Feature

Henry D. Pfister, Senior Member, IEEE | Duke University, Durham, NC 27708 USA | E-mail: henry.pfister@duke.edu
Christophe Piveteau | ETH Zürich, 8093 Zürich, Switzerland | E-mail: cpivetea@phys.ethz.ch
Joseph M. Renes, Member, IEEE | ETH Zürich, 8093 Zürich, Switzerland | E-mail: renes@phys.ethz.ch
Narayanan Rengaswamy, Member, IEEE | University of Arizona, Tucson, AZ 85721 USA | E-mail: narayananr@arizona.edu

Belief Propagation for Classical and Quantum Systems: Overview and Recent Results

Abstract—This article reviews belief propagation (BP) for classical inference problems and describes its extension to quantum systems, which is known as BP with quantum messages (BPQM). Since BP plays a key role in many low-complexity decoders for error-correcting codes, BPQM enables the practical extension of these decoders to classical quantum channels, such as the pure state channel.

Introduction

Belief propagation (BP) plays a key role in all known capacity-approaching coding schemes with low-complexity decoders. Introduced already in the 1960s, Gallager's low-density parity-check (LDPC) decoder [1] contains the essence of the BP algorithm, which later was formalized by Pearl in [2] and [3]. The discovery of turbo codes in 1993 [4] brought with it the turbo decoding algorithm, now recognized as an instance of BP [5]. In the late 1990s, these ideas led to formalization of factor graphs and the sum-product algorithm [6], [7]. More recently, Arikan's polar codes provide an elegant deterministic construction of capacity-achieving codes [8]. For polar codes, the low-complexity successive-cancellation decoder can also be seen as an instance of BP, albeit with a modified processing order that includes intermediate hard decisions.

Until recently, however, these amazing success stories from classical coding could not be translated into low-complexity capacity-achieving schemes for classical-quantum (CQ) channels. For example, in 2012 polar codes were shown to achieve the symmetric Holevo information rate for any CQ channel [9], but a low-complexity decoder for noncommuting outputs was explicitly mentioned as an interesting open problem. The missing element was a BP decoder for CQ channels.

In 2016, Renes defined BP with quantum messages (BPQM) as a quantum version of BP that applies to some inference problems on CQ channels [10]. In particular, Renes applied BPQM to binary linear codes defined by tree factor graphs that are transmitted over the CQ pure-state channel (PSC). The PSC is the simplest nontrivial CQ channel and can be seen as the quantum analog of the classical binary symmetric channel (BSC). The PSC is also a good model for a number of practical communication problems involving quantum states [11], [12]. In 2021, after some investigation and evolution, BPQM was shown to provide low-complexity minimum-error detection for this problem [13], [14]. As a result, the large body of practical work on LDPC and polar codes can now be applied to the PSC. Recently, BPQM has been extended to symmetric binary-input CQ channels, though the proposed approach is suboptimal in this case [15], [16].

The goal of this article is to provide a gentle introduction to BP for channels with classical inputs and quantum outputs (i.e., CQ channels). While we defer a precise discussion of this until after the review of quantum theory, the ith channel output in the CQ case is a quantum system whose state depends on the classical channel input x_i. These outputs should not be treated as *observations*, and there is no natural analog of a conditional distribution (say of X_i) given these outputs. Indeed, the 2022 Nobel prize in physics was awarded to Aspect, Clauser, and Zeilinger for experiments that definitively demonstrate the quantum-mechanical description of photons is incompatible with the possibility that all physical quantities associated with photons, e.g. polarization, always have definite values.

2692-4080 © 2023 IEEE
Digital Object Identifier 10.1109/MBITS.2023.3285848
Date of publication 13 June 2023; date of current version 23 August 2023.

Instead, one must extract classical information from the channel output via measurement in order to have any knowledge of X_i beyond the prior. Moreover, the receiver cannot determine the exact quantum state of these systems with certainty, as measurements are stochastic and fundamentally disturb the state. For that reason, we will discuss generalizations of BP that implement a collective measurement on the entire collection of channel outputs in order to estimate X_i (or the entire vector $X_1, \ldots, X_N$) with a minimal error probability. Belief-propagation with quantum messages (BPQM) [10], [13], [14], [15], [16] is the name given to a family of quantum algorithms (i.e., which are executed on a quantum computer) that can implement this measurement efficiently in some cases.

Notation

We denote the natural numbers by $\mathbb{N} = \{1, 2, \ldots\}$ and use the shorthand $[n] := \{1, \ldots, n\}$ for $n \in \mathbb{N}$. The ring of integers modulo-2 is denoted by $\mathbb{Z}_2$. Vectors are written in bold, $\boldsymbol{x} = (x_1, \ldots, x_N)$. For a subset $A = \{a_1, a_2, \ldots, a_{|A|}\} \subseteq [N]$ with $a_1 < a_2 < \cdots < a_{|A|}$, we define the subvector $\boldsymbol{x}_A = (x_{a_1}, x_{a_2}, \ldots, x_{a_{|A|}})$. The Kronecker delta function is denoted by δ_{ij}. Random variables are written as capital letters. Joint and conditional probability mass functions (pmfs) are denoted by P with a subscript indicating the order and conditioning (e.g., $P_{X_2,X_3|X_1}(x_2, x_3|x_1)$ equals the conditional probability that $X_2 = x_2$ and $X_3 = x_3$ given that $X_1 = x_1$).

Belief Propagation

Introduction

The belief propagation (BP) algorithm was introduced by Pearl [2], [3] in 1982 as an efficient inference algorithm for Bayesian belief networks. While it is important to note that similar ideas and special cases (e.g., [17], [18], [19]) were proposed earlier in other fields, this work is important because it formalized a general version of the idea and popularized it among computer scientists under the moniker *belief propagation* [6], [20].

From a contemporary viewpoint, the core idea is that a set of random variables (say $X_1, \ldots, X_N$) can be associated with a graph whose vertices represent the random variables and whose edges encode the conditional independence structure of the random variables. In particular, if a subset of variables forms a boundary that cuts the graph into two parts, then the random variables in the two parts are conditionally independent given the random variables on the boundary. If the associated graph is a tree, then the BP algorithm computes the exact marginal probability of each X_i by using local updating rules that pass messages along the edges. There are a few variations of the precise formulation (e.g., Markov random fields, Bayesian belief networks, factor graphs, and tensor networks) whose details differ but which are all equivalent when the graph is a tree [6], [19], [20], [21].

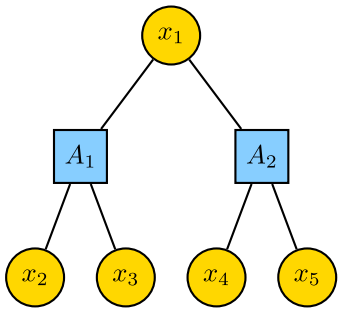

Figure 1
Factor graph associated with (1).

Example 1. Consider the case where $N = 5$, $\mathcal{X} = \{0, 1\}$, and the joint probability mass function (pmf) of $(X_1, \ldots, X_5) \in \mathcal{X}^5$ satisfies

$$P_{X_1,\ldots,X_5}(x_1, x_2, x_3, x_4, x_5) = P_{X_1}(x_1) P_{X_2,X_3|X_1}(x_2, x_3|x_1) P_{X_4,X_5|X_1}(x_4, x_5|x_1). \quad (1)$$

The factor graph associated with this representation is shown in Figure 1. Since the last term in (1) does not depend on x_2 or x_3, it follows that X_4, X_5 are conditionally independent of X_2, X_3 given X_1. This property is represented in the factor graph by the fact that the vertex labeled x_1 forms a boundary that cuts the graph into two parts, where the left part contains vertices x_2, x_3 and the right part contains vertices x_4, x_5.

More formally, a factor graph is defined as follows. Let $\mathcal{X}$ be a finite set, $[N] := \{1, 2, \ldots, N\}$, and $\mathcal{A}$ be a collection of subsets of $[N]$. A *factor graph* is a bipartite graph that represents a function $f : \mathcal{X}^N \to \mathbb{R}$ in the factored form

$$f(\boldsymbol{x}) := \prod_{A \in \mathcal{A}} f_A(\boldsymbol{x}_A) \quad (2)$$

where for $A \subseteq [N]$, we let $f_A : \mathcal{X}^{|A|} \to \mathbb{R}$ denote a function of $\boldsymbol{x}_A$ (i.e., that depends only on variables indexed by A). The factor graph has a *variable node* labeled $i \in [N]$ that is associated with the variable x_i and a *factor node* labeled $A \in \mathcal{A}$ that is associated with the subset $A \subseteq [N]$. The edges of the graph are formed by connecting each factor node to the variables in its associated subset A. In other words, there is an edge between variable node i and factor node A if and only if $i \in A$.

