within the data center of a centralized radio access network (C-RAN) [11] in the hope of speeding up the computing and maintaining throughput by solving the decoding problem there. In such a solution, the ML decoding problem can be first formulated as a Quadratic Unconstrained Binary Optimization (QUBO) problem [12], [13], suitable for a QA processing unit. In the end, the results are mapped back to decoded bits.

Our goal was to extend upon the methodology in [1] and adapt it to set of modulation schemes most relevant for advanced telecommunication scenarios. The paper presents our extension of the known decoding problem to 64-QAM modulation by the maximum likelihood detection. Furthermore, we present a comparative analysis focusing on embedding efficiency using the topologies of both the D-Wave 2000Q Quantum Processing Unit (QPU) of 2000 qubits and the recently released D-Wave Advantage platform based on a 5000-qubit QPU [14].

In the next section, we give a brief overview of Quantum Annealing and the MIMO ML decoding and its formulation as a QUBO/Ising problem. In Sec. IV, we present our extension of this problem formulation for a higher-order modulation. Sec. V-A gives a description of the available D-Wave topologies and embedding methods, while in Sec. V-C the Theoretical limits of the mapping is explained and the largest embedded MIMO scenarios are presented. Finally, in Sec. VI we summarize our results and discuss further possibilities.

II. THEORETICAL BACKGROUND

A. Ising and QUBO models

In order to use the D-Wave Quantum Annealer, one needs to state the optimization problem as a standard QUBO or its equivalent Ising model. The Ising model describes physical systems of discrete spin variables, where each variable is either -1 or 1 , i.e., the configuration space of dimension 2^N is:

$$\Omega_N := \{-1, +1\}^{\times N} = \{(s_1, \dots, s_N) : s_k = \pm 1\}. \quad (1)$$

¹Ericsson Research, Budapest, Hungary

²Budapest University of Technology and Economics, Budapest, Hungary

³Eötvös Loránd University, Budapest, Hungary

⁴Quantum Computing and Information Group, Wigner Research Centre for Physics, Budapest, Hungary

E-mail: {adam.marosits, zsolt.tabi, zsofia.kallus, peter.vaderna, istvan.godor}@ericsson.com; zimboras.zoltan@wigner.hu

The Ising spin glass model gives the energy function or Hamiltonian of a given spin configuration state $\mathbf{s} \in \Omega_N$ of the system as follows:

$$H(\mathbf{s}) = -\frac{1}{2} \sum_{i,j=1}^N J_{ij} s_i s_j - \sum_{i=1}^N h_i s_i, \quad (2)$$

where h_i is the i th qubit's interaction with the external field (bias), and J_{ij} is the strength of the interaction between qubits i and j (coupling strength). If the system prefers the pair of spins to be aligned ($s_i = s_j$) the interaction is called ferromagnetic coupling and, if the pair of spins to be anti-aligned ($s_i = -s_j$) the interaction is called antiferromagnetic. To follow the notation of [1], we shall rewrite the optimization from Eq. 2 to the following form:

$$\begin{aligned} \hat{\mathbf{s}} &= \hat{s}_1, \dots, \hat{s}_N = \\ &= \arg \min_{\mathbf{s} \in \{(s_1, \dots, s_N)\}} \left(\frac{1}{2} \sum_{i,j=1}^N g_{ij} s_i s_j + \sum_{i=1}^N f_i s_i \right), \end{aligned} \quad (3)$$

where $s_i \in \{-1, 1\}$ are the spin variables and f_i , g_{ij} contain the Ising model's coefficients corresponding to the biases and coupling strengths, respectively and $\hat{\mathbf{s}}$ is a minimum energy Ising spin configuration vector.

Since the QUBO model has equivalent expressing power to the Ising model, we can easily convert back-and-forth between the two. The QUBO description of an optimization problem is stated as:

$$\hat{\mathbf{q}} = \hat{q}_1, \dots, \hat{q}_N = \arg \min_{\mathbf{q} \in \{(q_1, \dots, q_N)\}} \frac{1}{2} \sum_{i,j=1}^N Q_{ij} q_i q_j, \quad (4)$$

where q_i are binary decision variables, Q is an symmetric matrix of coefficients and $\hat{\mathbf{q}}$ is the resulting bit string of the optimization.

Since q_i is binary, it has the property: $q_i^2 = q_i$, which can be very useful. With $q_i = (s_i + 1)/2$, one can convert between the two models effortlessly.

