Quantum Advantage via Qubit Belief Propagation

Narayanan Rengaswamy*, Kaushik P. Seshadreesan†, Saikat Guha†, and Henry D. Pfister*
*Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA
†College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA
Email: {narayanan.rengaswamy, henry.pfister}@duke.edu, {kaushiksesh, saikat}@arizona.edu

***Abstract*—Quantum technologies are maturing by the day and their near-term applications are now of great interest. Deep-space optical communication involves transmission over the *pure-state* classical-quantum channel. For optimal detection, a joint measurement on all output qubits is required in general. Since this is hard to realize, current (sub-optimal) schemes perform symbol-by-symbol detection followed by classical post-processing. In this paper we focus on a recently proposed belief propagation algorithm by Renes that passes *qubit* messages on the factor graph of a classical error-correcting code. More importantly, it only involves single-qubit Pauli measurements during the process. For an example 5-bit code, we analyze the involved density matrices and calculate the error probabilities on this channel. Then we numerically compute the optimal joint detection limit using the Yuen-Kennedy-Lax conditions and demonstrate that the calculated error probabilities for this algorithm appear to achieve this limit. This represents a first step towards achieveing quantum *communication* advantage. We verify our analysis using Monte-Carlo simulations in practice.**

***Index Terms*—Factor graphs, Helstrom measurement, classical-quantum channels, linear codes, belief propagation**

I. Introduction

The field of quantum computation and quantum information has made tremendous progress in the last three decades and, today, quantum computers are being considered for real-world applications [1]. Recently, Google and NASA physically demonstrated a quantum advantage over classical computers for a random circuit sampling task on their 53-qubit machine [2]. Despite IBM's dispute on the extent of the advantage [3], this milestone demonstration has reinforced the interest in finding useful near-term applications that can exploit such quantum advantages. Researchers have explored applications such as simulation of quantum systems, metrology for high-precision measurements, chemistry for nitrogen fixation, optimization for logistics, and prime factorization for digital security. However, quantum advantages in communication settings have been less explored. We make progress towards identifying a communication problem where low-complexity schemes that require only near-term quantum computing can provide a significant advantage over classical schemes.

The work of NR and HP was supported in part by the National Science Foundation (NSF) under Grant No. 1908730 and 1910571. KPS and SG acknowledge support from NSF, Grant No. 1855879, and the Office of Naval Research MURI program on Optical Computing, Grant No. N00014-14-1-0505. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these sponsors.

The pure-loss optical channel in deep-space optical communications [4] can be modeled as the so-called *pure-state* classical-quantum (CQ) channel. The ultimate Holevo capacity of this channel is well-known and CQ polar codes achieve this capacity under a quantum successive-cancellation decoder [5]. However, this decoder requires joint measurements on all the code qubits output by the channel, which is hard to realize in practice. Similarly, even for a given instance of a code transmitted over this channel, a collective measurement is required in general to optimally distinguish all the codewords. Therefore, an interesting open problem is to construct a low-complexity decoder that does not involve joint measurements.

Message-passing algorithms form a powerful class of computationally-efficient classical algorithms and are widely used to solve problems defined on graphs. In particular, *belief propagation (BP)* is a message-passing algorithm that is used to efficiently compute posterior marginal distributions in statistical-inference problems [6], [7]. For the decoding of linear codes, BP is executed on a *factor-graph (FG)* for the code. The FG is a bipartite graph that encodes the correlations between the bits of each codeword. When the FG is a tree, BP can perform bit-wise maximum-a-posteriori (bit-MAP) estimation and hence minimize the bit-error rate.

Recently, Renes [10] proposed a CQ generalization of the BP algorithm where the messages on the edges of the graph are now qubits instead of probabilities, and single-qubit measurements are performed along the way. We refer to this algorithm as *belief propagation with quantum messages (BPQM)*, and we refine the algorithm while retaining this name. Since one does not observe quantum data unless one measures it, the notion of a *posterior* in the case of CQ channels does not appear to be well-defined. Note that current receivers for deep-space optical communications effectively measure each qubit output by the channel and then perform classical post-processing on the resulting bits. Since this is sub-optimal, we primarily want to avoid measuring each qubit, which means the posterior might not yet be defined. Hence, the goal of BPQM remains unclear even though it retains the flavor of (quantum) statistical inference. Renes circumvents this problem by viewing BP as performing statistical inference locally on the "channel convolutions" induced at nodes of the factor graph, and then generalizing these channel convolutions to the quantum setting. We review this perspective shortly.

Although the algorithm is defined in [10], no simulation or analysis of the BPQM algorithm was presented. In this paper,

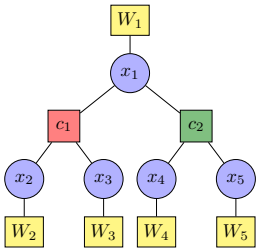

Fig. 1. Factor graph for the 5-bit linear code $\mathcal{C}$ in the running example.

we use a simple $[5, 3, 2]$ binary code, whose FG is a tree, as an example to understand the workings of the algorithm. We perform a detailed analysis of the involved density matrices and suggest some refinements to the algorithm. We also calculate both bit- and block-error probabilities for BPQM on this code and verify them by simulations[1]. Using the Yuen-Kennedy-Lax (YKL) conditions [11], [12] we numerically compute the fundamental joint (codeword) Helstrom limit for optimally distinguishing the 8 codewords of this code. Finally we show that the BPQM block-error rate appears to match this limit, i.e., it seems to be quantum-optimal. This represents a significant quantum advantage over current receivers, which measure each qubit and then perform classical post-processing.