This formalism is now used to connect Example 1 with the graph in Figure 1. To do this, we choose $\mathcal{X} = \{0, 1\}$,

$N = 5$, and $\mathcal{A} = \{\{1\}, \{1,2,3\}, \{1,4,5\}\}$. Then, we choose

$$f_1(x_1) = P_{X_1}(x_1)$$
$$f_{123}(x_1, x_2, x_3) = P_{X_2,X_3|X_1}(x_2, x_3|x_1)$$
$$f_{145}(x_1, x_4, x_5) = P_{X_4,X_5|X_1}(x_4, x_5|x_1).$$

Using these choices, we find that (2) implies $f(\boldsymbol{x}) = P_{\boldsymbol{X}}(\boldsymbol{x})$ for the model defined by (1).

While the random variables $\boldsymbol{X} = (X_1, \ldots, X_N)$ define the true system state, one typically has access only to noisy outputs $\boldsymbol{Y} = (Y_1, \ldots, Y_N) \in \mathcal{Y}^N$ from a memoryless channel described by a conditional pmf,

$$P_{Y_1,\ldots,Y_N|X_1,\ldots,X_N}(y_1, \ldots, y_N|x_1, \ldots, x_N) = Q(y_1, \ldots, y_N|x_1, \ldots, x_N) := \prod_{i=1}^{N} Q_i(y_i|x_i) \quad (3)$$

where $Q_i(y|x)$ is the conditional probability of output $y \in \mathcal{Y}$ given input $x \in \mathcal{X}$ for the ith channel. These additional channel outputs could simply be added to the original set of random variables without affecting any key properties (e.g., they cannot introduce cycles in the factor graph), but that approach is typically avoided due to the notational burden and because the channel outputs are treated as fixed observed values $\boldsymbol{y} = (y_1, \ldots, y_N)$. Then, BP can be used to compute marginals of the conditional distribution $P_{X_1,\ldots,X_N|Y_1,\ldots,Y_N}(x_1, \ldots, x_n|y_1, \ldots, y_N)$. In this work, we assume that $\mathcal{Y}$ is a finite set for simplicity, but most of the statements hold under more general conditions.

Channel Combining Perspective

In this section, we describe the channel combining perspective for classical BP decoding, which is the natural bridge to the BPQM decoding algorithm.

Restricting attention to cases (e.g., Example 1) where the factor graph is a tree rooted at x_1, the channel combining perspective allows one to easily construct the effective channel from the root x_1 of the tree to all the observations in the tree $\boldsymbol{y}$ as described by the conditional pmf $P_{\boldsymbol{Y}|X_1}$. The construction proceeds recursively from the leaves to the root, and the channel of any given node in the factor graph is formed by combining the channels of its children and any local observations. The initial channels at the leaf nodes are the physical noisy channels themselves. This procedure is equivalent to what is known as density evolution analysis on a tree [22], [23].

Given the effective channel $P_{\boldsymbol{Y}|X_1}$, the most likely value of x_1 can be computed, in principle, from the observed channel output $\boldsymbol{y}$. However, the number of possible values taken by $\boldsymbol{y}$ grows very rapidly with the depth of the tree and directly computing the most likely x_1 can be cumbersome. For the case of classical binary-input symmetric channels, this difficulty can be avoided because any channel can be represented as a weighted mixture of BSCs with different error rates [22], [23]. This allows for efficient representation of $P_{\boldsymbol{Y}|X_1}$ even for very large depth.

In the case of CQ channels, the effective channel can also be constructed via channel combining. The task of determining the most likely x_1 is then to make the appropriate measurement on the collection of all quantum outputs. However, this is potentially a quite cumbersome task. As with the BSC, for the case of CQ channels with outputs described by pure quantum states, a more efficient representation of the combined channels is possible. This in turn makes it possible to efficiently implement the desired measurement to determine x_1. The "Channel Combining for PSCs" section presents this in detail.

Channel Combining Details

We now describe the channel combining steps for tree factor graphs. In this case, for any factor node $A \in \mathcal{A}$, we define its parent variable node $x_{\mathrm{pa}(A)}$ to be the unique neighbor of A, which is closest to the root so that $\mathrm{pa}(A)$ is the index of the parent node. We also define the index set of its children, $\mathrm{ch}(A) = A \setminus \mathrm{pa}(A)$, to contain the indices of all the other adjacent variable nodes. We will also assume that the observer does not receive $\boldsymbol{X}$ but only the noisy channel outputs $\boldsymbol{Y}$ defined by (3). Thus, we define the joint function

$$f(\boldsymbol{x}, \boldsymbol{y}) := f(x_1, \ldots, x_N)Q(y_1, \ldots, y_N|x_1 \ldots, x_N)$$

and observe that, under our assumptions, this equals the joint probability $P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})$.

The channel-combining perspective starts by viewing the entire factor graph as a channel with input x_1, output $\boldsymbol{y} = (y_1, \ldots, y_N)$, and conditional output probability given by

$$P_{\boldsymbol{Y}|X_1}(\boldsymbol{y}|x_1) = \sum_{(x_2,\ldots,x_N)\in\mathcal{X}^{N-1}} P_{\boldsymbol{Y},X_2,\ldots,X_N|X_1}(\boldsymbol{y}, x_2, \ldots, x_N|x_1). \quad (4)$$

The key observation is that the same idea can be applied to any subtree. In particular, for the vertex labeled $i \in [N]$ which is associated with x_i, consider the subtree generated by keeping this vertex and all nodes and edges below it in the tree.

Let the set S_i contain the labels of all variable nodes in the subtree rooted at i (i.e., the labels of x_i and all variable below x_i in the tree). The channel for this subtree has input x_i, output $\boldsymbol{y}_{S_i}$, and conditional output probability

$$W_i(\boldsymbol{y}_{S_i}|x_i) := Q_i(y_i|x_i) \prod_{A:\mathrm{pa}(A)=i} W_A(\boldsymbol{y}_{A^*}|x_i) \quad (5)$$

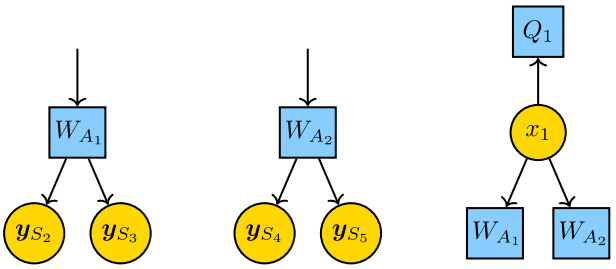

Figure 2

Effective channels defined by the channel combining for the factor graph in Figure 1.

where $W_A(\boldsymbol{y}_{A^*}|x_{\mathrm{pa}(A)})$ is the channel defined for the subtree rooted at factor node $A \in \mathcal{A}$ and the set $A^* = S_{\mathrm{pa}(A)} \setminus \mathrm{pa}(A)$ contains the indices of all variable nodes below x_i in the tree.

Equation (5) highlights a key property of tree factor graphs. A degree-j variable node x_i can be seen as the input to $j+1$ conditionally independent channels. The first is a direct channel with input x_i and conditional output probability $Q_i(y_i|x_i)$. The other j are subtree channels associated with the j child factor nodes.

