Available online at www.sciencedirect.com
ScienceDirect
Annals of Physics 323 (2008) 1899–1946
ANNALS of PHYSICS
www.elsevier.com/locate/aop

Quantum Graphical Models and Belief Propagation

M.S. Leifer [a,b], D. Poulin [c,*]

[a] *Institute for Quantum Computing, University of Waterloo, 200 University Avenue West, Waterloo Ont., Canada N2L 3G1*
[b] *Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo Ont., Canada N2L 2Y5*
[c] *Center for the Physics of Information, California Institute of Technology, 1200 E. California Boulevard, 107-81, Pasadena, CA 91125, USA*

Received 21 September 2007; accepted 7 October 2007
Available online 12 October 2007

Abstract

Belief Propagation algorithms acting on Graphical Models of classical probability distributions, such as Markov Networks, Factor Graphs and Bayesian Networks, are amongst the most powerful known methods for deriving probabilistic inferences amongst large numbers of random variables. This paper presents a generalization of these concepts and methods to the quantum case, based on the idea that quantum theory can be thought of as a noncommutative, operator-valued, generalization of classical probability theory. Some novel characterizations of quantum conditional independence are derived, and definitions of Quantum n-Bifactor Networks, Markov Networks, Factor Graphs and Bayesian Networks are proposed. The structure of Quantum Markov Networks is investigated and some partial characterization results are obtained, along the lines of the Hammersley–Clifford theorem. A Quantum Belief Propagation algorithm is presented and is shown to converge on 1-Bifactor Networks and Markov Networks when the underlying graph is a tree. The use of Quantum Belief Propagation as a heuristic algorithm in cases where it is not known to converge is discussed. Applications to decoding quantum error correcting codes and to the simulation of many-body quantum systems are described.

Keywords: Quantum information; Markov Networks; Bayesian Networks; Factor Graphs; Graphoids; Belief Propagation; Sum–product; Quantum error correction; Quantum many-body systems

* Corresponding author. Fax: +1 626 568 2764.
E-mail addresses: matt@mattleifer.info (M.S. Leifer), dpoulin@ist.caltech.edu (D. Poulin).

0003-4916/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.aop.2007.10.001

1. Introduction

Quantum theory is first and foremost a calculus for computing the probabilities of outcomes of measurements made on physical systems. Therefore, the generic problem in quantum theory is one of probabilistic inference, i.e. given a specified class of quantum states, compute the predicted probabilities of measurement outcomes and their correlations. For example, computing the correlation functions of a system in the ground state of a Hamiltonian, or computing the probabilities for the possible measurement outcomes after implementing a quantum circuit, are problems of this general type. Such quantum inferences present a formidable computational challenge as the number of subsystems becomes large, since the number of parameters needed to specify a quantum state grows exponentially with the number of subsystems, and the formulas for quantities of interest typically also involve an exponentially large number of terms.

A similar problem arises in classical probabilistic inference, since the number of terms required to specify a general probability distribution also grows exponentially with the number of random variables involved. A variety of algorithms for classical probabilistic inference have been discovered, of which Belief Propagation algorithms on Graphical Models are amongst the most powerful. Such algorithms are particularly interesting for two reasons. Firstly, they are highly parallelizable in the sense that they can be implemented by associating each random variable with a separate processor. Messages are received and sent by the processors along the links of a network corresponding to the edges of a graph and, importantly, the order in which the messages arrive does not matter. Secondly, Belief Propagation performs remarkably well as a heuristic algorithm, even in cases where it is not guaranteed to converge to the exact solution. Important examples include the near optimal decoding of low density [14] and turbo [8] error correction codes, spin glass models [33], and random satisfiability problems [34]. Understanding the reasons for this is currently an active area of research, but it is understood [58] to be related to a hierarchy of approximation schemes commonly used in statistical physics.

Due to the similarity between the classical and quantum problems, one might hope to leverage the power of Belief Propagation in the quantum case also, especially since quantum theory can be regarded as a noncommutative generalization of classical probability theory. This is indeed the case, and in this paper we develop the necessary theory of Quantum Belief Propagation and its associated Graphical Models.

This paper should be of interest to researchers in Graphical Models and Belief Propagation, as well as to researchers in quantum theory, particularly in quantum information and the simulation of quantum many-body systems. As such, it is intended to be as self-contained as possible, although we do assume familiarity with the basic formalism of quantum theory on finite dimensional Hilbert spaces, including the theory of density matrices, generalized measurements and completely positive maps, as used in quantum information theory. These are covered in detail in the textbook of Nielsen and Chuang [38], as well as in Preskill's lecture notes [43]. For further background on classical Graphical Models and Belief Propagation, we suggest the texts of Lauritzen [24], MacKay [31], and Neapolitan [36,37], as well as the review articles by Yedida et al. [58,59] and Aji and McEliece [6].

The remainder of this paper is structured as follows. In Section 2, the generic classical and quantum probabilistic inference problems are defined. In Section 3, we review the notions of classical and quantum conditional independence, which are crucial for the

development of Graphical Models and Belief Propagation algorithms. Section 3.1 outlines the entropic approach to conditional independence based on the vanishing of conditional mutual information and the associated constraints on conditional and mutual probability distributions. This entropic approach has a straightforward quantum generalization based on the equality conditions for strong subadditivity, which is described in Section 3.2. Section 3.3 introduces the quantum conditional and mutual density operators, which are analogous to classical conditional and mutual probability distributions, and Section 3.4 explains how quantum conditional independence can be characterized directly in terms of them.