In [13], we also provide a full circuit decomposition of BPQM (for this code) in terms of standard single-, two-qubit and Toffoli gates. Since the circuits for BPQM have a specific structure, this is an application that does not require a *universal* quantum computer to exploit this advantage. It remains to be seen if this advantage will persist for this code under the noise levels in current systems. But this is an opportunity for experimentalists to achieve the advantage by making this specific circuit sufficiently reliable.

II. Classical Belief Propagation (BP)

A. Decoding Linear Codes Using BP

Consider the $[5, 3, 2]$ code $\mathcal{C} \triangleq \{\underline{x} \in \{0,1\}^5 \colon H\underline{x}^T = \underline{0}^T\}$, where the *parity-check matrix* is given by

$$H = \begin{array}{c} \\ c_1 \\ c_2 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}. \tag{1}$$

The *factor graph* for $\mathcal{C}$ corresponding to this particular choice of H is a bipartite graph that consists of 5 *variable nodes* (circles) connected to 2 *factor nodes* (squares) according to the parity-checks defined by the two rows of H (see Fig. 1).

Let $W(y|x)$ represent a binary-input memoryless channel where $x \in \{0, 1\}$ is the input, $y \in \mathcal{Y}$ is the output for some output alphabet $\mathcal{Y}$, and $W(y|x)$ represents the channel transition probability. Given the output (vector) $\underline{y}$ from the channel, the decoder for the code $\mathcal{C}$ needs to determine the codeword $\underline{x}$ that was actually sent at the input. The *block-maximum-a-posteriori* (MAP) decoder, which is optimal for minimizing the average decoding error probability, calculates the posterior probability for each codeword in the code, given $\underline{y}$, and chooses the codeword with the maximum value.

[1]Our implementation for the 5-bit code: https://github.com/nrenga/bpqm.

Assuming all codewords are equally likely, $p(\underline{x}|\underline{y}) =$

$$\frac{\prod_{k=1}^{5} W(y_k|x_k) \cdot \mathbb{P}[\underline{x} \in \mathcal{C}]}{p(\underline{y})} \tag{2}$$

$$\propto W(y_1|x_1) \cdot [\mathbb{I}(x_1 \oplus x_2 \oplus x_3 = 0) W(y_2|x_2) W(y_3|x_3)] \cdot [\mathbb{I}(x_1 \oplus x_4 \oplus x_5 = 0) W(y_4|x_4) W(y_5|x_5)], \tag{3}$$

where the constant of proportionality is independent of $\underline{x}$. Then it determines the maximizing codeword according to $\hat{\underline{x}}^{\text{MAP}} \triangleq \text{argmax}_{\underline{x} \in \{0,1\}^5} p(\underline{x}|\underline{y})$. An alternative decoding scheme is the *bit-MAP* decoder which marginalizes $p(\underline{x}|\underline{y})$ for each bit and makes a decision bitwise. Hence, to decode bit x_1, the bit-MAP decoder computes

$$\hat{x}_1^{\text{MAP}} \triangleq \underset{x_1 \in \{0,1\}}{\text{argmax}} \sum_{x_2, x_3, x_4, x_5 \in \{0,1\}^4} p(\underline{x}|\underline{y}). \tag{4}$$

The idea of bit-MAP is that this marginalization can be done efficiently when the joint probability density $p(\underline{x}|\underline{y})$ factors into terms involving *disjoint* sets of variables. In our running example, we can use the distributive property of addition over multiplication and compute the sums involved in the two square brackets (from (3)) *simultaneously*. Then the results can be pooled in a final step that takes their product and multiplies the result with $W(y_1|x_1)$. This is exactly BP on this FG, since the two local sums can be interpreted as "local beliefs" about the variable x_1 that are propagated to be combined with the "belief" from the direct channel observation y_1.

B. Induced Channels in BP

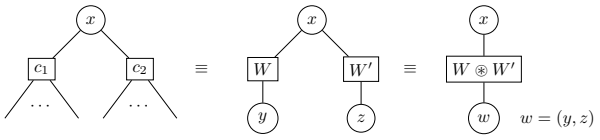

Fig. 2. Channel combining at a VN using the induced channels at the node.

The two induced channels can be combined into a channel

$$[W \circledast W'](y, z|x) = W(y|x) \cdot W'(z|x). \tag{5}$$

This is the *variable node (VN) convolution* of two channels.