For a child factor node $A \in \mathcal{A}$, the subtree channel is denoted by $W_A(\boldsymbol{y}_{A^*}|x_{\mathrm{pa}(A)})$, where the input has index $\mathrm{pa}(A)$ because the parent variable in A is the input to the combined channel. The input to this channel is $x_{\mathrm{pa}(A)}$ and the output is $\boldsymbol{y}_{A^*}$. The conditional probability of the channel output is

$$W_A(\boldsymbol{y}_{A^*}|x_{\mathrm{pa}(A)}) = Z_A^{-1} \sum_{\boldsymbol{x}_{\mathrm{ch}(A)} \in \mathcal{X}^{|\mathrm{ch}(A)|}} f_A(\boldsymbol{x}_A) \prod_{j \in \mathrm{ch}(A)} W_j(\boldsymbol{y}_{S_j}|x_j) \quad (6)$$

where Z_A is the normalization constant chosen so that the sum over $\boldsymbol{y}_{A^*}$ equals 1. This recursive definition terminates when all children of A are leaves because, if x_j is a leaf node, then $S_j = \{j\}$ and $W_j(\boldsymbol{y}_{S_j}|x_j)$ simplifies to $Q_j(y_j|x_j)$.

The operation in (6) is known as factor node combining and it represents the case where the symbol $x_{\mathrm{pa}(A)}$ is used to conditionally generate the values of the child variable nodes $x_{\mathrm{ch}(A)}$ and then each of these values are transmitted independently through subchannels $W_j(\cdot\,|\,x_j)$ for $j \in \mathrm{ch}(A)$. The two leftmost graphs in Figure 2 illustrate factor node combining for A_1 and A_2 in the example. In that case, BP decoding of x_1 is performed by first using (6) to compute W_{A_1} and W_{A_2} for the observed $\boldsymbol{y}$ values and all $x_1 \in \mathcal{X}$.

The operation in (5) is known as variable node combining and it represents the case where the symbol x_i is transmitted independently through multiple channels (i.e., Q_i and all W_A with $\mathrm{pa}(A) = i$). The variable node combining operation for x_1 is depicted by the rightmost graph in Figure 2. This operation combines the values of W_{A_1} and W_{A_2} computed earlier with the observation of x_1 through Q_1. Thus, the value of (4) can be computed by applying channel combining operations like this starting from the leaves. Then, using Bayes' rule, one can compute $P_{X_1|\boldsymbol{Y}}$ and this is equivalent to performing BP decoding on the tree.

If the received values are known (e.g., when decoding observed channel outputs), then the inputs and outputs of the recursive steps are vectors in $\mathbb{R}^{|\mathcal{X}|}$ that specify the relevant conditional probabilities of the actual observations for each possible channel input. This is equivalent to standard BP decoding on tree factor graphs.

If the received values have not yet been observed (e.g., when analyzing the effective channels themselves), then the inputs and output of this recursion are transition probability matrices for finite output channels with $\mathcal{X}$ inputs. The recursion relies implicitly on the fact that, for any finite set of channels with input alphabet $\mathcal{X}$ and a finite output alphabet, the new channel formed by combining also has input alphabet $\mathcal{X}$ and a finite output alphabet. As noted earlier, this is equivalent to what is known as density evolution analysis on a tree [22], [23].

Finally, the channel combining perspective does not extend naturally to factor graphs with cycles whereas the standard BP algorithm is well defined. However, in that case, BP does not compute the marginals exactly [24].

Symmetric Binary-Input Channels

BP and density evolution simplify considerably for the case of symmetric binary-input channels. The simplest channel of this type is the binary symmetric channel (BSC), which is defined by $\mathcal{Y} = \mathcal{X} = \{0,1\}$ and a parameter $p \in [0,(1/2)]$ that specifies the probability the output is not equal to the input (i.e., the transition probability is $V_p(y|x) = (1/2) + (-1)^{x+y}((1/2) - p)$).

For a random input $X \in \mathcal{X}$, the random output $Y \in \mathcal{Y}$ of any symmetric binary-input channel can be mapped into a sufficient statistic $(\hat{X}, R) \in \mathcal{X} \times [0,(1/2)]$, where $\hat{X} = \hat{x}(Y)$ is the maximum-likelihood (ML) estimate of X given Y and $R = r(Y) := \Pr(X \neq \hat{x}(Y))$ is the probability the ML estimate is wrong. The channel is symmetric if R is independent of the channel input X. Using this, any symmetric binary-input channel $Q(y|x)$ is equivalent to a mixture of BSCs and its conditional pmf can be written as

$$Q(y|x) = P_R(r(y)) V_{r(y)}(\hat{x}(y)|x).$$

Variable Node Combining

Using this result, it is sufficient to define the variable node combining rule for two BSCs and then extend it to two

general symmetric binary-input channels by considering all pairs of BSCs they can generate. Thus, we first consider two BSCs, with outputs $\hat{x}_1, \hat{x}_2$ and error probabilities r_1, r_2. The ML estimate of the combined channel is denoted by $\hat{x}$ and its error probability denoted by $r(\hat{x}_1, \hat{x}_2)$. From (5), we see that the combined channel will have conditional pmf

$$W(\hat{x}_1, \hat{x}_2|x) = V_{r_1}(\hat{x}_1|x)V_{r_2}(\hat{x}_2|x).$$

This implies that the ML estimate from the BSC with the smaller error probability will determine the ML estimate for the combined channel and we have

$$\hat{x} = \begin{cases} \hat{x}_1 & \text{if } r_1 \leq r_2 \\ \hat{x}_2 & \text{otherwise.} \end{cases}$$

Once the ML estimate has been determined, one can compute the error probability for the combined channel using the BSC error probabilities. The two channel outputs agree (i.e., $\hat{x}_1 = \hat{x}_2$) with probability $r_1 r_2 + (1 - r_1)(1 - r_2)$ and, in this case, the error probability of the combined channel is

$$r(\hat{x}_1, \hat{x}_2) = \frac{r_1 r_2}{r_1 r_2 + (1 - r_1)(1 - r_2)}$$

because it makes an error only if both BSCs make errors. The two channel outputs disagree (i.e., $\hat{x}_1 \neq \hat{x}_2$) with probability $r_1(1 - r_2) + (1 - r_1)r_2$ and, in this case, the error probability of the combined channel is

$$r(\hat{x}_1, \hat{x}_2) = \frac{\min\{r_1(1 - r_2), (1 - r_1)r_2\}}{r_1(1 - r_2) + (1 - r_1)r_2}$$

because it makes an error only if the more reliable BSC makes an error and the less reliable BSC does not.

Check Node Combining

Check node combining is a specific example of factor node combining used for the decoding of binary linear codes [10], [23]. It corresponds to the case where $\mathcal{X} = \{0, 1\}$ and the factor node (say A) is defined by the even-parity function f_A, where $f_A(x_1, \ldots, x_k) = 1$ if $(x_1, \ldots, x_k)$ contains an even number of ones and f_A equals 0 otherwise.

Here, we will focus on the case where $k = 3$ with $A = \{1, 2, 3\}$ and $\mathrm{pa}(A) = 1$. Using (6), we see that the combined channel takes the input x_1, draws the child node values x_2, x_3 according to

$$P_{X_2,X_3|X_1}(x_2, x_3|x_1) = \frac{f_A(x_1, x_2, x_3)}{\sum_{(x_2', x_3') \in \mathcal{X}^2} f_A(x_1, x_2', x_3')}$$

and then transmits those values through the two channels. Like the variable node combining case, we will assume the two channels are BSCs because symmetric binary-input channels are equivalent to mixtures of BSCs.

Since $f_A(x_1, x_2, x_3)$ equals 1 for even-weight inputs and 0 otherwise, this is equivalent to picking X_3 to be a uniform random bit and then computing $X_2 = x_1 \oplus X_3$ to satisfy the overall parity constraint. Then, the bits x_2 and x_3 are transmitted through BSCs with error probabilities r_1 and r_2. The channel outputs are denoted as $\hat{x}_2$ and $\hat{x}_3$. The goal of check node combining is to represent all the information about x_1 contained in $\hat{x}_2$ and $\hat{x}_3$ as a channel from x_1 to a new output y. Since $x_1 = x_2 \oplus x_3$, it turns out that the ML estimate of x_1 is given by $\hat{x}_2 \oplus \hat{x}_3$ and the ML estimate is wrong if and only if exactly one of $\hat{x}_2$ and $\hat{x}_3$ is wrong. Thus, the combined channel is a BSC with output $y = \hat{x}_2 \oplus \hat{x}_3$ and error probability $r_1(1 - r_2) + r_2(1 - r_1)$.