In Section 4, we develop the theory of quantum Graphical Models. Section 4.1 reviews the definition of classical Markov Networks and the Hammersley–Clifford theorem, which gives an explicit representation of the probability distributions supported on them. Motivated by this, Section 4.2 defines the class of quantum n-Bifactor Networks, which are the most general class of networks on which our Belief Propagation algorithms operate. Section 4.3 reviews the theory of dependency models and graphoids, which are abstractions of the conditional independence relation, and a quantum graphoid is defined based on quantum conditional independence. Section 4.4 uses the quantum graphoid to define quantum Markov Networks and gives some partial characterization theorems, along the lines of the classical Hammersley–Clifford theorem, which connect quantum Markov Networks to n-Bifactor Networks. Section 4.5 briefly discusses quantum generalizations of two other classical Graphical Models: Factor Graphs and Bayesian Networks. Fig. 10 sketches the relation between some of these Graphical Models, and summarizes the Quantum Belief Propagation algorithm's domain of convergence.

Section 5 discusses the Quantum Belief Propagation algorithms. In Section 5.1, QBP algorithms are described for n-Bifactor Networks. In Section 5.2, QPB is shown to converge for 1-Bifactor Networks on trees and for general Bifactor Networks on trees that are also Quantum Markov Networks. Section 6 discusses some methods for using QBP as a heuristic algorithm in cases where it is not known to converge. These are coarse graining Section 6.1, sliding window QBP Section 6.2 and the method of replicas Section 6.3.

Section 7 presents two applications of QBP: to decoding quantum error correcting codes in Section 7.1 and to simulating many-body quantum systems in Section 7.2. In particular, Section 7.2 explains how projected entangled-pair states, which have been successfully used in statistical physics as approximations to the ground states of a wide class of Hamiltonians, can be incorporated into the framework of Bifactor Networks.

To conclude, Section 8 discusses the relationship to other quantum generalizations of Graphical Models and Belief Propagation that have been proposed and Section 9 describes open questions and future research directions suggested by this work.

Note that a slightly unconventional notation for probability distributions on sets of random variables and for quantum states on tensor products of quantum systems is used throughout. This is very convenient for describing Graphical Models and is reviewed in Appendix A.

2. Classical and quantum probabilistic inference

Classical Graphical Models are designed to be used as tools for making probabilistic inferences amongst large numbers of correlated random variables. Consider a set random variables, $V = \{v_1, v_2, \ldots, v_N\}$, each of which takes a finite number of integer values

$\{1, 2, \ldots, d\}$. To specify a general probability distribution, $P(V)$, over the variables requires $O(d^N)$ parameters. On learning that some subset of the variables $U \subseteq V$ take particular values, denoted $\widetilde{U} = \{u = j_u\}_{u \in U}$, an important task is to update the probability for some other disjoint subset of variables $W \subseteq V$ via Bayes rule

$$P(W|\widetilde{U}) = \frac{P(\widetilde{U} \cup W)}{P(\widetilde{U})} = \frac{\sum_{V-(U\cup W)} P(\widetilde{U} \cup (V-U))}{\sum_{V-U} P(\widetilde{U} \cup (V-\widetilde{U}))}. \tag{1}$$

This immediately raises two problems. Firstly, the number of parameters needed to specify the input to the computation, i.e. the probability distribution itself, is exponential in N. We would like to specify a well-defined computational problem in which N measures the input size. Therefore, it is not feasible to consider the full set of probability distributions over N variables, and attention must be restricted to families of distributions that can be specified with a number of parameters that grows only polynomially in N. Secondly, assuming that the sizes of U and W are held constant as N increases, Eq. (1) involves sums over a number of terms that is exponential in N. Thus, a straightforward evaluation of the formula would not give an efficient algorithm. The restriction on the class of probability distributions must somehow be used to find an alternative method of computation that is efficient.

Classical Graphical Models are designed to provide an efficient representation of classes of probability distributions and Belief Propagation algorithms are designed to solve the corresponding inference problem.

In quantum theory, the random variables are replaced by a set of N quantum systems $V = \{v_1, v_2, \ldots, v_N\}$, each associated with a Hilbert space of dimension d. Again, it takes an exponential in N number of parameters to specify a general density operator ρ_V. The analog of the inference in Eq. (1) is to perform a positive operator valued measure (POVM) $\left\{E_U^{(j)}\right\}$ on a subsystem $U \subseteq V$ and, on obtaining outcome j, update the state of some disjoint subsystem $W \subseteq V$ according to

$$\rho_{U|E_U^{(j)}} = \frac{\mathrm{Tr}_U\left(E_U^{(j)} \rho_{U\cup W}\right)}{\mathrm{Tr}\left(E_U^{(j)} \rho_U\right)} = \frac{\mathrm{Tr}_{V-W}\left(E_U^{(j)} \rho_V\right)}{\mathrm{Tr}\left(E_U^{(j)} \rho_V\right)}. \tag{2}$$

It should be noted that this quantum problem reduces to the classical case when all the operators involved commute and are diagonal in a product basis of the systems in V. In this sense Eq. (2) is a noncommutative generalization of Eq. (1) and this correspondence provides the guiding principle that we use to generalize the classical theory.