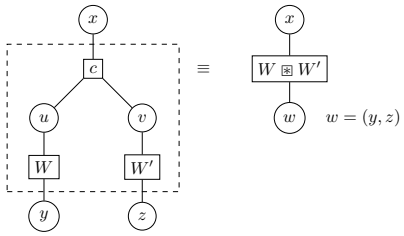

Fig. 3. Channel combining at a FN using the induced channels at the node.

Likewise, the *factor node (FN) convolution* of two channels is

$$[W \circledast W'](y, z|x) = \frac{1}{2} W(y|x) \cdot W'(z|0) + \frac{1}{2} W(y|x \oplus 1) \cdot W'(z|1). \tag{6}$$

Hence, during the node updates, BP is simply performing local inference over these local channels $[W \circledast W'], [W \circledcirc W']$, i.e., calculating the local posterior for x given (y, z). For some more discussion on these induced channels, see [13].

This perspective on BP crucially aids us in defining the quantum channel combining operations for a classical-quantum (CQ) channel [14]. As discussed in [5] and [10], for a CQ channel denoted by $W(x) \triangleq W(|x\rangle\langle x|), x \in \{0,1\}$, we can define the variable and factor node convolutions as

$$[W \circledast W'](x) \triangleq W(x) \otimes W'(x), \tag{7}$$

$$[W \circledcirc W'](x) \triangleq \frac{1}{2} W(x) \otimes W'(0) + \frac{1}{2} W(x \oplus 1) \otimes W'(1). \tag{8}$$

III. BP with Quantum Messages (BPQM)

A. Pure-State Channel

The *pure-state* classical-quantum (CQ) channel is defined for classical inputs $x \equiv |x\rangle\langle x|, x \in \{0,1\}$, as

$$\begin{aligned} W(x) &\triangleq |\theta\rangle\langle 0| \cdot |x\rangle\langle x| \cdot |0\rangle\langle\theta| + |-\theta\rangle\langle 1| \cdot |x\rangle\langle x| \cdot |1\rangle\langle -\theta| \\ &= |(-1)^x\theta\rangle\langle(-1)^x\theta|, \end{aligned} \tag{9}$$

$$|\pm\theta\rangle \triangleq \cos\frac{\theta}{2}|0\rangle \pm \sin\frac{\theta}{2}|1\rangle. \tag{10}$$

Hence, the Kraus operators for the channel can be taken to be $M_0 = |\theta\rangle\langle 0|, M_1 = |-\theta\rangle\langle 1|$. The *fidelity* of this channel is

$$F(W) \triangleq |\langle\theta|-\theta\rangle|^2 = \cos^2\theta, \ \cos\theta = 2\cos^2\frac{\theta}{2} - 1. \tag{11}$$

Unless $\theta = \pi/2$, the two output states are not perfectly distinguishable, which introduces uncertainty at the receiver. The ultimate Holevo capacity of this channel is given by [4]

$$C_\infty(W) = h_2\left(\frac{1+\sqrt{F(W)}}{2}\right). \tag{12}$$

The *Helstrom measurement* [15], [16] to optimally distinguish between density matrices ρ_0 and ρ_1 is given by the positive operator-valued measurement (POVM) $\{\Pi_{\text{Hel}}, \mathbb{I} - \Pi_{\text{Hel}}\}$:

$$\Pi_{\text{Hel}} \triangleq \sum_{i:\ \lambda_i \geq 0} |i\rangle\langle i|, \quad (\rho_0 - \rho_1)|i\rangle = \lambda_i |i\rangle. \tag{13}$$

For the pure state channel it is easy to calculate that $\rho_0 - \rho_1 = |\theta\rangle\langle\theta| - |-\theta\rangle\langle -\theta| = \sin\theta \cdot X$, so that the Helstrom measurement is projecting onto the Pauli X basis, i.e., the POVM is $\{|+\rangle\langle +|, |-\rangle\langle -|\}$. In practice, the *Dolinar receiver* [17] for the BPSK modulated pure-loss optical channel optimally measures each output qubit. Hence, it induces the binary symmetric channel BSC($P_{\min}$) with $P_{\min} = (1-\sqrt{1-F(W)})/2$, which is the minimum probability of error to distinguish between the states $\{|\theta\rangle, |-\theta\rangle\}$ [4]. If we implemented a classical optimal (block-MAP) decoder on this induced BSC, then the capacity that is attainable is $C_1(W) = 1 - h_2(P_{\min})$, where the subscript "1" indicates that we are performing symbol-by-symbol measurements and not a collective measurement. It can be easily checked that $C_1(W) \ll C_\infty(W)$, and classical-quantum polar codes equipped with a quantum successive-cancellation decoder close this gap [4], [5]. However, this decoder is hard to realize in the lab. Hence, an open problem is to analyze if this gap is closed by BPQM, which can be mapped into a "successive-cancellation-type" decoder [10].

B. Node Operations in BPQM

Given the convolution output $[W \circledast W']$ at a VN, a specific unitary $U_\circledast(\theta, \theta')$ (see [10], [13]) is applied to "compress" the information into one qubit and force the other system into $|0\rangle$:

$$U_\circledast(\theta,\theta')\left(|\pm\theta\rangle \otimes |\pm\theta'\rangle\right) = \left|\pm\theta^\circledast\right\rangle \otimes |0\rangle, \tag{14}$$

where $\cos\theta^\circledast \triangleq \cos\theta\cos\theta'$. The VN update is then to pass the qubit from the first system and discard the second system.