Introduction to Quantum Theory

This article is targeted at researchers with little or no background in quantum mechanics, but it does assume some familiarity with tensor linear algebra. Therefore, we start with a brief review of the necessary quantum theory. It is worth noting that, while quantum mechanics is a physical theory, this article treats quantum theory from a purely mathematical perspective where each operation is defined without reference to how it might be implemented physically.

Quantum States and Dirac Notation

An isolated quantum system with d perfectly distinguishable states can be represented by a "pure state" vector ψ in the Hilbert space $\mathbb{C}^d$ with unit length. An important special case is where $d = 2$ because the basic unit of quantum information (based on two-level systems) is a *quantum bit* or, simply, a *qubit*. A qubit that is in a deterministic quantum state is called *pure state* and represented mathematically by a unit vector in $\mathbb{C}^2$. In this section, we try to provide most of the background required for this work. For a more complete picture, see [25], [26], [27].

Let $d = 2^n$ for some integer $n \geq 1$ and recall that a complex length-d vector $\psi \in \mathbb{C}^d$ can be expressed as a linear combination of standard basis vectors. For the standard basis vector e_v, all elements are zero except the entry indexed by v which equals 1. For $n = 1$, the two basis vectors e_0 and e_1 are denoted by $|0\rangle$ and $|1\rangle$, respectively, using the Dirac's "braket" notation. These are to be read as "ket 0" and "ket 1," respectively, and they are length $d = 2$ column vectors.

Conjugate transposes are denoted by $\langle 0| := |0\rangle^\dagger$ and $\langle 1| := |1\rangle^\dagger$, respectively, which are to be read as "bra 0" and "bra 1." This naming was chosen so that the inner product $e_i^\dagger e_j = \langle i|j\rangle = \delta_{ij}$, where $i, j \in \{0, 1\}$, appears like a bracket ("braket"). Therefore, any length $d = 2$ complex vector ψ can be written as "ket psi" $|\psi\rangle$, where

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle = \alpha_0 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \alpha_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \in \mathbb{C}^2.$$

Mathematically, such a qubit ($n = 1$) pure state is simply a unit vector in $\mathbb{C}^2$, which means it can be represented as $|\psi\rangle$ as above with the additional constraint that $|\alpha_0|^2 + |\alpha_1|^2 = 1$.

For $n \geq 1$ qubits, the standard basis vectors are defined by $e_v = e_{v_1} \otimes e_{v_2} \otimes \cdots \otimes e_{v_n}$, where $v = (v_1, \ldots, v_n) \in \mathbb{Z}_2^n$ and $\otimes$ denotes the Kronecker product. In Dirac notation, we write $|v\rangle = e_v = |v_1\rangle \otimes |v_2\rangle \otimes \cdots \otimes |v_n\rangle$. Thus, a general n-qubit pure state is represented as

$$|\psi\rangle = \sum_{v \in \mathbb{Z}_2^n} \alpha_v |v\rangle \tag{7}$$

where

$$\||\psi\rangle\|^2 := \langle\psi|\psi\rangle = \sum_{v \in \mathbb{Z}_2^n} |\alpha_v|^2 = 1. \tag{8}$$

This set of standard basis vectors is called the *computational basis* of a quantum system. If $|\phi\rangle = \sum_{v \in \mathbb{Z}_2^n} \beta_v |v\rangle$ is another pure state, then its inner product with $|\psi\rangle$ is given by

$$\langle\phi|\psi\rangle = \sum_{v \in \mathbb{Z}_2^n} \beta_v^* \alpha_v = \langle\psi|\phi\rangle^\dagger$$

where β_v^* is the complex conjugate of $\beta_v \in \mathbb{C}$. This inner product is called the *overlap* between the two states.

All reversible operations on a quantum system can be represented by unitary evolution. In particular, if $|\psi\rangle \in \mathbb{C}^d$ and $U \in \mathbb{C}^{d \times d}$ is unitary matrix, then applying U to $|\psi\rangle$ to get $|\psi'\rangle = U|\psi\rangle$ is assumed to be a physically realizable quantum operation.

Consider two independent (i.e., noninteracting) quantum systems in pure states $|\psi\rangle \in \mathbb{C}^{d_1}$ and $|\phi\rangle \in \mathbb{C}^{d_2}$. Using the Kronecker product, these two systems can be combined into a common Hilbert space of dimension $d_1 d_2$. The combined system will be the pure state $|\psi\rangle \otimes |\phi\rangle \in \mathbb{C}^{d_1 d_2}$. After combining, one can apply quantum operations that act jointly on the two systems. We also note that, for $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^{d_1}$ and $|\phi_1\rangle, |\phi_2\rangle \in \mathbb{C}^{d_2}$, the implied inner product on $\mathbb{C}^{d_1 d_2}$ satisfies

$$\begin{aligned}&(|\psi_1\rangle \otimes |\phi_1\rangle)^\dagger (|\psi_2\rangle \otimes |\phi_2\rangle) \\ &\quad = (\langle\psi_1| \otimes \langle\phi_1|)(|\psi_2\rangle \otimes |\phi_2\rangle) = \langle\psi_1|\psi_2\rangle\langle\phi_1|\phi_2\rangle. \end{aligned} \tag{9}$$

From this, one can interpret (7) as joining the Hilbert spaces of n independent qubits so that one can operate on them jointly.

Measurement of Pure States

The only way to obtain classical information about a quantum system is *measurement*. The simplest type of measurement is called a *von Neumann* (or projective) measurement and it is the only type we consider in this article. A von Neumann measurement is defined by a set of orthogonal projection matrices $\{\Pi_i\}_{i \in [M]}$, where $\Pi_i \in \mathbb{C}^{d \times d}$, such that

$$\Pi_i \Pi_j = \delta_{ij} \Pi_i, \qquad \sum_{i=1}^{M} \Pi_i = I_d,$$

where δ_{ij} is the Kronecker delta and I_d is the $d \times d$ identity matrix.

To explain the effect of this measurement on a pure state $|\psi\rangle$, we describe how to simulate the measurement process on a classical computer. One starts with the implied orthogonal decomposition

$$|\psi\rangle = \sum_{i=1}^{M} \underbrace{\Pi_i |\psi\rangle}_{|\psi_i\rangle}$$

where $p_i = \|\Pi_i|\psi\rangle\|^2 = \langle\psi|\Pi_i|\psi\rangle$ is the squared Euclidean norm of the ith component. From the definition of the measurement projectors, it is evident that

$$\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} \langle\psi|\Pi_i|\psi\rangle = \langle\psi| \left(\sum_{i=1}^{M} \Pi_i \right) |\psi\rangle = 1$$

and $\langle\psi|\Pi_i|\psi\rangle = \langle\psi_i|\psi_i\rangle = p_i$. Thus, we can treat $\{p_i\}_{i \in [M]}$ as a probability distribution on $[M]$ and draw a sample j from it. The value j is called the *outcome* of the measurement. By the *Born rule* of quantum mechanics, the postmeasurement state $|\psi'\rangle$ equals the normalized projection of $|\psi\rangle$ onto the range of Π_j and we get

$$|\psi'\rangle = \frac{|\psi_j\rangle}{\||\psi_j\rangle\|} = \frac{\Pi_j|\psi\rangle}{\sqrt{p_j}}.$$

A special case of the von Neumann measurement is the scenario where all the projectors have rank 1, i.e., $\Pi_i = |\phi_i\rangle\langle\phi_i|$ for a set of orthogonal quantum states $\{|\phi_i\rangle\}_{i \in [M]}$ and $M = d$.

Note that the change in the original state due to measurement is a distinguishing feature of quantum systems. For classical systems, it is assumed that one can read the value of a register without any possibility of disturbing its value. In reality, however, this assumption is based on the fact that the probability of disturbing the value is negligible.

Now, we introduce notation for a few special quantum states. The plus and minus states are defined by

$$|+\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \quad \text{and} \quad |-\rangle := \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

Also, for $\theta \in [0, \pi/2]$ and $x \in \mathbb{Z}_2$, we define the parameterized qubit pure state

$$|\psi_x(\theta)\rangle := \cos\left(\frac{\theta}{2}\right)|0\rangle + (-1)^x \sin\left(\frac{\theta}{2}\right)|1\rangle$$

and note that θ will be dropped when it is clear from the context. See Figure 3 for a diagram of the PSC.