The quantum problem raises the same sort of issues as in the classical case, since it takes an exponential in N number of parameters to specify a state on N subsystems and the trace and partial trace in Eq. (2) involve sums over an exponential number of terms. In quantum many-body theory, physical considerations are often used to motivate solutions to the representation problem, e.g. we may restrict attention to the ground or Gibbs states of some class of efficiently specifiable Hamiltonians. In this paper, we take a different approach and instead generalize the sort of constraints that are used in defining classical Graphical Models. The reasons for this are twofold. Firstly, with the advent of quantum information science, it is relevant to solve instances of Eq. (2) that are of broader scope than those typically considered in statistical physics. For example, we may be interested in states that are the output of a class of polynomial quantum circuits, or in the code states of a quan-

tum error correction code. The most natural way to phrase such constraints is not always in terms of Hamiltonians, although it may be possible to do so. Secondly, by focussing on constraints with a clear probabilistic and information theoretic meaning, the connection between the classical and quantum problems is elucidated and the results of the vast literature on the classical inference problem can be called into play.

3. Conditional independence

The formal construction of classical Graphical Models is based on the idea of placing conditional independence constraints on sets of random variables. In this section, the relevant classical definitions are reviewed and their quantum generalizations are introduced. In Section 3.1, the entropic approach to conditional independence is outlined and the corresponding constraints on conditional and mutual probability distributions are reviewed. In Section 3.2, the entropic definition is straightforwardly generalized to the quantum case by replacing the Shannon entropy with the von Neumann entropy. In order to provide constraints on density operators that are analogous to those for classical conditional and mutual probability distributions, conditional and mutual density operators are defined in Section 3.3 and quantum conditional independence is expressed in terms of them in Section 3.4.

3.1. Classical conditional independence

For a set V of classical random variables with joint distribution $P(V)$, the marginal distribution for any $U \subseteq V$ is defined as $P(U) = \sum_{V-U} P(V)$ and for any two disjoint sets $U, W \subseteq V$, the conditional distribution of U given W is defined as

$$P(U|W) = \frac{P(U \cup W)}{P(W)}. \tag{3}$$

The Shannon entropy of any $U \subseteq V$ is defined as

$$H(U) = -\sum_{U} P(U) \log_2 P(U). \tag{4}$$

For disjoint $U, W \subseteq V$, the conditional entropy of U given W is defined as

$$H(U|W) = -\sum_{U \cup W} P(U \cup W) \log_2 P(U|W), \tag{5}$$

and satisfies the identity

$$H(U|W) = H(U \cup W) - H(W). \tag{6}$$

The mutual information between U and W is defined to be

$$H(U : W) = H(U) - H(U|W) \tag{7}$$

$$= H(U) + H(W) - H(U \cup W). \tag{8}$$

Note that $H(U : W) = 0$ iff $P(U \cup W) = P(U)P(W)$. For three disjoint sets $U, W, X \subseteq V$, the conditional mutual information between U and W, given X is defined to be

$$H(U:W|X) = H(U|X) - H(U|W \cup X) \tag{9}$$
$$= H(U \cup X) + H(W \cup X) - H(X) - H(U \cup W \cup X). \tag{10}$$

The condition $H(U:W|X) = 0$ is known as *conditional independence* of U and W given X and it is equivalent to any of the following conditions

$$P(U|W \cup X) = P(U|X) \tag{11}$$
$$P(W|U \cup X) = P(W|X) \tag{12}$$
$$P(U \cup W|X) = P(U|X)P(W|X) \tag{13}$$
$$P(U \cup W \cup X) = P(U|X)P(W|X)P(X). \tag{14}$$

Example 3.1. Consider a Markov chain consisting of three random variables $u - x - w$. The defining condition for such a process is that u and w are conditionally independent given x. Thus, Eq. (14) immediately implies that the joint probability distribution has the form

$$P(u,x,w) = P(u|x)P(w|x)P(x). \tag{15}$$

In general, a joint distribution of three variables can be written as $P(u,x,w) = P(w|u,x)P(x|u)P(u) = P(u|x,w)P(x|w)P(w)$ and so Eqs. (11) and (12) imply that $P(u,x,w)$ can also be written as

$$P(u,x,w) = P(w|x)P(x|u)P(u) \tag{16}$$
$$P(u,x,w) = P(u|x)P(x|w)P(w). \tag{17}$$

The three equivalent decompositions given in Eqs. (15)–(17) are suggestive of three different types of causal scenario that might give rise to such a Markov chain:

(15) suggests x is a common cause of u and w: $u \leftarrow x \rightarrow w$

(16) suggests u causes x and then x causes w: $u \rightarrow x \rightarrow w$

(17) suggests w causes x and then x causes u: $u \leftarrow x \leftarrow w$.

The common feature of these three scenarios is that in each case all the correlations between u and w are mediated by x. Ultimately, conditional independence captures this common feature rather than implying any specific causal scenario.