At a FN, the induced mixed state $[W \circledcirc W'](x)$ can be turned into the CQ state $\sum_{j\in\{0,1\}} p_j \left|\pm\theta_j^\circledcirc\right\rangle\left\langle\pm\theta_j^\circledcirc\right| \otimes |j\rangle\langle j|$ by performing $U_\circledcirc \triangleq \text{CNOT}_{W\to W'}$, the controlled-NOT gate with W as control and W' as target. Hence,

$$U_\circledcirc\left([W \circledcirc W'](x)\right)U_\circledcirc^\dagger = \sum_{j\in\{0,1\}} p_j \left|\pm\theta_j^\circledcirc, j\right\rangle\left\langle\pm\theta_j^\circledcirc, j\right|,$$

$$p_0 \triangleq \frac{1}{2}(1+\cos\theta\cos\theta'), \ p_1 \triangleq 1 - p_0,$$

$$\cos\theta_0^\circledcirc \triangleq \frac{\cos\theta+\cos\theta'}{1+\cos\theta\cos\theta'}, \ \cos\theta_1^\circledcirc \triangleq \frac{\cos\theta-\cos\theta'}{1-\cos\theta\cos\theta'}. \tag{15}$$

Thus, the FN update measures the second system and passes the resulting qubit from the first system as the message, along with the result of the classical measurement. For reversibility, we will now describe BPQM as a coherent operation that does not measure or discard qubits along the way at the nodes.

IV. BPQM on the 5-bit Code

A. Decoding Bit x_1

Observe that the codewords belonging to the code are $\mathcal{C} = \{00000, 00011, 01100, 01111, 10101, 10110, 11001, 11010\}$. We assume that all the codewords are equally likely to be transmitted, just as in classical BP. Then the task of decoding the value of x_1 involves distinguishing between the density matrices $\rho_1^{(0)}$ and $\rho_1^{(1)}$, which are uniform mixtures of the states corresponding to the codewords that have $x_1 = 0$ and $x_1 = 1$, respectively, i.e., using (9) and taking $\pm \equiv (-1)^{x_1}$,

$$\rho_1^{(x_1)} = |\pm\theta\rangle\langle\pm\theta| \otimes \frac{1}{4} \sum_{\underline{c}\in\mathcal{C}:\ c_1 = x_1} \bigotimes_{i=2}^{5} W(c_i). \tag{16}$$

These density matrices can be written using notation in (8) as $\rho_1^{(x_1)} = \rho_\pm = |\pm\theta\rangle\langle\pm\theta|_1 \otimes [W \circledcirc W](x_1)_{23} \otimes [W \circledcirc W](x_1)_{45}$.

The full BPQM circuit to decode all 5 bits of this code is given in Fig. 4. From the perspective of decoding x_1, the input state to the circuit is $\rho_\pm$. We will track this state through each marked stage in Fig. 4. Define the unitaries and angles

$$U \triangleq \sum_{j,k\in\{0,1\}^2} U_\circledast(\theta_j^\circledcirc, \theta_k^\circledcirc)_{23} \otimes |jk\rangle\langle jk|_{45}, \tag{17}$$

$$\cos\theta_{jk}^\circledast \triangleq \cos\theta_j^\circledcirc \cos\theta_k^\circledcirc, \tag{18}$$

$$V \triangleq \sum_{j,k\in\{0,1\}^2} U_\circledast(\theta, \theta_{jk}^\circledast)_{12} \otimes |jk\rangle\langle jk|_{45}, \tag{19}$$

$$\cos\varphi_{jk}^\circledast \triangleq \cos\theta\cos\theta_{jk}^\circledast. \tag{20}$$

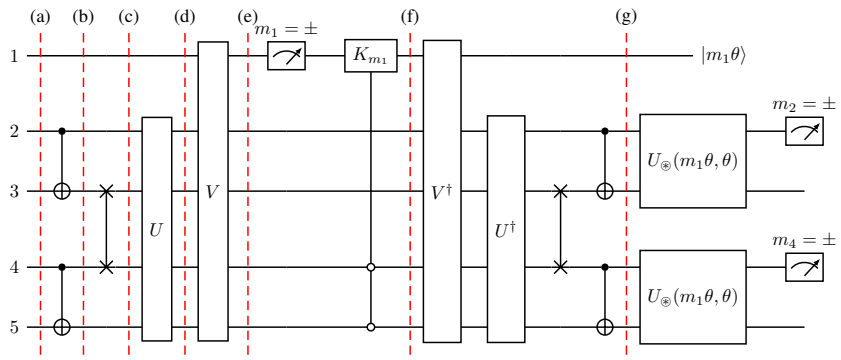

Fig. 4. The full BPQM circuit to decode all bits of the 5-bit code in Fig. 1. The decoded values are related to the measurement results as $m_1 = (-1)^{\hat{x}_1}, m_2 = (-1)^{\hat{x}_2}, m_4 = (-1)^{\hat{x}_4}$, and $\hat{x}_3 = \hat{x}_1 \oplus \hat{x}_2, \hat{x}_5 = \hat{x}_1 \oplus \hat{x}_4$. The open-circled controls indicate that K_{m_1} is coherently controlled by the last two qubits being in the state $|00\rangle_{45}$. The solid line before K_{m_1} indicates that the controlled unitary is applied to the post-measurement state.