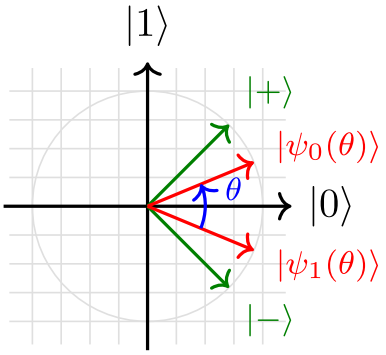

Figure 3

Diagram of the $|0\rangle, |1\rangle$ plane showing the output vectors for the PSC (with $\theta = \pi/4$) along with projectors from Example 2.

Definition 1. *The* pure-state channel *(PSC) is a binary-input CQ channel with parameter $\theta \in [0, \pi/2]$ whose output is a qubit in a deterministic pure state. If the transmitter sends 0, then the receiver's qubit is in the pure state $|\psi_0(\theta)\rangle$ and if the transmitter sends 1, then the receiver's qubit is in the pure state $|\psi_1(\theta)\rangle$. This channel can also be parameterized by its overlap*

$$\langle\psi_0(\theta)|\psi_1(\theta)\rangle = \cos^2\left(\frac{\theta}{2}\right) - \sin^2\left(\frac{\theta}{2}\right) = \cos(\theta). \quad (10)$$

At this point, we can already appreciate the challenge that the laws of quantum mechanics impose when transmitting classical information over a CQ channel such as the PSC. While the state of the output qubit is fully deterministic, it is impossible for the receiver to learn with certainty whether the state is $|\psi_0(\theta)\rangle$ or $|\psi_1(\theta)\rangle$, since measurements necessarily perturb the system. The following example describes a strategy for the receiver to guess the state of the PSC's qubit output with the highest success probability, assuming an equiprobable input.

Example 2. To distinguish between $|\psi_0\rangle$ and $|\psi_1\rangle$, one can use the 2×2 projection matrices

$$\Pi_0 = |+\rangle\langle+| = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \Pi_1 = |-\rangle\langle-| = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Using these, the conditional probability of outcome y given the channel input x is given by

$$q_{x,y} = \|\Pi_y \psi_x\|^2 = \frac{1}{2}\left(\cos\frac{\theta}{2} + (-1)^{x+y}\sin\frac{\theta}{2}\right)^2 = \frac{1}{2} + (-1)^{x+y}\frac{\sin(\theta)}{2}.$$

This measurement minimizes the error probability when the channel inputs are equally likely.

Perhaps the most straightforward strategy for transmitting information over a collection of PSCs would be to perform individual optimal measurements (as in Example 2) on each channel output, effectively converting each PSC to a BSC. While this approach is simpler, its performance is degraded because the classical capacity of the resulting BSC is strictly lower than the quantum (Holevo) capacity of the original PSC if $\theta \in (0, \pi/2)$. In this case, a truly quantum decoder, which performs quantum information processing on the channel outputs, is required to achieve optimal performance.

Mixed Quantum States

A quantum system can also be in a *mixed state*, which is equivalent to a classical random mixture of pure states. In particular, it can be in one of several distinct states $\{|\psi_j\rangle\}_{j\in[J]}$ with associated probabilities p_j. For such a "bag of states" model, a succinct description of the quantum state is given by its *density matrix* $\rho \in \mathbb{C}^{d\times d}$, which is defined by

$$\rho := \sum_{j=1}^{J} p_j |\psi_j\rangle\langle\psi_j|. \quad (11)$$

It is easy to verify that this matrix is positive semidefinite and has trace 1.

Let $\mathcal{D}(d)$ denote the set of $d \times d$ positive semidefinite matrices with unit trace. Any $\rho \in \mathcal{D}(d)$ is a valid density matrix for some quantum system with d distinguishable states. While the density matrix of a "bag of states" model is unique, the interpretation of a density matrix as a mixture of pure states is not necessarily unique. Moreover, it is not possible to distinguish between mixed states whose density matrices are equal even if they were generated by different "bag of states" models. We note that it is typical in quantum to refer to mixed state simply as a state and instead use the modifier "pure state" to distinguish between the two cases.

When a system in a mixed state $\rho \in \mathcal{D}(d)$ undergoes a reversible evolution according to the unitary $U \in \mathbb{C}^{d\times d}$, its state evolves to $U\rho U^H$. This can be understood by observing that, for all $j \in [J]$, the state $|\psi_j\rangle$ in our bag of states [e.g., see (2)] separately evolves to $U|\psi_j\rangle$.

Consider the effect of a von Neumann measurement $\{\Pi_i\}$ on a system described by the state $\rho = \sum_{j=1}^{J} p_j |\psi_j\rangle\langle\psi_j|$. Let $q_{j,i} = \langle\psi_j|\Pi_i|\psi_j\rangle$ denote the conditional probability of measurement outcome i given that the system is in pure state $|\psi_j\rangle$. If the outcome is i, then the overall

postmeasurement state is

$$\rho_i = \sum_{j=1}^{J} \frac{p_j q_{j,i}}{p_i} \left|\psi_j^i\right\rangle \left\langle\psi_j^i\right| = \frac{\Pi_i \rho \Pi_i}{p_i} \tag{12}$$

where $p_j q_{j,i}/p_i$ equals the probability that the system is in pure state $|\psi_j\rangle$ given measurement outcome i and the associated pure state is

$$\left|\psi_j^i\right\rangle = \frac{\Pi_i |\psi_j\rangle}{\sqrt{q_{j,i}}}. \tag{13}$$

Computing the trace on both sides of (12) and using the cyclic property of the trace shows that

$$p_i = \sum_{j'=1}^{J} p_{j'} q_{j',i} = \mathrm{Tr}[\Pi_i \rho]. \tag{14}$$

Therefore, the density matrix encodes all the necessary information about the system in order to track its evolution through arbitrary quantum processes, i.e., unitary operations and measurements.

Now, we can define a CQ channel to have a classical input and quantum output.

Definition 2. *A* CQ channel W *is defined by an input alphabet* $\mathcal{X}$*, an output dimension* d*, and a mapping* $W : \mathcal{X} \to \mathcal{D}(d)$ *that specifies the mixed state channel output for each input.*

Definition 3. *Two CQ channels* $W_1 : \mathcal{X} \to \mathcal{D}(d)$ *and* $W_2 : \mathcal{X} \to \mathcal{D}(d)$ *are* unitarily equivalent *if there exists a unitary matrix* $U \in \mathbb{C}^{d \times d}$ *such that* $UW_1(x)U^H = W_2(x)$ *for all* $x \in \mathcal{X}$*.*

If the channel W_1 is unitarily equivalent to W_2, then we can effectively realize W_2 by postprocessing the output of W_1 with the appropriate unitary transformation (and vice versa). For this reason, two unitarily equivalent channels are identical from an information theoretic perspective.

Combining Quantum Systems

Consider two independent (i.e., noninteracting) quantum systems in mixed states $\rho_1 \in \mathcal{D}(d_1)$ and $\rho_2 \in \mathcal{D}(d_2)$. These two systems can be combined into a common Hilbert space of dimension $d_1 d_2$. The combined system will have the density matrix $\rho = \rho_1 \otimes \rho_2 \in \mathcal{D}(d_1 d_2)$ whose "marginalization" (i.e., partial trace) down to either component system equals the original density matrix for that system. In the combined space, one can apply operations that act jointly on the two systems.

Similarly, consider two independent CQ channels W_1, W_2 with input alphabets $\mathcal{X}_1, \mathcal{X}_2$ and output dimensions d_1, d_2. These can be combined into a single "product channel" $W : \mathcal{X} \to \mathcal{D}(d)$, with $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and $d = d_1 d_2$, that is defined by

$$W(x_1, x_2) = W_1(x_1) \otimes W_2(x_2).$$

This is the quantum analog of the classical channel $W(y_1, y_2|x_1, x_2) = W(y_1|x_1)W(y_2|x_2)$ defined by the independent uses of classical channels $W_1(y_1|x_1)$ and $W_2(y_2|x_2)$.

Channel Combining for PSCs

Now we are in a position to consider the effect of applying the channel combining operations to binary-input pure-state channels, as first introduced in [10].