The example shows that care should be taken when interpreting a decomposition of a joint probability distribution into conditional and marginal distributions. Conditional independence is about the structure of correlations between random variables rather than their specific causal relations. For this reason it is often useful to replace conditional probabilities with an object that is more closely connected with correlation.

The *mutual probability distribution* of disjoint $U, W \subseteq V$ is given by

$$P(U:W) = \frac{P(U \cup W)}{P(U)P(W)} = \frac{P(U|W)}{P(U)}. \tag{18}$$

As the name implies, this is related to the mutual information and it is easy to check that Eq. (7) can be rewritten as

$$H(U:V) = \sum_{U \cup W} P(U,W) \log_2 P(U:W). \tag{19}$$

The conditional independence conditions Eqs. (11)–(14) can be re-expressed in terms of mutual distributions as

$$P(U:W \cup X) = P(U:X) \tag{20}$$
$$P(W:U \cup X) = P(W:X) \tag{21}$$
$$P(U \cup W:X) = P(U:X)P(W:X) \tag{22}$$
$$P(U \cup W \cup X) = P(U:X)P(W:X)P(X)P(U)P(W). \tag{23}$$

Example 3.2. Returning to the Markov chain of Example 3.1, the decompositions Eqs. (15)–(17) can all be rewritten in terms of mutual distributions by replacing each conditional probability with the product of a marginal and a mutual distribution using the relation $P(U|W) = P(U:W)P(U)$. All three decompositions reduce to the same expression:

$$P(u,x,w) = P(u)P(x)P(w)P(u:x)P(x:w). \tag{24}$$

This decomposition clearly shows that all correlations between u and w are mediated by x and avoids the causal ambiguities that are implicit in the use of conditional probabilities.

3.2. Quantum conditional independence

Turning now to the quantum case, if V is a set of subsystems then the joint state is a density operator ρ_V. For $U \subseteq V$, the analog of a marginal distribution is the reduced state obtained by taking a partial trace over $V - U$, i.e. $\rho_U = \mathrm{Tr}_{V-U}(\rho_V)$. The Shannon entropy is replaced by the von Neumann entropy, defined as

$$S(U) = -\mathrm{Tr}(\rho_U \log_2 \rho_U). \tag{25}$$

Quantum analogs of conditional and mutual probability distributions are not commonly discussed in the literature, but they are needed to obtain decompositions of the joint density operator analogous to Eqs. (11)–(14) and (20)–(23), so they are introduced in the next section. For now, note that the quantum conditional entropy, mutual information and conditional mutual information can already be defined by simply replacing H with S in the expressions (6), (8), and (10), since these expressions only involve joint and marginal probability distributions.

By comparison with the classical case, it is natural to consider $S(U:W|X) = 0$ as a definition of *quantum conditional independence*. In fact, the inequality $S(U:W|X) \geqslant 0$ always holds and is known as strong subadditivity, so quantum conditional independence is simply the equality condition for strong subadditivity. This equality condition has been investigated extensively and has been shown [20] to be equivalent to the existence a decomposition of the Hilbert space $\mathcal{H}_X$ of the form

$$\mathcal{H}_X = \bigoplus_{j=1}^{d} \left(\mathcal{H}_{X_j^L} \otimes \mathcal{H}_{X_j^R} \right), \tag{26}$$

(the superscripts L and R indicate the left and right sector of the tensor product) such that the joint density operator $\rho_{U \cup W \cup X}$ can be written as

$$\rho_{U\cup W\cup X} = \sum_{j=1}^{d} p_j \sigma_{UX_j^L} \otimes \tau_{X_j^R W}, \tag{27}$$

where $0 \leqslant p_j \leqslant 1$, $\sum_{j=1}^{d} p_j = 1$, and $\sigma_{UX_j^L}$ and $\tau_{X_j^R W}$ are density operators on $\mathcal{H}_U \otimes \mathcal{H}_{X_j^L}$ and $\mathcal{H}_{X_j^R} \otimes \mathcal{H}_W$, respectively.

Less explicit formulations of the equality condition have also been found [45], such as the operator equality

$$\log \rho_{UWX} + \log \rho_X = \log \rho_{UX} + \log \rho_{WX}, \tag{28}$$

where the logarithms are restricted to the supports of the operators.

3.3. Conditional and mutual density operators

Quantum conditional independence can be expressed in a form closer to the classical conditions Eqs. (11)–(14) and (20)–(23) by introducing definitions of *conditional and mutual density operators*. For this purpose, it is convenient to define a family of products for pairs of operators A, B as follows.

$$A \star^{(n)} B = \left(A^{\frac{1}{2n}} B^{\frac{1}{n}} A^{\frac{1}{2n}} \right)^n. \tag{29}$$

An important property of the $\star^{(n)}$ products is that if A and B are both positive operators then $A\star^{(n)}B$ is also positive. In what follows, the most frequently used of these products are $A \star B = A\star^{(1)}B$ and

$$A \odot B = \lim_{n\to\infty} (A\star^{(n)}B). \tag{30}$$

Note that whilst $\odot$ is commutative and associative, $\star^{(n)}$ is neither in general, so particular attention must be paid to the ordering of operators.

The product $\odot$ was previously introduced in [56], in the context of a Bayesian calculus for quantum theory, and it satisfies the formula

$$A \odot B = \exp(\log A + \log B), \tag{31}$$

whenever A and B are strictly positive. If A and B are semi-positive, then Eq. (31) may be extended by restricting the action of the logarithm to the supports of the operators.