Finding the global minimum of a given Hamiltonian is an NP-hard task, i.e., for large problems it will take exponentially long time in the size of the problem to compute the exact solution on a classical computer.

Therefore, one often employs heuristic algorithms, such as Simulated Annealing (SA) [15] to produce approximate results for large problems in polynomial time.

B. Quantum Annealing

QA is a heuristic method for finding a global minimum of an objective function using quantum mechanical evolution. QA is similar to Simulated Annealing in a sense that it randomly searches through the energy landscape of the optimization problem. However, unlike SA, QA does not use a temperature parameter to traverse the energy landscape, instead it slowly tunes the parameters of an Ising model Hamiltonian with transverse field that governs the quantum mechanical evolution of the system. The system starts from a superposition of all possible computational basis states with equal amplitudes,

all possible computational basis states with equal amplitudes, which is the ground state of the purely transverse field Hamiltonian. During the time-evolution the system approximately continues to be in the lowest energy state of the transverse-field Ising model with coupling strengths varying in time. As the coupling strength of the transverse field is approaching zero, the system evolves into the ground state of the original problem Ising Hamiltonian [16].

D-Wave's Quantum Annealer is a superconducting QPU that realizes the Ising spin system in a transverse field. Its qubits and couplers are individually controllable via digital-to-analog converters and have time-dependent control in order to implement the transverse-field Ising Hamiltonian [17]. The qubits of a D-Wave QPU are superconducting flux qubits, where the states are determined by whether the current is flowing clockwise or counterclockwise, or in the superposition of these. The interconnection between the qubits are called *couplers* and have less control circuits than the qubits. Their control represent the coupling strength (J_{ij} of Eq. 2). As the system is susceptible to noise (e.g., cross-talk, environment), the results produced by the QPU might not always represent the solution to the original problem. Furthermore, since the QPU is an analog device with limited precision, some problems might not be presentable at all. In the D-Wave QPU, the physical lattice of qubits and couplers has a limited connectivity and can be described special graph structures called *Chimera* and *Pegasus*. These architectures will be described in more detail in Sec. V-A.

III. MIMO DECODING AS A QUBO PROBLEM

A. Maximum Likelihood Detection for MIMO Decoding

In a multiple user MIMO (MU-MIMO) system there are multiple antennas that can simultaneously transmit to multiple recipients and vice-versa. The transmission goes through the channel matrix of $N_t \times N_r$ in case of N_t transmit and N_r receive antennas. The receiver then has to decode the vector of complex receive symbols ($\mathbf{y} \in \mathbb{C}^{N_r}$) to restore the originally transmitted bits. Such a MIMO system can be modelled as: $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where the vector of complex transmit symbols $\mathbf{x} \in \mathbb{C}^{N_t}$ is affected by the complex channel matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ and the additive Gaussian white noise $\mathbf{n} \in \mathbb{C}^{N_r}$. In this text, we refer to such a system as a MIMO setup (or scenario) of $N_t \times N_r$.

Other than spatial multiplexing, digital modulation is also present in these communication scenarios. [18] This means that each symbol can represent multiple bits (dependent on the modulation scheme), where the bit-to-symbol mapping is usually given by the *constellation* (\mathcal{O}).

The MIMO ML decoding [19] is a search in a space of $|\mathcal{O}|^{N_t}$ for some symbol vector $\hat{\mathbf{v}}$ that minimizes the symbol errors, with variable \mathbf{v} representing all the possible vector of transmitted symbols:

$$\hat{\mathbf{v}} = \arg \min_{\mathbf{v} \in \mathcal{O}^{N_t}} \|\mathbf{y} - \mathbf{H}\mathbf{v}\|^2. \quad (5)$$

The result is the decoded symbol vector $\hat{\mathbf{v}}$, which is mapped to the decoded bit-string $\hat{\mathbf{b}}$ according to the used constellation.