Variable Node Combining

Consider the binary-input PSCs W_1 and W_2 with parameters θ_1 and θ_2. The variable node combining operation sends the same input through both channels and combines their output spaces using a tensor product. If the channel input is x, then the first PSC will output $|\psi_x(\theta_1)\rangle$ and the second PSC will output $|\psi_x(\theta_2)\rangle$. Thus, for input x, the output of both channels in the combined Hilbert space $\mathbb{C}^4$ will be

$$|\phi_x\rangle := |\psi_x(\theta_1)\rangle \otimes |\psi_x(\theta_2)\rangle.$$

Since the tensor product of pure states is a pure state, it will become evident below that the combined channel must be unitarily equivalent to a PSC (e.g., with parameter θ).

To consolidate these two channel outputs into a single channel output for the new PSC whose output will be stored in the first qubit, we need a unitary operation $V \in \mathbb{C}^{4\times 4}$ that satisfies

$$V|\phi_x\rangle = |\psi_x(\theta)\rangle \otimes |0\rangle \quad \text{for all } x \in \mathcal{X}.$$

Since V is unitary, this operation will preserve all the information in the two channel outputs.

Here, we note that unitary transforms preserve inner products between vectors and, for any four unit vectors $|\chi_1\rangle, |\chi_2\rangle, |\chi_3\rangle, |\chi_4\rangle \in \mathbb{C}^d$ with $\langle\chi_1|\chi_2\rangle = \langle\chi_3|\chi_4\rangle$, there is a unitary transformation mapping $|\chi_1\rangle$ to $|\chi_3\rangle$ and $|\chi_2\rangle$ to $|\chi_4\rangle$. Thus, we can use (9) and (10) to see that the unitary U exists if and only if

$$\begin{aligned}\langle\phi_0|\phi_1\rangle &= \underbrace{\langle\psi_0(\theta_1)|\psi_1(\theta_1)\rangle}_{\cos(\theta_1)} \underbrace{\langle\psi_0(\theta_2)|\psi_1(\theta_2)\rangle}_{\cos(\theta_2)} \\ &= \underbrace{\langle\psi_0(\theta)|\psi_1(\theta)\rangle}_{\cos(\theta)}.\end{aligned}$$

This demonstrates that lossless variable node combining is possible for the PSC and that the new PSC will

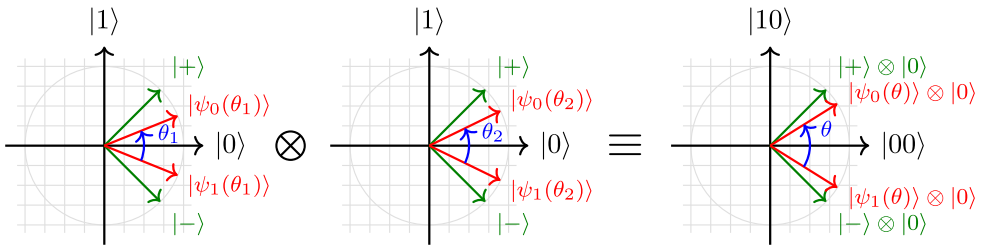

Figure 4

Diagram illustrating variable node combining operation for two PSCs with parameters $\theta_1 = 45°$ and $\theta_2 = 52°$. The result is a PSC with parameter $\theta = \cos^{-1}(\cos(\theta_1)\cos(\theta_2)) \approx 64°$.

have $\theta = \cos^{-1}(\cos(\theta_1)\cos(\theta_2))$. We note that the matrix V is not unique and it depends on the values of θ_1 and θ_2. Following [10, Eqn. (6)], we will denote one such V by

$$U_{\circledast}(\theta_1,\theta_2) := \begin{bmatrix} a_+ & 0 & 0 & a_- \\ a_- & 0 & 0 & -a_+ \\ 0 & b_+ & b_- & 0 \\ 0 & b_- & -b_+ & 0 \end{bmatrix} \tag{15}$$

$$a_\pm := \frac{1}{\sqrt{2}} \frac{\cos\left(\frac{\theta_1-\theta_2}{2}\right) \pm \cos\left(\frac{\theta_1+\theta_2}{2}\right)}{\sqrt{1+\cos(\theta_1)\cos(\theta_2)}}$$

$$b_\pm := \frac{1}{\sqrt{2}} \frac{\sin\left(\frac{\theta_1+\theta_2}{2}\right) \mp \sin\left(\frac{\theta_1-\theta_2}{2}\right)}{\sqrt{1-\cos(\theta_1)\cos(\theta_2)}}.$$

This idea generalizes easily to combining any finite number of PSCs and is illustrated in Figure 4.

Check Node Combining

Now, we describe check node combining for two PSCs by building on the approach for two BSCs in Section "Symmetric Binary-Input Channels." In particular, we let x_1 denote the input to the combined channel and let x_2 and x_3 denote the inputs to the channels which are being combined. As before, due to the even-parity constraint, $x_2 = x_1 \oplus x_3$ and the combined channel output only depends on x_1 and x_3. Specifically, it equals the pure state

$$\begin{aligned} \left|\phi_{x_1,x_3}\right\rangle :=& \left|\psi_{x_1\oplus x_3}(\theta_1)\right\rangle \otimes \left|\psi_{x_3}(\theta_3)\right\rangle \qquad (16) \\ =& \left(\cos\frac{\theta_1}{2}|0\rangle + (-1)^{x_1+x_3}\sin\frac{\theta_1}{2}|1\rangle\right) \\ & \otimes \left(\cos\frac{\theta_2}{2}|0\rangle + (-1)^{x_3}\sin\frac{\theta_2}{2}|1\rangle\right) \\ =& \cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2}|00\rangle + (-1)^{x_3}\cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2}|01\rangle \\ & + (-1)^{x_1+x_3}\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2}|10\rangle \\ & + (-1)^{x_1}\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}|11\rangle \qquad (17) \end{aligned}$$

where we have expanded the tensor product in anticipation of the next step. Our goal is to map all the information about x_1 into the first qubit and treat this qubit as the output of a new PSC.

Observe that, if we group the first and last terms in (17) (i.e., terms involving $|00\rangle$ and $|11\rangle$), then together they form a scaled PSC with input x_1. Likewise, if we group the second and third terms in (17) (i.e., terms involving $|01\rangle$ and $|10\rangle$), then together they form a scaled PSC with input x_1 (since $(-1)^{x_3}$ is in both terms). Now, we exploit these observations in two steps.

First, we apply a unitary V that maps $|00\rangle \mapsto |00\rangle$, $|11\rangle \mapsto |10\rangle$, $|01\rangle \mapsto |01\rangle$, and $|10\rangle \mapsto |11\rangle$ (i.e., $|v_1, v_2\rangle \mapsto |v_1, v_1 \oplus v_2\rangle$) so that the second qubit distinguishes between the two pairs we observed and the first qubit can act as the new PSC output. The unitary V is also known as $\mathrm{CNOT}_{1\to 2}$ and, at this point, we have constructed a superposition of PSCs with common input x_1, where the common output is the first (primary) qubit and the channel parameter depends conditionally on the second (auxiliary) qubit.

To verify the last statement and compute the PSC parameters associated with this construction, we will now analyze projective measurement of the second qubit defined by $\Pi_0 = |00\rangle\langle 00| + |10\rangle\langle 10|$ and $\Pi_1 = |01\rangle\langle 01| + |11\rangle\langle 11|$. Using the results of Section "Measurement of Pure States," we see that the probabilities of the two outcomes are given by

$$\begin{aligned} p_0 &= \left\langle \phi_{x_1,x_3}\middle| V^H \Pi_0 \, V \middle|\phi_{x_1,x_3}\right\rangle \\ &= \cos^2\left(\frac{\theta_1}{2}\right)\cos^2\left(\frac{\theta_2}{2}\right) + \sin^2\left(\frac{\theta_1}{2}\right)\sin^2\left(\frac{\theta_2}{2}\right) \\ &= \frac{1}{2} + \frac{1}{2}\cos(\theta_1)\cos(\theta_2) \qquad (18) \\ p_1 &= \left\langle \phi_{x_1,x_3}\middle| V^H \Pi_1 \, V \middle|\phi_{x_1,x_3}\right\rangle \\ &= \cos^2\left(\frac{\theta_1}{2}\right)\sin^2\left(\frac{\theta_2}{2}\right) + \sin^2\left(\frac{\theta_1}{2}\right)\cos^2\left(\frac{\theta_2}{2}\right) \\ &= \frac{1}{2} - \frac{1}{2}\cos(\theta_1)\cos(\theta_2) \qquad (19) \end{aligned}$$