The $\star^{(n)}$ products can be used to define a family of conditional density operators. Let V be a set of quantum systems in a state ρ_V and let $U, W \subseteq V$ be disjoint. Define

$$\rho_{U|W}^{(n)} = \rho_W^{-1} \star^{(n)} \rho_{U\cup W}, \tag{32}$$

where superscript -1 denotes the Moore–Penrose pseudoinverse.[1] Note that if $W = \emptyset$, so that $\mathcal{H}_W = \mathbb{C}$ is the trivial Hilbert space, then $\rho_{U|W}^{(n)} = \rho_U$. The conditional density operators used most frequently in this paper are $\rho_{U|W} = \rho_{U|W}^{(1)}$ and $\rho_{U|W}^{(\infty)} = \rho_W^{-1} \odot \rho_{U\cup W}$.

The operator $\rho_{U|W}^{(\infty)}$ was originally introduced [11] because it allows the quantum conditional entropy to be expressed via a formula analogous to Eq. (5)

[1] In the present case this means that ρ_W^{-1} is the inverse of ρ_W when restricted to the support of ρ_W and has the same null space as ρ_W.

$$S(U|W) = -\mathrm{Tr}\left(\rho_{U\cup W}\log_2\rho^{(\infty)}_{U|W}\right). \tag{33}$$

The operator $\rho_{U|W}$ was introduced in [28,27,7] and also exhibits strong analogies with classical conditional probability.

The corresponding family of *mutual density operators* is defined similarly via

$$\rho^{(n)}_{U:W} = \left(\rho_U^{-1}\otimes\rho_W^{-1}\right)\star^{(n)}\rho_{U\cup W} = \rho_U^{-1}\star^{(n)}\rho^{(n)}_{U|W}, \tag{34}$$

with $\rho^{(\infty)}_{U:W}$ and $\rho_{U:W}$ defined in the obvious way.

The operator $\rho^{(\infty)}_{U:W}$ was introduced [11] in order to express the quantum mutual information via a formula analogous to Eq. (19)

$$S(U:W) = -\mathrm{Tr}\left(\rho_{U\cup W}\log_2\rho^{(\infty)}_{U:W}\right). \tag{35}$$

3.4. Constraints on conditional and mutual density operators

In this section, quantum conditional independence is shown to be equivalent to constraints on conditional and mutual density operators analogous to Eqs. (11)–(14) and (20)–(23).

Theorem 3.3. *If $S(U:W|X) = 0$ then the following conditions hold*:

$$\rho^{(n)}_{U|X\cup W} = \rho^{(n)}_{U|X}\otimes P_W \tag{36}$$

$$\rho^{(n)}_{W|X\cup U} = \rho^{(n)}_{W|X}\otimes P_U \tag{37}$$

$$\rho^{(n)}_{U\cup W|X} = \rho^{(n)}_{U|X}\rho^{(n)}_{W|X} \tag{38}$$

$$\rho_{U\cup W\cup X} = \rho_X\star^{(n)}\left(\rho^{(n)}_{U|X}\rho^{(n)}_{W|X}\right), \tag{39}$$

where P_W is the projector onto the support of ρ_W and P_U is the projector onto the support of ρ_U.

Proof. These conditions are a direct consequence of the decomposition given in Eq. (27). Since each $\mathcal{H}_{X^L_j}$ is a factor in a direct sum decomposition of $\mathcal{H}_X$, it follows that the operators $\sigma_{UX^L_j}$ have disjoint support. Similarly, the operators $\tau_{WX^R_j}$ have disjoint support. Hence, to prove Eq. (36) note that

$$\rho_{W\cup X} = \sum_{j=1}^{d} p_j\sigma_{X^L_j}\otimes\tau_{X^R_jW}, \tag{40}$$

and hence

$$\rho^{(n)}_{U|W\cup X} = \rho^{-1}_{WX}\star^{(n)}\rho_{UWX} \tag{41}$$

$$= \sum_{j=1}^{d}\left(\sigma^{-1}_{X^L_j}\star^{(n)}\sigma_{UX^L_j}\right)\otimes\left(\tau^{-1}_{X^R_jW}\star^{(n)}\tau_{X^R_jW}\right) \tag{42}$$

$$= \sum_{j=1}^{d}\sigma^{(n)}_{U|X^L_j}\otimes P_{X^R_jW} \tag{43}$$

$$= \rho^{(n)}_{U|X}\otimes P_W, \tag{44}$$

where $P_{X^R_jW}$ is the projector onto the support of $\tau_{X^R_jW}$.

Eqs. (37) and (38) are proved similarly, with the proviso that the decomposition given in Eq. (27) implies that $\rho^{(n)}_{U|X}$ and $\rho^{(n)}_{W|X}$ commute, which is necessary to prove Eq. (38). Finally, (39) is equivalent to (38) via the definition a conditional density operator. □

It is straightforward to adapt the proof in order to arrive at analogous decompositions in terms of mutual density operators.