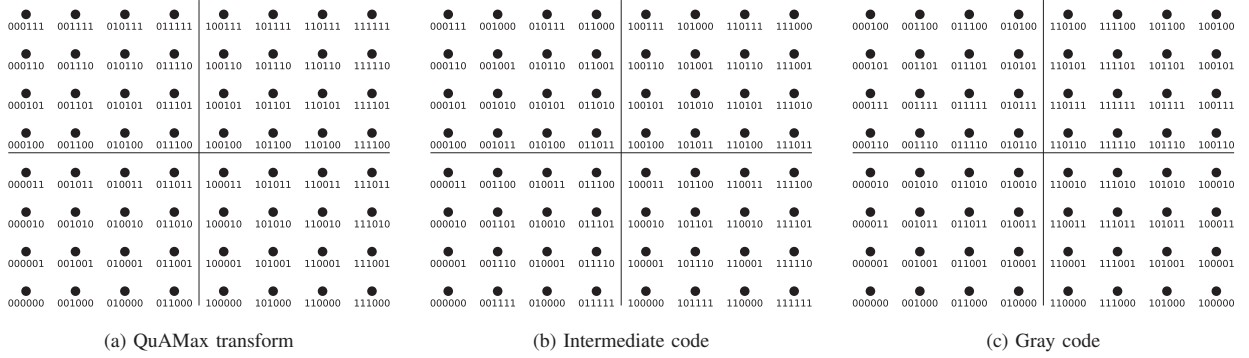


Fig. 1. The process of converting the *QuAMax* encoded bits to the original Gray-coded transmission with 64-QAM. The *QuAMax* constellation (1a) encodes non-Gray-coded bit-strings in order to retain linear QUBO translation, but the transmission is done using the Gray code constellation (1c), causing disparity. By using the Intermediate code's (1b) constellation, we can easily convert back to the original sent data on the output of the minimization procedure (the *QuAMax* code).

The idea of ML decoding is to maximize throughput via minimizing the bit error rate, however, classical algorithms (such *Sphere Decoding* [20]) can hit a computational wall quickly as the complexity grows exponentially with the number of transmitters [21]. For this reason, practical (polynomial-time) MIMO decoding algorithms (like *Zero Forcing* [22]) often settle for sub-optimal solutions that are easier to obtain.

Kim et al. [1] show in their work how QA can speed up solution of ML MIMO decoding. We derive the explicit QUBO formulation for 64-QAM and present a study of embedding capacity of two currently available D-Wave QPUs of different size and topology. Through this experimental study, we uncover the limiting problem sizes that can be implemented on real-world Quantum Annealers for Massive-MIMO problems. In this study we also highlighted the efficiency of the embedding methods via length statistics of chains representing logical qubits – a crucial factor in annealing performance in non-fully connected Quantum Annealer topologies. For this we also use the non-trivial QUBO expansion of Eq. 5. Furthermore, we implement enhanced embedding algorithms to further improve the found limiting problem sizes of each scenarios. In this work we extend on the symmetric problem statement of [1], where $N_t = N_r$, corresponding to the state-of-the-art commercially available Massive MIMO equipment [23].

However, the methodology can easily be applied to $N_t \neq N_r$ scenarios. In these cases, the number of QUBO variables always equals to N_t , therefore the embedding of any $N_t \times N_r$ MIMO ML decoding corresponds to the structure of $N_t \times N_t$.

B. Overview of Ising Formulation of the ML MIMO Decoding

In order to use a D-Wave QPU for solving MIMO ML decoding, we should formulate the QUBO formula of the optimization problem. Some basic modulation techniques are already investigated by Kim. et al. [1], these will be briefly described in the following subsections. Following that, we provide an extension of the ML to QUBO conversion (*QuAMax transform*) for the 64-QAM modulation.

In general, the *QuAMax* has a qubit requirement of $N_t \log_2 M$ for a symmetric $N_t \times N_r$ MIMO setup and a constellation of size M . One can arrive at the QUBO coefficients

by substituting the *QuAMax*-transformed v_i symbols to their respective equations with QUBO variables. The exact formulas to get the QUBO coefficients are described in [1].

1) *BPSK*: In the case of Binary Phase Shift Keying modulation, each symbol consists of one bit. There is a 180° phase-shifting between two possible states. The BPSK is the easiest modulation technique to convert into the QUBO form, since each symbol $v_i \in \{-1, 1\}$ can be mapped to $2q_i - 1$, where q_i is the i th QUBO variable.

2) *QPSK*: The Quadrature Phase Shift Keying modulation transmits two bits as one symbol and the phase shifting between possible states is 90° . In case of QPSK, each symbol is a complex number number: $v_i = v_i^I + jv_i^Q$, and $v_i \in \{\pm 1 \pm 1j\}$, therefore we need 2 qubits to encode a single symbol: $v_i = (2q_{2i-1} - 1) + j(2q_{2i} - 1)$.