First we can decompose $U_\circledast(\theta, \theta')$ into standard gates [18]. Then these coherently controlled unitaries can be decomposed into standard single-, two-qubit, and Toffoli gates (see [13]).

The density matrices at stages (a)-(e) in Fig. 4 are:

$$\text{(a) } \rho_{\pm,a} = |\pm\theta\rangle\langle\pm\theta|_1 \otimes [W \circledast W](x_1)_{23} \otimes [W \circledast W](x_1)_{45},$$

$$\text{(b) } \rho_{\pm,b} = \sum_{j,k\in\{0,1\}^2} p_j p_k \; \mathcal{T}(\pm\theta, \pm\theta_j^{\circledast}, j, \pm\theta_k^{\circledast}, k),$$

$$\text{(c) } \rho_{\pm,c} = \sum_{j,k\in\{0,1\}^2} p_j p_k \; \mathcal{T}(\pm\theta, \pm\theta_j^{\circledast}, \pm\theta_k^{\circledast}, j, k),$$

$$\text{(d) } \sigma_{\pm} = \sum_{j,k\in\{0,1\}^2} p_j p_k \; \mathcal{T}(\pm\theta, \pm\theta_{jk}^{\circledast}, 0, j, k),$$

$$\text{(e) } \Psi_{\pm} = \sum_{j,k\in\{0,1\}^2} p_j p_k \left|\pm\varphi_{jk}^{\circledast}\right\rangle\left\langle\pm\varphi_{jk}^{\circledast}\right|_1 \otimes |00jk\rangle\langle 00jk|.$$

Here, for brevity, $\mathcal{T}(\cdots)$ denotes 5 qubits with the individual subsystems being pure states described by the respective arguments, e.g., $\pm\theta_j^{\circledast} \mapsto \left|\pm\theta_j^{\circledast}\right\rangle\left\langle\pm\theta_j^{\circledast}\right|$. We emphasize that, at each stage, the density matrix is the *expectation* over all pure states that correspond to transmitted codewords with the first bit taking value $x_1 \in \{0, 1\}$. The operations U and V are effectively two-qubit unitary operations, albeit controlled ones, and this phenomenon extends to any factor graph. Evidently, BPQM compresses all the quantum information into system 1 and the problem reduces to distinguishing between $\Psi_{\pm}^{(1)} = \sum_{j,k\in\{0,1\}^2} p_j p_k \left|\pm\varphi_{jk}^{\circledast}\right\rangle\left\langle\pm\varphi_{jk}^{\circledast}\right|_1$, since the other systems are either trivial or completely classical and independent of x_1. Finally, system 1 is measured by projecting onto the Pauli X basis, which we know from the discussion in Section III-A (after (13)) to be the Helstrom measurement to optimally distinguish between the states $\Psi_{\pm}^{(1)}$.

The optimal success probability of distinguishing between $\rho_1^{(0)}$ and $\rho_1^{(1)}$ using a *collective* Helstrom measurement is

$$P_{\text{succ},1}^{\text{Hel}} = \frac{1}{2} + \frac{1}{4}\left\|\rho_1^{(0)} - \rho_1^{(1)}\right\|_1, \quad \|M\|_1 \triangleq \text{Tr}\left(\sqrt{M^\dagger M}\right). \tag{21}$$

The action of BPQM until the final measurement is unitary and the trace norm $\|\cdot\|_1$ is invariant under unitaries. Thus, BPQM does not lose optimality until the final measurement. Since the final measurement is also optimal for distinguishing $\Psi_{\pm}$, BPQM is indeed optimal in decoding the value of x_1.

Thus, by only performing a *single-qubit* measurement at the end of a sequence of unitaries motivated by the FG structure and induced channels in classical BP, BPQM is still optimal to determine x_1. The success probability is $\text{Tr}\left[\Psi_+^{(1)}|+\rangle\langle+|\right]$ and the full post-measurement state by quantum mechanics is

$$\Phi_{m_1} = \sum_{j,k\in\{0,1\}^2} \frac{p_j p_k \left|\left\langle m_1| \pm \varphi_{jk}^{\circledast}\right\rangle\right|^2}{\text{Tr}\left[\Psi_{\pm}^{(1)}|m_1\rangle\langle m_1|\right]} \mathcal{T}(m_1, 0, 0, j, k),$$

where $m_1 \triangleq (-1)^{\hat{x}_1}$. Although BPQM is optimal for x_1, in order to execute the collective Helstrom POVM for distinguishing $\rho_1^{(0)}$ and $\rho_1^{(1)}$, BPQM must apply the inverse of the sequence of operations in (a)-(e) to the state Φ_{m_1}. However, it is actually beneficial to coherently rotate Φ_{m_1} before applying these inverses in order to set up a (slightly) better state discrimination problem for x_2. We think this coherent rotation might be applicable for BPQM in general.