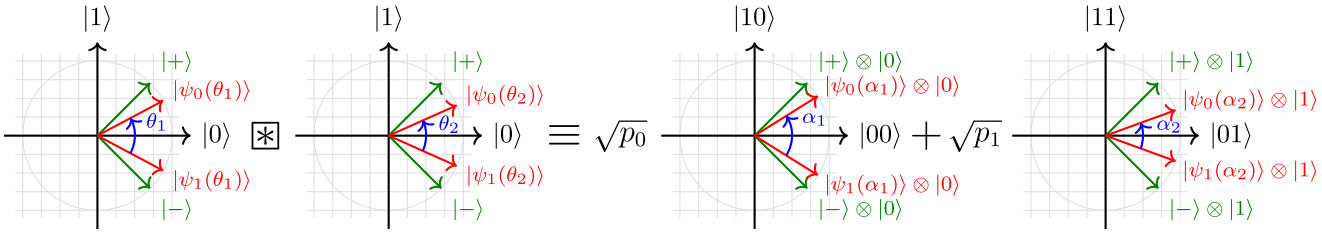

Figure 5

Diagram illustrating the check node combining operation for two PSCs with parameters θ_1 and θ_2. The result is unitarily equivalent to the weighted superposition, with weights $\sqrt{p_0}$ and $\sqrt{p_1}$, of two PSCs with parameters α_1 and α_2.

where the last step in each derivation follows from applying the product identities for sine and cosine a few times. Thus, the postmeasurement states for the two outcomes will be

$$\begin{aligned}
&\frac{1}{\sqrt{p_0}}\Pi_0\ V\left|\phi_{x_1,x_3}\right\rangle \\
&= \frac{1}{\sqrt{p_0}}\left(\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2}|00\rangle + (-1)^{x_1}\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}|10\rangle\right) \\
&= |\psi_x(\alpha_0)\rangle \otimes |0\rangle \\
&\frac{1}{\sqrt{p_1}}\Pi_1\ V\left|\phi_{x_1,x_3}\right\rangle \\
&= \frac{(-1)^{x_3}}{\sqrt{p_1}}\left(\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2}|01\rangle + (-1)^{x_1}\cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2}|11\rangle\right) \\
&= (-1)^{x_3}|\psi_x(\alpha_1)\rangle \otimes |1\rangle
\end{aligned}$$

where α_0 and α_1 are the parameters of the postmeasurement PSCs associated with Π_0 and Π_1, respectively. We also note that the overall phase factor of $(-1)^{x_3}$ on the second state does not affect the measurement error rate (e.g., see Example 2) or future processing. By computing the overlap of these channels and applying trigonometric identities, one can also compute the two possible channel parameters

$$\cos(\alpha_0) = \frac{\cos(\theta_1) + \cos(\theta_2)}{1 + \cos(\theta_1)\cos(\theta_2)} \tag{20}$$

$$\cos(\alpha_1) = \frac{\cos(\theta_1) - \cos(\theta_2)}{1 + \cos(\theta_1)\cos(\theta_2)}. \tag{21}$$

While the previous calculation is based on measuring the second qubit, this can be avoided by using the quantum principle of deferred measurement [10], [13], [14]. The primary drawback of deferring measurements is that all unitary operations done after the deferred measurement must be implemented as conditional unitary operations that depend on the system that was not measured.

This operation is illustrated in Figure 5. It can also be extended to any finite number of PSCs by sequentially combining the channels as one would compute a cumulative sum.

Decoding With BPQM

For mixtures of PSC channels, both channel combining operations result in mixtures of PSC channels, just as combining mixtures of BSCs results in mixtures of BSCs. Thus, the effective channel from x_1 to all the PSC outputs can be succinctly described as a mixture of PSCs. This yields a method of performing the optimal measurement on the actual PSC output systems in order to determine x_1.

Working from the leaves toward the root, the quantum operations associated with the two combinations are recursively applied to the appropriate qubits. At variable nodes, the unitary in (15) is applied, using the classical information describing the parameters of the two PSCs, and one qubit is discarded. At check nodes, a CNOT gate is applied and the auxiliary qubit is measured. The measurement result determines the PSC parameter of the primary qubit, and this information is used in subsequent variable node operations. The sequence of operations defines a quantum circuit consisting of unitary operations and measurements. The output of the circuit is a single qubit, associated with the root node, which is then measured to determine the most-likely channel input. The depth of the resulting quantum circuit is precisely the depth of the factor graph.

If the goal is to decode all the bits in the factor graph, then this algorithm can be applied multiple times each with a different variable node chosen as the root. In this case, however, instead of measuring the auxiliary qubits during check node combining, these measurements are deferred so that the quantum state is preserved for the decoding of the next bit.

Combining General CQ Channels

Now, we will apply the channel combining perspective to factor graphs where the classical random variables are transmitted through CQ channels. First, we generalize (3) to the quantum case. Consider the case where the ith factor graph variable is transmitted through the CQ channel $Q_i : \mathcal{X} \to \mathcal{D}(d_i)$, where $\mathcal{D}(d_i)$ is the set of density matrices for the ith quantum system of dimension d_i. In this case, the conditional

independence of the observations implies that the overall density matrix for input $\boldsymbol{x} = (x_1, \ldots, x_N)$ is given by the tensor product

$$Q(\boldsymbol{x}) := Q_1(x_1) \otimes Q_2(x_2) \otimes \cdots \otimes Q_N(x_N).$$

Using this, the quantum generalization of (4) is

$$\begin{aligned} W_1(x_1) &= \sum_{(x_2,\ldots,x_N)\in\mathcal{X}^{N-1}} P_{X_2,\ldots,X_N|X_1}(x_2,\ldots,x_N|x_1)\, Q(\boldsymbol{x}) \\ &= \frac{1}{P_{X_1}(x_1)} \sum_{(x_2,\ldots,x_N)\in\mathcal{X}^{N-1}} f(\boldsymbol{x})\, Q(\boldsymbol{x}). \end{aligned}$$

This gives the density matrix for the entire joint quantum system as a function of the root variable x_1. Given the set of density matrices $\{W_1(x_1)\}_{x_1\in\mathcal{X}}$, one can solve for an optimal quantum measurement whose outcome maximizes the guessing probability for x_1.

The goal of BPQM is to implement an optimal measurement efficiently using unitary operations acting on small numbers of qubits. To do this, one needs to identify simple quantum operations that implement the combining transformations one step at a time. For example, the classical variable node combining rule in (5) generalizes to the quantum rule

$$W_i(x_i) = Q_i(x_i) \otimes \left(\otimes_{A:\mathrm{pa}(A)=i} W_A(x_i)\right) \tag{22}$$

where $W_i(x)$ is the combined CQ channel for variable node i. Similarly, the classical factor node combining rule in (6) generalizes to the quantum rule

$$W_A(x_{\mathrm{pa}(A)}) = Z_A^{-1} \sum_{\boldsymbol{x}_{\mathrm{ch}(A)}\in\mathcal{X}^{|\mathrm{ch}(A)|}} f_A(x_A)\left(\otimes_{j\in\mathrm{ch}(A)} W_j(x_j)\right) \tag{23}$$

where $W_A(x)$ is the combined CQ channel factor node A and Z_A is the normalization constant chosen so that the resulting matrix has unit trace.

These recursions allow one to analyze general tree-like factor graphs, where the variables are observed through independent CQ channels. However, the dimensions of the density matrices may grow so rapidly that this approach becomes infeasible. This is similar to growth in output alphabet, which is challenging for the classical case. For binary-input PSCs and check node constraints, however, the combining rules outlined in Section "Channel Combining for PSCs" can be used to efficiently represent these density matrices as weighted mixtures of PSCs with different parameters. This can be seen as a special case of the techniques introduced in [15] and [16]. Still, this problem is much less studied in the quantum case and it would be interesting to find efficient methods that work in more generality.