3) *16-QAM*: 16-QAM is a Quadrature Amplitude Modulation with $|\mathcal{O}| = 16$, that can transmit 4 bits per symbol. As both dimensions can have 4 possible values ($v_i^I, v_i^Q \in \{\pm 1, \pm 3\}$), we need 2 qubits per dimension to describe each symbol. Since we want a linear transformation, the mapping: $v_i = (4q_{4i-3} + 2q_{4i-2} - 3) + j(4q_{4i-1} + 2q_{4i} - 3)$ seems like a good choice.

However, since in wireless communication we often use Gray code¹ as a mean to avoid bit errors, we need to consider a mapping that takes this into account.

Nevertheless, as pointed out by Kim et al. [1], the mapping of a Gray-coded constellation to the QUBO form will always incorporate higher-order terms, that are not allowed in our Ising model. For this, following [1], we can retain a linear *QuAMax* transform with a non-Gray-coded bit encoding and use a post-processing technique to regain the original bits. We will elaborate on this technique in Sec. IV.

IV. EXTENSION TO THE 64-QAM MODULATION

Here we present the method for extension of the QUBO formulation of [1] to the 64-QAM case. 64-QAM is an amplitude modulation that can transmit 6 bits per symbol (Fig. 1a).

¹Gray code is an encoding technique where each subsequent symbol is encoded by a bit pattern that only differs in one bit in order to make error correction more robust.

Exploring Embeddings for MIMO Channel Decoding on Quantum Annealers

TABLE I
QUBIT REQUIREMENTS OF DIFFERENT ML ENCODED MIMO CONFIGURATIONS.

Config.	BPKS	QPSK	16-QAM	64-QAM
10 × 10	10	20	40	60
20 × 20	20	40	80	120
30 × 30	30	60	120	180
40 × 40	40	80	160	240
60 × 60	60	120	240	360
80 × 80	80	160	320	480
100 × 100	100	200	400	600
120 × 120	120	240	480	720
140 × 140	140	280	560	840
160 × 160	160	320	640	960
180 × 180	180	360	720	1080

Following [1], in order for the QUBO form to remain linear, we use Gray coding for the transmitted symbols and interpret the result according the QuAMax transform constellation. The QA process solves for the QuAMax constellation which then needs to be mapped back to the Gray code to recover the original message.

In our extended formula, to accommodate both dimensions ($v_i^I, v_i^Q \in \{\pm 1, \pm 3, \pm 5, \pm 7\}$), we need 6 qubits (3 per dimension) and so we chose the suitable linear transformation:

$$v_i = (8q_{6i-5} + 4q_{6i-4} + 2q_{6i-3} - 7) + j(8q_{6i-2} + 4q_{6i-1} + 2q_{6i} - 7) . \quad (6)$$

From this, we get the conversion formulae to the Ising energy function values by expanding Eq. 5. The complete description is given in Ref. [24]. The constellations used to map of symbols from QuAMax to the Intermediate code and finally to Gray code is illustrated in Fig. 1.

V. PROBLEM EMBEDDING ONTO D-WAVE QPUS

A. D-Wave architectures

D-Wave currently has two types of publicly available QPUs. The older model is the D-Wave 2000Q, with up to 2048 physical qubits in a Chimera topology. The newer model is the D-Wave Advantage, with up to 5640 physical qubits in a Pegasus topology. The Chimera C_{16} topology has $K_{4,4}$ graphs in a 16×16 lattice for which the sub-graph is shown in Fig. 2a. The recently released D-Wave Advantage QPU has a Pegasus P_{16} topology, which is more connected than the C_{16} since it has degree 15 (each qubit is connected to 15 other qubits via couplers), while the older model has degree 6.

B. Embedding methods

In the D-Wave programming model the linear and quadratic coefficients of the QUBO problem can be mapped to qubits and the connections of qubits, respectively. The direct mapping is rarely possible, since the QPU has a sparse graph topology, therefore we need to use embedding methods transforming

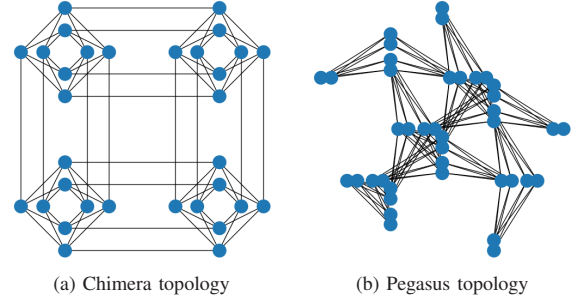


Fig. 2. Unit cells and their connections of currently available D-Wave hardware topology. The older QPU (D-Wave 2000Q) uses a less connected Chimera C_{16} architecture (2a) with degree 6, while the new Advantage System uses the denser Pegasus P_{16} architecture with degree of 15 (2b). Here, only C_2 and a P_2 graphs are depicted; the complete working graphs have around 2000 and 5000 qubits, respectively (some qubits are disabled due to manufacturing imperfection).

logical qubits to a chain of connected physical qubits [25]. The standard D-Wave MinorMiner (MM) embedding algorithm [26] works by searching loops in a set of logical qubits and interactions and mapping these qubits to the physical topology creating physical qubit chains (see Fig. 3).