Theorem 3.4. *If $S(U:W|X)=0$ then the following conditions hold*:

$$\rho^{(n)}_{U:X\cup W} = \rho^{(n)}_{U:X}\otimes P_W \tag{45}$$

$$\rho^{(n)}_{W:X\cup U} = \rho^{(n)}_{W:X}\otimes P_U \tag{46}$$

$$\rho^{(n)}_{U\cup W:X} = \rho^{(n)}_{U:X}\rho^{(n)}_{W:X} \tag{47}$$

$$\rho_{U\cup W\cup X} = (\rho_U\otimes\rho_W\otimes\rho_X)\star^{(n)}\left(\rho^{(n)}_{U:X}\rho^{(n)}_{W:X}\right). \tag{48}$$

It remains to determine whether any converse implications hold, i.e. which of the conditions Eqs. (36)–(39) and (45)–(48) imply that $S(U:W|X)=0$. For this purpose, it is only necessary to consider Eqs. (36)–(38) because Eqs. (45)–(48) are equivalent to Eqs. (36)–(39) via the definition of a mutual density operator and Eq. (39) is equivalent to Eq. (38) via the definition of a conditional density operator. In general, the situation appears to be more complicated than in the classical case and we are only able to obtain tight converse results for the cases $n\to\infty$ and $n=1$.

Theorem 3.5. *In the limit, $n\to\infty$, all the converse implications hold, i.e. any of the conditions (36)–(38) imply that $S(U:W|X)=0$.*

Proof. These results are simple consequences of the equality condition given in Eq. (28). For Eq. (36) we have

$$\rho^{-1}_{W\cup X}\odot\rho_{U\cup W\cup X} = \rho^{-1}_X\odot\rho_{U\cup X}. \tag{49}$$

Using Eq. (31) gives

$$\exp(\log\rho_{U\cup W\cup X}-\log\rho_{W\cup X}) = \exp(\rho_{U\cup X}-\rho_X). \tag{50}$$

Taking logarithms and rearranging gives Eq. (28). The proofs for Eqs. (37) and (38) follow by similar arguments. □

For the $n=1$ case, Eqs. (36) and (37) imply converse results.

Theorem 3.6. *If $\rho_{U|X\cup W}=\rho_{U|X}$ or $\rho_{W|X\cup U}=\rho_{W|X}$ then $S(U:W|X)=0$.*

Proof. As explained in [20], Uhlman's theorem [50], implies that $S(U:W|X)=0$ iff there exists a trace preserving, completely positive map $\mathcal{E}_{U\cup X\cup W|U\cup X}:\mathfrak{L}(\mathcal{H}_U\otimes\mathcal{H}_X)\to\mathfrak{L}(\mathcal{H}_U\otimes\mathcal{H}_X\otimes\mathcal{H}_W)$, such that both

$$\mathcal{E}_{U\cup X\cup W|U\cup X}(\rho_U\otimes\rho_X) = \rho_U\otimes\rho_{X\cup W} \tag{51}$$

$$\mathcal{E}_{U\cup X\cup W|U\cup X}(\rho_{U\cup X}) = \rho_{U\cup X\cup W} \tag{52}$$

hold simultaneously. In the present case, this can be achieved via a map of the form $\mathcal{E}_{U\cup X\cup W|U\cup X}=\mathcal{I}_U\otimes\mathcal{F}_{X\cup W|X}$, where $\mathcal{I}_U$ is the identity superoperator on $\mathfrak{L}(\mathcal{H}_U)$ and $\mathcal{F}_{X\cup W|X}:\mathfrak{L}(\mathcal{H}_X)\to\mathfrak{L}(\mathcal{H}_X\otimes\mathcal{H}_W)$ is a trace preserving completely positive

map. $\mathcal{F}_{X\cup W|X}$ is defined via a Kraus representation $\mathcal{F}_{X\cup W|X}(\sigma_X) = \sum_j M^{(j)}_{X\cup W|X}\sigma_X M^{(j)\dagger}_{X\cup W|X}$, where

$$M^{(j)}_{X\cup W|X} = \rho^{\frac{1}{2}}_{X\cup W}|j\rangle_W \rho_X^{-\frac{1}{2}}, \tag{53}$$

and $|j\rangle_W$ are basis vectors for $\mathcal{H}_W$.

It is straightforward to check that $\sum_j M^{(j)\dagger}_{X\cup W|X} M^{(j)}_{X\cup W|X} = P_X$, where P_X is the projector onto the support of ρ_X. This can easily be extended to be a trace preserving map by adding an extra Kraus operator that has support only in the subspace orthogonal to the support of ρ_X, but this can be omitted for the present purpose since it does not change the action of $\mathcal{E}_{U\cup X\cup W|U\cup X}$ on $\rho_U\otimes\rho_X$ or $\rho_{U\cup X}$. It is straightforward to check that $\mathcal{F}_{X\cup W|X}(\rho_X) = \rho_{X\cup W}$, so the first condition is satisfied. The action on $\rho_{U\cup X}$ is given by

$$\mathcal{I}_U\otimes\mathcal{F}_{X\cup W|X}(\rho_{U\cup X}) = \sum_j I_U\otimes M^{(j)}_{X\cup W|X}\rho_{U\cup X} I_U\otimes M^{(j)}_{X\cup W|X} \tag{54}$$

$$= \rho^{\frac{1}{2}}_{X\cup W}\sum_j |j\rangle_W\langle j|_W \rho_X^{-\frac{1}{2}}\rho_{U\cup X}\rho_X^{-\frac{1}{2}}\rho^{\frac{1}{2}}_{X\cup W} \tag{55}$$

$$= \rho^{\frac{1}{2}}_{X\cup W}\rho_{U|X}\rho^{\frac{1}{2}}_{X\cup W} \tag{56}$$