For general CQ channels, the decoding problem on a quantum computer is even more challenging. While the rules outlined in the "Channel Combining for PSCs" section were introduced in [10] and studied further in [13], [14], and [28], it is still not clear if there are efficient and optimal BPQM decoding rules for channels besides the PSC. Suboptimal rules have been introduced and analyzed for symmetric binary-input CQ channels in [15] and [16]. The primary challenge is finding ways to compress the relevant information, from all the CQ observations in a subtree, into the small number of qubits to be passed as messages.

Classical BP also extends to factor graphs with cycles (as an approximate marginalization algorithm) by simply continuing to pass messages around loops in the factor graph [6], [24]. While something similar can be managed for BPQM [14], another peculiarity of quantum mechanics makes this extension much more challenging. The "no cloning" theorem says that quantum states cannot be cloned exactly. Thus, if a variable is measured the first time around a loop, its quantum information collapses and leaves only the classical outcome for subsequent rounds. One approach that provides some gain in experiments is to use "approximate" cloning [14], but this is really the only method that has been explored so far.

Finally, we mention that there is a powerful theory of duality for CQ channels that relates the performance of optimal inference for problems utilizing a code and a channel to related problems utilizing the dual code and dual channel [29]. For this work, it is particularly relevant to note that the dual channel for the BSC is the PSC and vice versa. This special case is discussed in more detail in [30] and [31].

Conclusions and Open Problems

This article notes the importance of BP in efficient error-correction schemes for classical channels and discusses the extension of BP to BPQM in order to apply the same ideas to CQ channels. While the extension to PSCs is well understood, many questions remain for more general CQ channels. In particular, we list here some interesting open questions:

1) Can we apply BPQM to loopy factor graphs and achieve gains or understanding beyond [14]?
2) What assumptions are required to generalize BPQM for PSCs to nonbinary CQ channels with pure-state outputs?
3) Is there a BPQM algorithm that is lossless, or with arbitrarily small loss, for general symmetric binary-input CQ channels with qubit outputs, where the messages consist of a bounded number of qubits?
4) Is there a BPQM algorithm that is lossless, or with arbitrarily small loss, for general CQ channels where the message dimension is a bounded multiple of the channel output dimension?

Acknowledgments

This work was supported in part by the National Science Foundation under Grant 1908730, Grant 2106189, Grant 2106213 , and Grant 2212437. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF. The work of Christophe Piveteau and Joseph M. Renes was supported by the ETH Quantum Center and the Swiss National Science Foundation Sinergia under Grant CRSII5_186364. Henry D. Pfister would like to acknowledge useful discussions on this topic with S. Brandsen and Avijit Mandal.

References

[1] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA, USA: MIT Press, 1963.

[2] J. Pearl, "Reverend Bayes on inference engines: A distributed hierarchical approach," in *Proc. AAAI Conf. Artif. Intell.*, 1982, pp. 133–136.

[3] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Reasoning*, San Francisco, CA, USA: Morgan Kaufmann, 1988.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun.*, 1993, vol. 2, pp. 1064–1070.

[5] R. J. McEliece, D. J. C. MacKay, and J. Cheng, "Turbo decoding as an instance of Pearl's "belief propagation" algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 140–152, Feb. 1998.

[6] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[7] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 325–343, Mar. 2000.

[8] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[9] M. M. Wilde and S. Guha, "Polar codes for classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 1175–1187, Feb. 2013.

[10] J. M. Renes, "Belief propagation decoding of quantum channels by passing quantum messages," *New J. Phys.*, vol. 19, no. 7, 2017, Art. no. 072001.

[11] M. P. da Silva, S. Guha, and Z. Dutton, "Achieving minimum-error discrimination of an arbitrary set of laser-light pulses," *Phys. Rev. A*, vol. 87, no. 5, 2013, Art. no. 052320.

[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, Dec. 2015, Art. no. 062333.

[13] N. Rengaswamy, K. P. Seshadreesan, S. Guha, and H. D. Pfister, "Belief propagation with quantum messages for quantum-enhanced classical communications," *NPJ Quantum Inf.*, vol. 7, 2021, Art. no. 97.

[14] C. Piveteau and J. M. Renes, "Quantum message-passing algorithm for optimal and efficient decoding," *Quantum*, vol. 6, 2022, Art. no. 784.

[15] S. Brandsen, A. Mandal, and H. D. Pfister, "Belief propagation with quantum messages for symmetric classical-quantum channels," in *Proc. IEEE Inf. Theory Workshop*, 2022, pp. 494–499.

[16] A. Mandal, S. Brandsen, and H. D. Pfister, "Belief-propagation with quantum messages for polar codes on classical-quantum channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2023.

[17] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[18] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[19] R. Kinderman and J. Snell, *Markov Fields and Their Application of Contemporary Mathematics*, vol. 1. Rhode Island, USA: American Math. Soc., 1980.

[20] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[21] Y.-Y. Shi, L.-M. Duan, and G. Vidal, "Classical simulation of quantum many-body systems with a tree tensor network," *Phys. Rev. A*, vol. 74, Aug. 2006, Art. no. 022320.

[22] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

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

[24] A. T. Ihler, J. W. Fisher III, A. S. Willsky, and D. M. Chickering, "Loopy belief propagation: Convergence and effects of message errors," *J. Mach. Learn. Res.*, vol. 6, no. 5, pp. 905–936, 2005.

[25] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[26] M. M. Wilde, *Quantum Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[27] J. M. Renes, *Quantum Information Theory: Concepts and Methods*. Munich, Germany: De Gruyter Oldenbourg, 2022.

[28] N. Rengaswamy, K. P. Seshadreesan, S. Guha, and H. D. Pfister, "Quantum advantage via qubit belief propagation," in *Proc. IEEE Int. Symp. Inf. Theory*, 2020, pp. 1824–1829.

[29] J. M. Renes, "Duality of channels and codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 577–592, Jan. 2018.

[30] N. Rengaswamy and H. D. Pfister, "On the duality between the BSC and quantum PSC," in *Proc. IEEE Int. Symp. Inf. Theory*, 2021, pp. 2232–2237.

[31] N. Rengaswamy and H. D. Pfister, "A semiclassical proof of duality between the classical BSC and the quantum PSC," 2021, *arXiv:2103.09225*.

Henry D. Pfister (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of California, San Diego, CA, USA, in 2003.

He is currently a Professor with the Electrical and Computer Engineering Department, Duke University, with a secondary appointment in Mathematics. Prior to that, he was an Associate Professor with Texas A&M University (2006–2014) and a postdoctoral fellow with the École Polytechnique Fédérale de Lausanne (2005–2006). He is a coauthor of the 2007 IEEE COMSOC best paper in Signal Processing and Coding for Data Storage, a coauthor of a 2016 Symposium on the Theory of Computing best paper. Dr. Pfister is a recipient of the 2021 Information Theory Society Paper Award.

Christophe Piveteau received the B.Sc. degree in physics and mathematics and the M.Sc. degree in physics from ETH Zürich, Switzerland. Since 2020, he has been working toward the Ph.D. program in quantum information theory at ETH Zürich.

He worked with IBM Research Zürich to develop algorithms for in-memory computing and deep learning acceleration with memristive crossbar arrays. His research interests include, among others, quantum error correction, quantum error mitigation, circuit knitting, and machine learning.

Joseph M. Renes (Member, IEEE) received the B.S. degree from Caltech, Pasadena, CA, USA, in 1999, and the Ph.D. degree from the University of New Mexico, Albuquerque, NM, USA, in 2004, both in physics.

Between 2005 and 2007, he was an Alexander von Humboldt Research Fellow, first at the University of Erlangen, and later at the Technical University of Darmstadt. He is currently a Senior Scientist with the Institute of Theoretical Physics, ETH Zürich, Zurich, Switzerland. He is the author of the book *Quantum Information Theory: Concepts and Methods* (DeGruyter, 2022).

Narayanan Rengaswamy (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Duke University, Durham, NC, USA, in 2020, under the supervision of Henry Pfister and Robert Calderbank.

He is an Assistant Professor of Electrical and Computer Engineering with The University of Arizona. He also works with the NSF-ERC Center for Quantum Networks. Before this, he was a Postdoctoral Research Associate with Bane Vasić in the same department. He worked with Laurent Schmalen and Vahid Aref as a graduate research intern at Alcatel-Lucent Bell Labs, Stuttgart, Germany. His research interests are classical and quantum error correction, quantum computing, quantum networking, and information theory.