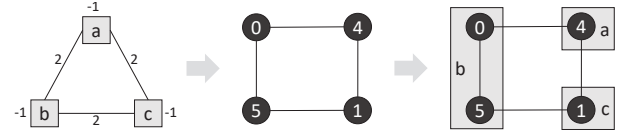


Fig. 3. The MinorMiner embedding method, source: [27]. A heuristic process which tries to find the best way to map source graph to the structure of the target graph. In case of D-Wave Quantum Annealers, we need to map QUBO problems (with up K_n structure) to either a Chimera graph or a Pegasus graph, both of which have limited connectivity.

The D-Wave open-source SDK [28] implements this algorithm for both Chimera and Pegasus architectures. Since embedding can have a direct effect on the quality of the solution of the annealing, finding close to optimal embedding is a crucial part of the problem solving. One measure of embedding quality is the chain lengths of physical qubits that represent a single logical QUBO variable. The shorter the chains, the closer the solution of the embedded problem to the original one, since current architectures are noisy and have imperfect qubits. Hence, in our experiments, we used improved embedding methods.

The Clique-Based MinorMiner (CLMM) and the Spring-Based MinorMiner (SPMM) embeddings, presented in [29], improve upon embeddings by providing heuristic initial chains to start the search from. We have experimented with these techniques to extend the maximal problem sizes that could be embedded into the hardware graph. SPMM uses a standard graph layout of the QUBO variables matched to the physical qubits on the same plane as initial chains, while CLMM constructs a native clique embedding with uniform chains as a starting point. Both algorithms pass the resulting initial chains to MM to find the final embedding.

C. MIMO ML decoding embedding onto the Chimera C-16 and Pegasus P-16 Architectures

To establish a baseline on what problem sizes map easily to the QPU, we used the native clique embedding for Chimera and Pegasus topologies [30] implemented in the D-Wave Ocean SDK. For the Chimera C_{16} , the largest native clique is K_{64} , while for the Pegasus P_{16} , that is K_{180} .

The QUBO equation of an N -QAM $N_t \times N_r$ MIMO ML setup is (almost) equivalent to a $K_{N_t \log_2 N}$ clique [1]. We summarize our results of the exact embedding requirements of the extended set of MIMO optimization problems showing the level of superiority of the P_{16} topology in Table I. For different ML encoded MIMO configurations (with native clique embedding) the green cells indicate feasible mapping to both QPU architectures, yellow cells are infeasible on C_{16} and red cells indicate non-feasibility on both the 2000Q and the Advantage system. Using the new, more connected architecture, the physical qubit requirement decreases significantly for each problem size, making possible the embedding of larger scenarios.

D. Enhanced embedding via heuristic algorithms

To extend the baseline limits, we now discuss the possibilities using D-Wave's heuristic algorithm and its extensions. We were able to supersede the native clique embedding by using the CLMM algorithm. According to our results for extension of base MM for complete graphs, CLMM yields embeddings of larger MIMO setups with fewer physical qubit requirements in most of the problem sizes. We find SPMM method to be inferior in case of the clique embeddings. These limits exceed the ones published in [1] on the Chimera C_{16} architecture.

TABLE II
UPPER LIMITS OF QUBO-FORM ML MIMO DECODING PROBLEMS
MAPPING ONTO THE CHIMERA C_{16} AND PEGASUS P_{16} ARCHITECTURES.

Arch	Method	BPSK	QPSK	16-QAM	64-QAM
C_{16}	CLIQUE	64×64	32×32	16×16	10×10
	CLMM	65×65	33×33	16×16	11×11
P_{16}	CLIQUE	180×180	90×90	45×45	30×30
	CLMM	182×182	91×91	45×45	30×30

Furthermore, using the new Pegasus P_{16} topology, we could more than double the number of users for each modulation and still embed the problem. This is of course, due to the new QPU architecture being larger (in qubit count) and more connected.