In order to run BPQM for x_1 in reverse to get "as close" to the channel outputs as possible, we need to make sure that the state Φ_{m_1} is modified to be compatible with the (angles used to define the) unitaries V and U in Fig. 4. Since we can keep track of the intermediate angles deterministically, we can conditionally rotate subsystem 1 to be $\left|m_1\varphi_{00}^{\circledast}\right\rangle\left\langle m_1\varphi_{00}^{\circledast}\right|_1$ for $|jk\rangle\langle jk|_{45} = |00\rangle\langle 00|_{45}$. Note that in $\Psi_{\pm}$, when either of j or k is 1 (or both), $\varphi_{jk}^{\circledast} = \pi/2$ and hence $\left|\pm\varphi_{jk}^{\circledast}\right\rangle\left\langle\pm\varphi_{jk}^{\circledast}\right| = |\pm\rangle\langle\pm|$. Therefore, if $\hat{x}_1 \neq x_1$, then $\langle m_1|\pm\rangle = 0$ and the superposition in Φ_{m_1} collapses to just the term $j = k = 0$.

More precisely, we can implement the unitary operation

$$M_{m_1} \triangleq (K_{m_1})_1 \otimes |00\rangle\langle 00|_{45} + (I_2)_1 \otimes (I_4 - |00\rangle\langle 00|)_{45},$$

where K_+ and K_- are unitaries chosen to satisfy $K_+|+\rangle = \left|\varphi_{00}^{\circledast}\right\rangle$ and $K_-|-\rangle = \left|-\varphi_{00}^{\circledast}\right\rangle$. Hence, at stage (f) we have

$$\tilde{\Psi}_{m_1} = \sum_{j,k\in\{0,1\}^2} \frac{p_j p_k \left|\left\langle m_1| \pm \varphi_{jk}^{\circledast}\right\rangle\right|^2}{\text{Tr}\left[\Psi_{\pm}^{(1)}|m_1\rangle\langle m_1|\right]} \mathcal{T}(m_1\varphi_{jk}^{\circledast}, 0, 0, j, k). \tag{22}$$

Now we reverse the initial operations and arrive at stage (g). Assuming $\hat{x}_1 = x_1$ so that the superposition in $\tilde{\Psi}_{m_1}$ does not collapse, by comparing $\tilde{\Psi}_{m_1}$ ((f)) with $\Psi_\pm$ ((e)) we realize that the state at (g) cannot be exactly equal to the channel output, due to the additional factor involving m_1. We prove this in [13] by explicitly calculating the additional "error term" at (g) when compared to (the state at) (a). However, we show that when averaged over both cases $\hat{x}_1 = x_1$ and $\hat{x}_1 \neq x_1$, we obtain the following 5-qubit state for the system at (g):

$$\tilde{\rho}_{m_1,a} = |m_1\theta\rangle\langle m_1\theta|_1 \otimes [W \circledast W](\hat{x}_1)_{23} \otimes [W \circledast W](\hat{x}_1)_{45}, \tag{23}$$

where $P_{\text{succ},1}^{\text{BPQM}} \triangleq \text{Tr}\left[\Psi_+^{(1)}|+\rangle\langle+|\right] = \text{Tr}\left[\Psi_-^{(1)}|-\rangle\langle-|\right]$.

B. Decoding Bit x_2

At this point, we have decoded $\hat{x}_1 = 0$ if $m_1 = +$ and $\hat{x}_1 = 1$ if $m_1 = -$. We can absorb the value of $\hat{x}_1$ in the FG by updating the parity checks c_1 and c_2 to impose $x_2 \oplus x_3 = \hat{x}_1$ and $x_4 \oplus x_5 = \hat{x}_1$, respectively. Now we have two disjoint FGs after having removed x_1. It suffices to decode x_2 and x_4 since $\hat{x}_3 = \hat{x}_2 \oplus \hat{x}_1$ and $\hat{x}_5 = \hat{x}_4 \oplus \hat{x}_1$. Also, due to symmetry, it suffices to analyze the success probability of decoding x_2 (resp. x_4) and x_3 (resp. x_5). For this reduced FG, we need to split $\tilde{\rho}_{m_1,a}$ into two density matrices corresponding to the hypotheses $x_2 = 0$ and $x_2 = 1$. Recollect that for the hypotheses $\rho_1^{(0)}$ and $\rho_1^{(1)}$ for x_1, the 5-qubit state at the channel output was exactly their uniform superposition $\frac{1}{2}\rho_1^{(0)} + \frac{1}{2}\rho_1^{(1)}$. Hence, for x_2, we accordingly split $[W \circledast W](\hat{x}_1)_{23}$ in $\tilde{\rho}_{m_1,a}$:

$$\tilde{\Phi}_{x_2=\hat{x}_1}(\hat{x}_1) = |m_1\theta\rangle\langle m_1\theta|_2 \otimes |\theta\rangle\langle\theta|_3 \otimes [W \circledast W](\hat{x}_1)_{45},$$
$$\tilde{\Phi}_{x_2\neq\hat{x}_1}(\hat{x}_1) = |-m_1\theta, -\theta\rangle\langle -m_1\theta, -\theta|_{23} \otimes [W \circledast W](\hat{x}_1)_{45}$$

are the two hypotheses states. We can deterministically apply (Pauli) $Z^{\hat{x}_1}$ to system 2 in order to map these into

$$\Phi_\pm(\hat{x}_1) = |\pm\theta\rangle\langle\pm\theta|_2 \otimes |\pm\theta\rangle\langle\pm\theta|_3 \otimes [W \circledast W](\hat{x}_1)_{45},$$

where $\pm \equiv (-1)^{x_2 \oplus \hat{x}_1}$. Clearly, we can process systems 2 and 3 separately to decide x_2. Similarly, systems 4 and 5 can be processed separately to decide x_4. It is also clear that by performing the variable node operation $U_\circledast(\theta, \theta)$, we compress all the information into system 2, i.e., produce $|\pm\theta^\circledast\rangle\langle\pm\theta^\circledast|_2 \otimes |0\rangle\langle 0|_3$, which can be optimally distinguished by measuring in the X-basis. This agrees with the definition of node operations in BPQM as well because now the factor node $\tilde{c}_1$ has degree 2, and hence the optimal processing is to perform the variable node convolution between qubits 2 and 3. We can incorporate the operation $Z^{\hat{x}_1}$ into BPQM by performing $U_\circledast(m_1\theta, \theta)$ on systems 2 and 3 (and similarly on systems 4 and 5). Although $U_\circledast(m_1\theta, \theta) \neq U_\circledast(\theta, \theta) \cdot (Z^{\hat{x}_1} \otimes I_2)$, the two operations act identically on the states $\tilde{\Phi}_{x_2=\hat{x}_1}(\hat{x}_1)$ and $\tilde{\Phi}_{x_2\neq\hat{x}_1}(\hat{x}_1)$.

Based on this perspective, we can calculate

$$\mathbb{P}[\hat{x}_2 \neq x_2] = \frac{1}{2} - \frac{1}{4}\left\|\Phi_+(\hat{x}_1) - \Phi_-(\hat{x}_1)\right\|_1 \tag{24}$$
$$= \frac{1}{2} - \frac{1}{4} \cdot \sin\theta^\circledast \left\|X\right\|_1 = \frac{1 - \sin\theta^\circledast}{2}. \tag{25}$$

However, from simulations of BPQM we observe that this is slightly lower than the actual BPQM error rate for x_2.

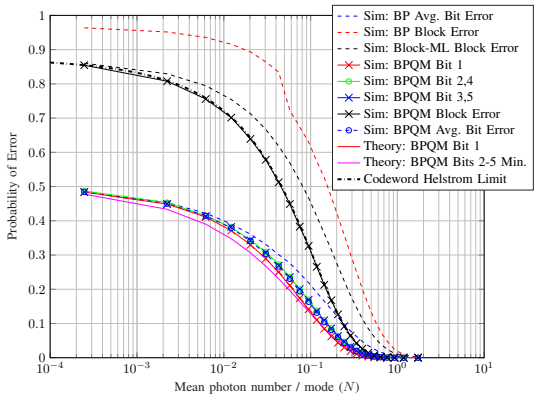

Fig. 5. The BPQM bit and block error probabilities, and the YKL limit, plotted against the mean photon number per mode N ($\cos\theta = e^{-2N}$ [4]).

C. Correct Analysis for Bit x_2

In order to understand the discrepancy, we start from the channel outputs and construct the hypotheses states $\rho_2^{(x_2)}$ for x_2 assuming we first decode x_2. Similar to $\rho_1^{(x_1)}$, $\rho_2^{(x_2)}$ is the uniform superposition of states corresponding to all codewords with the second bit taking the value x_2. Since we actually decode x_1 first, we track $\rho_2^{(x_2)}$ through the full BPQM circuit in Fig. 4. During this process, we show that the state $\tilde{\rho}_{2,m_1=+}^{(0)}$ we obtain at stage (g) (for, say, $\hat{x}_1 = 0$) is not exactly equal to $\tilde{\Phi}_{x_2=\hat{x}_1}(\hat{x}_1)$ but only two of the distinct entries differ (slightly) [13]. However, most importantly, we observe that

$$\frac{1}{2}\tilde{\rho}_{2,m_1=+}^{(0)} + \frac{1}{2}\tilde{\rho}_{2,m_1=+}^{(1)} = \frac{1}{2}\tilde{\Phi}_{x_2=0}(0) + \frac{1}{2}\tilde{\Phi}_{x_2=1}(0). \tag{26}$$