By assumption, $\rho_{U|X\cup W} = \rho_{U|X}$, so it follows that $\rho_{U\cup X\cup W} = \rho^{\frac{1}{2}}_{X\cup W}\rho_{U|X}\rho^{\frac{1}{2}}_{X\cup W}$, as required. The result for $\rho_{W|X\cup U} = \rho_{W|X}$ follows by symmetry. □

For $n < \infty$, it is not true that (38) implies conditional independence, even in the case $n = 1$. This is illustrated by the following counterexample.

Example 3.7. Let U and W be single qubits, and X be composed of two qubits labeled X^L and X^R. For $\epsilon > 0$, consider the normalized state

$$\rho_{U\cup X\cup W} = \frac{4}{(1-\epsilon)^{\frac{1}{n}} + 3(\epsilon/3)^{\frac{1}{n}}}\rho_X \bigstar^{(n)} (P^-_{U\cup X^L}\otimes P^-_{W\cup X^R}), \tag{57}$$

where

$$\rho_X = (1-\epsilon)P^-_{X^L\cup X^R} + \frac{\epsilon}{3}P^+_{X^L\cup X^R} \tag{58}$$

and where $P^{\pm}_{A\cup B}$ denote the projector onto the symmetric and anti-symmetric subspaces of $\mathcal{H}_A\otimes\mathcal{H}_B$. The conditional states are

$$\rho^{(n)}_{U|X} = \frac{2}{\sqrt{(1-\epsilon)^{\frac{1}{n}} + 3(\epsilon/3)^{\frac{1}{n}}}} P^-_{U\cup X^L}\otimes I_{X^R} \text{ and} \tag{59}$$

$$\rho^{(n)}_{W|X} = \frac{2}{\sqrt{(1-\epsilon)^{\frac{1}{n}} + 3(\epsilon/3)^{\frac{1}{n}}}} I_{X^L}\otimes P^-_{W\cup X^R}. \tag{60}$$

By construction, condition (39) is easily verified $\rho_{U\cup X\cup W} = \rho_X \bigstar^{(n)} (\rho^{(n)}_{U|X}\rho^{(n)}_{W|X})$. In the limit $\epsilon\to 0$, the state $\rho_{U\cup X\cup W}\to P^-_{U\cup W}\otimes P^-_{X^L\cup X^R}$, which has $S(U:W|X) = 2$. By continuity, we

claim that for all $n < \infty$, there exists an $\epsilon > 0$ such that $\rho_{U\cup X\cup W}$ is a density operator that does not saturate strong subadditivity.

The preceding example shows that some of the conditions given in Eqs. (36)–(39) are not sufficient to imply quantum conditional independence on their own. Therefore, additional constraints need to be imposed in order to obtain converse results. Two alternative approaches are considered here, one based on additional commutation conditions that hold for conditionally independent states and one based on the algebraic structure of such states. The approach based on commutation conditions is perhaps more elegant, but the algebraic conditions are also relevant because they are used in Theorem 4.11 in Section 4.4 to provide a characterization result for quantum Markov Networks on trees. The following sequence of results provides the approach based on commutation conditions.

Theorem 3.8. *For a fixed n, if* $\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}$ *and its adjoint commute with* $\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}$, *then the conditions given in Eqs. (36)–(39) are all equivalent.*

Proof. We start by showing that $\rho^{(n)}_{U|W\cup X} = \rho^{(n)}_{U|X}$ is equivalent to $\rho^{(n)}_{W|U\cup X} = \rho^{(n)}_{W|X}$. The first of these can be written explicitly in terms of joint and reduced density operators as

$$\rho_{W\cup X}^{-\frac{1}{2n}}\rho_{U\cup W\cup X}^{\frac{1}{n}}\rho_{W\cup X}^{-\frac{1}{2n}} = \rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{n}}\rho_X^{-\frac{1}{2n}}. \tag{61}$$

Left and right multiplying by $\rho_{W\cup X}^{\frac{1}{2n}}$ gives

$$\rho_{U\cup W\cup X}^{\frac{1}{n}} = \rho_{W\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}. \tag{62}$$

Now, define $T = \rho_{W\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}$ so that $\rho_{U\cup W\cup X}^{\frac{1}{n}} = TT^\dagger$. In a similar fashion, $\rho^{(n)}_{W|U\cup X} = \rho_{W|X}$ can be shown to be equivalent to $\rho^n_{U\cup W\cup X} = T^\dagger T$. Now,

$$\begin{aligned} T^\dagger &= \rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}} && (63)\\ &= \rho_X^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}} && (64)\\ &= \rho_X^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}} && (65)\\ &= \rho_{W\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}} && (66)\\ &= T, && (67) \end{aligned}$$

where the assumption that $\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}$ commutes with $\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}$ has been used to derive Eq. (65). Hence, T is Hermitian and the two conditions are equivalent. For the remaining condition note that $\rho^{(n)}_{U\cup W|X} = \rho^{(n)}_{U|X}\rho^{(n)}_{W|X}$ is equivalent to

$$\begin{aligned} \rho_{U\cup W\cup X}^{\frac{1}{n}} &= \rho_{U\cup X}^{\frac{1}{n}}\rho_X^{-\frac{1}{n}}\rho_{W\cup X}^{\frac{1}{n}} && (68)\\ &= \rho_{U\cup X}^{\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}} && (69) \end{aligned}$$