We highlighted the performance comparison of all embedding methods in case of the 64-QAM modulation in Fig. 4, testing all embeddable $2^n \times 2^n$ MIMO scenarios. As mentioned earlier the CLMM algorithm had the best performance, with less number of physical qubit requirement for each problem size.

Table II summarizes the found theoretical and heuristic limits for the largest MIMO scenarios for each modulation comparing the baseline (native clique embedding) to the best-performing heuristic algorithm (CLMM).

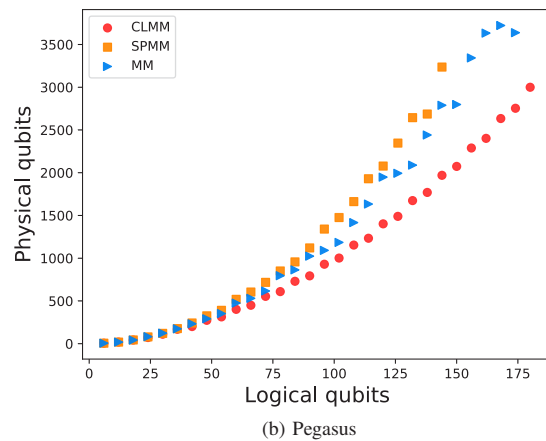
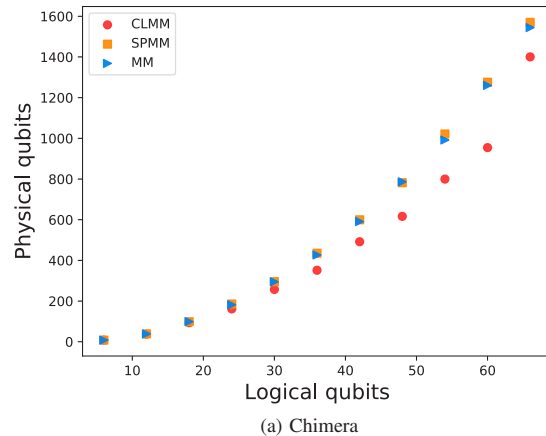


Fig. 4. Logical and physical qubits requirements of the embedding of the QUBO-formulated 64-QAM ML MIMO decoding problem with clique based minor miner for Chimera C_{16} (4a) and Pegasus P_{16} (4b) architectures. The number of logical qubits correspond to $6N$, $2 \leq N \leq N_{max}$, where N is number of transmitters (and receivers) in the MIMO setup and N_{max} is the largest MIMO scenario that the given QPU can handle (11 and 30 for C_{16} and P_{16} , respectively).

VI. CONCLUSIONS AND OUTLOOK

In this paper, we presented the extension of the MIMO ML decoding as QUBO problem to the 64-QAM modulation scheme. By creating the required constellation diagrams, we showed that one can use a linear QuAMax formulation and still employ Gray code to reduce the number of bit errors during transmission.

Furthermore, we extended the range of embedding of MIMO ML decoding problems in both the dimension of modulation complexity and transmitter number. We used the Pegasus P_{16} architecture of the new D-Wave Advantage system to show the limits of each modulation scheme on a state-of-the-art QPU. Additionally, we explored heuristic embedding methods (such as MM, SPMM, CLMM) to further improve the range of MIMO setups that could benefit from quantum speedup. As a result, we were able to double the embeddable problem sizes compared to the earlier work.

Exploring Embeddings for MIMO Channel Decoding on Quantum Annealers

The currently available commercial Massive-MIMO solutions [23] have already reached the point of being able to decode transmissions beyond 64-QAM modulation in 64×64 -MIMO setups. However, the potential advantage of QA lies in the problem sizes that are prohibitive for classical computers. Since the new Pegasus architecture shows promising scalability improvements regarding maximal problem size, we believe that future QPUs will stand as viable alternative to these classical solutions.

We believe that these result could further be improved by using manual embedding. We leave the study of this possibility to future work.

For next steps, we would like to study the performance of the D-Wave Advantage system on the ML decoding problems. Furthermore, since the 64-QAM case has never been tested before on any QPU, we see potential in testing it on both the D-Wave 2000Q and Advantage systems.

ACKNOWLEDGMENT

Zsolt Tabi and Zoltán Zimborás would like to thank the support of the Hungarian Quantum Technology National Excellence Program (Project No. 2017-1.2.1-NKP-2017-00001), and acknowledge also support from the Hungarian National Research, Development and Innovation Office (NKFIH) within the Quantum Information National Laboratory of Hungary and through Grants No. FK 135220, K124176, KH129601.