This explains that while the full density matrix $\tilde{\rho}_{m_1,a}$ was correct, we had split it incorrectly to arrive at the two hypotheses $\tilde{\Phi}_{x_2=\hat{x}_1}(\hat{x}_1)$ and $\tilde{\Phi}_{x_2\neq\hat{x}_1}(\hat{x}_1)$. Based on this, we verify that

$$P_{\text{succ},2}^{\text{BPQM}} = \text{Tr}\left[U_\circledast(m_1\theta, \theta)\tilde{\rho}_{2,m_1}^{(x_2)}U_\circledast(m_1\theta, \theta)^\dagger \cdot |\pm\rangle\langle\pm|_2\right]$$
$$= \frac{1}{2} + \frac{1}{4}\left\|L\left(\rho_2^{(0)} - \rho_2^{(1)}\right)L^\dagger\right\|_1 \tag{27}$$
$$= \frac{1}{2} + \frac{1}{4}\left\|\rho_2^{(0)} - \rho_2^{(1)}\right\|_1 = P_{\text{succ},2}^{\text{Hel}}, \tag{28}$$

where L is the sequence of operations in Fig. 4 from (a)-(g).

V. Simulations and Conclusion

We simulated the BPQM circuit in Fig. 4 and averaged the bit and block error rates of BPQM over 10^5 uniformly random codeword transmissions. These results are plotted in Fig. 5 against the mean photon number per mode [4]. We used the Yuen-Kennedy-Lax (YKL) conditions [11], [12] to numerically calculate the fundamental joint Helstrom limit, which is also shown in Fig. 5. We observe that BPQM appears to exactly achieve the YKL limit and hence be quantum optimal. The equalities relating $P_{\text{succ},2}^{\text{BPQM}}$ and $P_{\text{succ},2}^{\text{Hel}}$ are yet to be proven rigorously, although they were verified directly using the density matrices. It also remains to be shown analytically that the block error rate of BPQM meets the YKL limit exactly.

REFERENCES

[1] E. Conover, "Quantum computers are about to get real," *Science News*, vol. 191, no. 13, p. 28, 2017, https://www.sciencenews.org/article/quantum-computers-are-about-get-real.

[2] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019. [Online]. Available: http://www.nature.com/articles/s41586-019-1666-5

[3] E. Pednault, J. A. Gunnels, G. Nannicini, L. Horesh, and R. Wisnieff, "Leveraging Secondary Storage to Simulate Deep 54-qubit Sycamore Circuits," *arXiv preprint arXiv:1910.09534*, 2019. [Online]. Available: http://arxiv.org/abs/1910.09534

[4] S. Guha and M. M. Wilde, "Polar coding to achieve the Holevo capacity of a pure-loss optical channel," in *Proc. IEEE Int. Symp. Inform. Theory*, 2012, pp. 546–550. [Online]. Available: https://arxiv.org/abs/1202.0533

[5] M. M. Wilde and S. Guha, "Polar Codes for Classical-Quantum Channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 2, pp. 1175–1187, 2013. [Online]. Available: https://arxiv.org/abs/1109.2591

[6] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. New York, NY: Cambridge University Press, 2008.

[7] M. Mezard and A. Montanari, *Information, Physics, and Computation*. New York, NY: Oxford University Press, 2009.

[8] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, 2001. [Online]. Available: http://ieeexplore.ieee.org/document/910572/

[9] H. A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Mag.*, vol. 21, no. 1, pp. 28–41, 2004.

[10] J. M. Renes, "Belief propagation decoding of quantum channels by passing quantum messages," *New Journal of Physics*, vol. 19, no. 7, p. 072001, 2017. [Online]. Available: http://arxiv.org/abs/1607.04833

[11] H. Yuen, R. Kennedy, and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory," *IEEE Trans. Inform. Theory*, vol. 21, no. 2, pp. 125–134, 1975. [Online]. Available: http://ieeexplore.ieee.org/document/1055351/

[12] H. Krovi, S. Guha, Z. Dutton, and M. P. da Silva, "Optimal measurements for symmetric quantum states with applications to optical communication," *Phys. Rev. A*, vol. 92, no. 6, p. 062333, Dec 2015. [Online]. Available: http://arxiv.org/abs/1507.04737

[13] N. Rengaswamy, K. P. Seshadreesan, S. Guha, and H. D. Pfister, "Belief Propagation with Quantum Messages for Quantum-Enhanced Classical Communications," *arXiv preprint arXiv:2003.04356*, 2020. [Online]. Available: https://arxiv.org/abs/2003.04356

[14] M. M. Wilde, *Quantum Information Theory*. Cambridge University Press, 2013.

[15] C. W. Helstrom, "Quantum detection and estimation theory," *Journal of Statistical Physics*, vol. 1, no. 2, pp. 231–252, 1969.

[16] C. W. Helstrom, J. W. Liu, and J. P. Gordon, "Quantum-mechanical communication theory," *Proc. of the IEEE*, vol. 58, no. 10, pp. 1578–1598, 1970.

[17] S. Dolinar Jr., "An Optimum Receiver for the Binary Coherent State Quantum Channel," *MIT Res. Lab. Electron. Q. Prog. Rep.*, vol. 111, pp. 115–120, 1973. [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/56414/RLE{_}QPR{_}111{_}VII.pdf?sequence=1

[18] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.