REFERENCES

- [1] M. Kim, D. Venturelli, and K. Jamieson, "Leveraging quantum annealing for large MIMO processing in centralized radio access networks," in *Proceedings of the ACM Special Interest Group on Data Communication*, 2019, pp. 241–255.
- [2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [3] S. E. Gaily and S. Imre, "Quantum optimization of resource distribution management for multi-task, multi-subtasks," *Infocommunications Journal*, vol. XI, no. 4, pp. 47–53, December 2019. doi: 10.36244/ICJ.2019.4.7
- [4] T. Kadowaki and H. Nishimori, "Quantum annealing in the transverse Ising model," *Physical Review E*, vol. 58, no. 5, p. 5355, 1998. doi: 10.1103/PhysRevE.58.5355
- [5] E. Farhi, J. Goldstone, and S. Gutmann, "Quantum adiabatic evolution algorithms versus simulated annealing," *preprint arxiv:quantph/0201031*, 2002.
- [6] M. W. Johnson et al., "Quantum annealing with manufactured spins," *Nature*, vol. 473, no. 7346, pp. 194–198, 2011. doi: 10.1038/nature10012
- [7] E. Dahlman, S. Parkvall, and J. Sköld, "Chapter 24 - new 5G radioaccess technology," in *4G LTE-Advanced Pro and The Road to 5G (Third Edition)*, 3rd ed., E. Dahlman, S. Parkvall, and J. Sköld, Eds. Academic Press, 2016, pp. 547–573. ISBN 978-0-12-804575-6
- [8] G. Fodor, L. Pap, and M. Telek, "Recent advances in acquiring channel state information in cellular mimo systems," *Infocommunications Journal*, vol. XI, no. 3, pp. 2–12, September 2019. doi: 10.36244/ICJ.2019.3.2
- [9] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE communications magazine*, vol. 52, no. 2, pp. 186–195, 2014. doi: 10.1109/MCOM.2014.6736761
- [10] B. Trotobas, A. Nafkha, and Y. Loüet, "A review to massive MIMO detection algorithms: Theory and implementation," in *Advanced Radio Frequency Antennas for Modern Communication and Medical Systems*. IntechOpen, 2020.
- [11] K. Chen, "C-RAN: The road towards green RAN," China Mobile Research Institute, Beijing, China, China Mobile Research Institute, Tech. Rep., Oct 2011.
- [12] A. Lucas, "Ising formulations of many NP problems," *Frontiers in Physics*, vol. 2, p. 5, 2014.
- [13] F. Glover, G. Kochenberger, and Y. Du, "A tutorial on formulating and using QUBO models," *preprint arXiv:1811.11538*, 2018.
- [14] K. Boothby, P. Bunyk, J. Raymond, and A. Roy, "Next-generation topology of D-Wave quantum processors," *arXiv preprint arXiv:2003.00133*, 2020.
- [15] M. Pincus, "Letter to the editor monte carlo method for the approximate solution of certain types of constrained optimization problems," *Operations research*, vol. 18, no. 6, pp. 1225–1228, 1970. doi: 10.1287/opre.18.6.1225
- [16] C. C. McGeoch, "Adiabatic quantum computation and quantum annealing: Theory and practice," *Synthesis Lectures on Quantum Computing*, vol. 5, no. 2, pp. 1–93, 2014.
- [17] D-Wave, "Technical description of the D-Wave quantum processing unit, user manual," https://docs.dwavesys.com/docs/latest/doc_qpu.html, 2021.
- [18] P. Botsinis, S. X. Ng, and L. Hanzo, "Low-complexity iterative quantum multi-user detection in SDMA systems," in *2014 IEEE International Conference on Communications (ICC)*. IEEE, 2014. pp. 5592–5597. doi: 10.1109/ACCESS.2014.2322013.
- [19] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on information theory*, vol. 49, no. 10, pp. 2389–2402, 2003. doi: 10.1109/TIT.2003.817444
- [20] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bolcskei, "VLSI implementation of MIMO detection using the sphere decoding algorithm," *IEEE Journal of solid-state circuits*, vol. 40, no. 7, pp. 1566–1577, 2005. doi: 10.1109/JSSC.2005.847505
- [21] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE transactions on signal processing*, vol. 53, no. 8, pp. 2806–2818, 2005. doi: 10.1109/TSP.2005.850352
- [22] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE transactions on signal processing*, vol. 52, no. 2, pp. 461–471, 2004. doi: 10.1109/TSP.2003.821107
- [23] "Advanced antenna systems for 5G networks," Ericsson, Tech. Rep., November 2018. [Online]. Available: "https://www.ericsson.com/4a8a87/assets/local/reports-papers/white-papers/10201407_wp_advanced_antenna_system_nov18_181115.pdf"
- [24] Z. Tabi et al., "Evaluation of quantum annealer performance for massive MIMO channel decoding problem," *to be published*, 2021.
- [25] L. Lovász, "Graph minor theory," *Bulletin of the American Mathematical Society*, vol. 43, no. 1, pp. 75–86, 2006. doi: 10.1090/S0273-0979-05-01088-8
- [26] J. Cai, W. G. Macready, and A. Roy, "A practical heuristic for finding graph minors," *arXiv preprint arXiv:1406.2741*, 2014.
- [27] "Minor-embedding a problem onto the QPU," https://docs.dwavesys.com/docs/latest/c_gs_7.html, accessed: 2020-01-30.
- [28] "D-Wave ocean software documentation," <https://docs.ocean.dwavesys.com/en/stable/>, accessed: 2020-01-30.
- [29] S. Zbinden, A. Bärttschi, H. Djidjev, and S. Eidenbenz, "Embedding algorithms for quantum annealers with chimera and pegasus connection topologies," in *International Conference on High Performance Computing*. Springer, 2020. pp. 187–206. doi: 10.1007/978-3-030-50743-5_10
- [30] T. Boothby, A. D. King, and A. Roy, "Fast clique minor generation in chimera qubit connectivity graphs," *Quantum Information Processing*, vol. 15, no. 1, pp. 495–508, 2016. doi: 10.1007/s11128-015-1150-6



Ádám Marosits received his BSc degree in electrical engineering from the Budapest University of Technology and Economics (BME), Budapest, Hungary, and he is currently a MSc student in electrical engineering. Ádám works as a student researcher at Ericsson Research. His research interests include optimization with quantum annealing, furthermore quantum communication and quantum random number generation.



Zsolt I. Tabi was born in Budapest, Hungary in 1990. He received the BSc and MSc degrees in computer science from Eötvös Loránd University, Budapest, Hungary in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree in computer science at Eötvös Loránd University.

From 2016 to 2020, he was a Developer on the MINILINK project with Ericsson Hungary. Since 2020 he is working with the Ericsson Research Team as part of the ETH Quantum Project. He is teaching programming

languages at Eötvös Loránd University. His research interests include programming language design and adiabatic quantum computing.



Zsófia Kallus, PhD works as a master researcher at Ericsson Research, Traffic Analysis and Network Performance Laboratory in Hungary. She holds a PhD degree in Physics from the Doctoral School of Physics of Eötvös Loránd University, Budapest.

She studied open quantum systems and the structure and dynamics of large-scale real-world networks. She joined ER in 2014. Currently, she is focused on emerging quantum technologies, quantum computing applications and explainable machine learning solutions

for optimization and automation of highperformance telecommunication systems. She is also interested in sustainability and urban sciences.



Péter Vaderna works as a master researcher at Ericsson Research, Traffic Analysis and Network Performance Laboratory in Budapest. He received his Ph.D. in physics from the Department of Physics of Complex Systems at Eötvös Loránd University, Budapest in 2008, in the area of traffic modeling in communication networks. Currently, he focuses on AI in network analytics, network management and network automation. His research interests also include emerging technologies that can potentially be involved

in various business areas of telecommunication such as quantum AI, camera based positional tracking and AR/VR.



Dr. István Gódor (M'01–SM'20) is a research leader at Ericsson Research, Traffic Analysis and Network Performance Laboratory of Ericsson Hungary. He is a senior member of IEEE and a member of public body of the Hungarian Academy of Sciences (MTA). He holds a Ph.D. from Budapest University of Technology and Economics. He was awarded the 2014 IEEE Communications Society Fred W. Ellersick Prize.

His research interests include network design, traffic analysis and modeling, self-organizing networks, energy efficiency, collective sensing and industrial IoT & smart manufacturing.



Zoltán Zimborás is the head of the Quantum Computing and Information Group at the Wigner Research Centre for Physics (Budapest). He obtained his PhD at the Eötvös Loránd University in Budapest and worked later at ISI Torino, University College London, and Freie Universität Berlin before returning to Hungary. He is member of several editorial boards and is also a board member of QWorld, a nonprofit organization that aims to develop open-source software for quantum programming at all levels. His research interests include Quantum Computing, Quantum Control and Quantum Statistical Physics.