The commutativity of $\rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}$ and $\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}$ then gives

$$\begin{aligned} \rho_{U\cup W\cup X}^{\frac{1}{n}} &= \rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}} && (70)\\ &= \rho_X^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}, && (71) \end{aligned}$$

and the commutativity of $\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}$ and $\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}$ gives

$$\rho_{U\cup W\cup X}^{\frac{1}{n}} = \rho_X^{+\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}} \tag{72}$$

$$= \rho_{W\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{n}}\rho_X^{-\frac{1}{n}}\rho_{W\cup X}^{\frac{1}{2n}}, \tag{73}$$

which is equivalent to $\rho^{(n)}_{U|W\cup X} = \rho^{(n)}_{U|X}$. □

Theorem 3.8 relates the conditions Eqs. (36)–(38) for a fixed value of n, but the conditions for different values of n can also be related via the following corollary.

Corollary 3.9. *For fixed n, if $\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}$ and its adjoint commute with $\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}$, then $\rho^{(n)}_{U|W\cup X} = \rho^{(n)}_{U|X}$ implies $\rho^{(2n)}_{U\cup W|X} = \rho^{(2n)}_{U|X}\rho^{(2n)}_{W|X}$.*

Proof. In the preceding proof it was shown that $\rho^{(n)}_{U|W\cup X} = \rho^{(n)}_{U|X}$ is equivalent to $\rho_{U\cup W\cup X}^{\frac{1}{n}} = TT^\dagger$, where $T = \rho_{W\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}$, and that the commutativity conditions imply that T is Hermitian. Therefore, $\rho_{U\cup W\cup X}^{\frac{1}{n}} = (T^\dagger)^2$, which implies $\rho_{U\cup W\cup X}^{\frac{1}{2n}} = T^\dagger = \rho_{U\cup X}^{\frac{1}{2n}}\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}$. The latter is straightforwardly equivalent to $\rho^{(2n)}_{U\cup W|X} = \rho^{(2n)}_{U|X}\rho^{(2n)}_{W|X}$. □

Putting these results together leads to a set necessary and sufficient condition for conditional independence.

Corollary 3.10. *If $\rho_X^{-\frac{1}{2n}}\rho_{U\cup X}^{\frac{1}{2n}}$ and its adjoint commute with $\rho_X^{-\frac{1}{2n}}\rho_{W\cup X}^{\frac{1}{2n}}$ for every n, then any of the conditions given in Eqs. (36)–(38) imply that $S(U:W|X) = 0$.*

Proof. Under these commutativity conditions, Theorem 3.8 implies that Eqs. (36)–(38) are equivalent for any fixed m and Corollary 3.9 shows that $\rho^{(2m)}_{U\cup W|X} = \rho^{(2m)}_{U|X}\rho^{(2m)}_{W|X}$ can be derived from $\rho^{(m)}_{U|W\cup X} = \rho^{(m)}_{U|X}$. By applying Theorem 3.8 with $n = 2m$, it follows that $\rho^{(m)}_{U|W\cup X} = \rho^{(m)}_{U|X}$ implies $\rho^{(2m)}_{U|W\cup X} = \rho^{(2m)}_{U|X}$. By induction, this implies that $\rho^{(2^s m)}_{U|W\cup X} = \rho^{(2^s m)}_{U|X}$ for any positive integer s. Taking the limit $s\to\infty$ gives $\rho^{(\infty)}_{U|W\cup X} = \rho^{(\infty)}_{U|X}$, which implies $S(U:W|X) = 0$ by Theorem 3.5. □

We now turn to the algebraic approach to proving converse results. Firstly, note that Eq. (38) implies that $\rho^{(n)}_{U|X}$ and $\rho^{(n)}_{W|X}$ commute, since $\rho^{(n)}_{U\cup W|X}$ is Hermitian. It can be shown that whenever two operators $A_{U\cup X}\otimes I_W$ and $I_U\otimes B_{W\cup X}$ commute there exists a decomposition of $\mathcal{H}_X$ as in Eq. (26) such that

$$A_{UX} = \sum_{j=1}^{d} a_{UX_j^L}\otimes I_{X_j^R} \text{ and} \tag{74}$$

$$B_{WX} = \sum_{j=1}^{d} I_{X_j^L}\otimes b_{X_j^R W}, \tag{75}$$

so Eq. (38) implies that $\rho^{(n)}_{U|X}$ and $\rho^{(n)}_{W|X}$ have this structure, as would be expected if the joint state is conditionally independent and hence satisfies Eq. (27). However, Eq. (27) implies an additional constraint that has not been used so far, namely that ρ_X also respects the same tensor product structure on $\mathcal{H}_X$, i.e. ρ_X is of the form

$$\rho_X = \sum_{j=1}^{d} p_j \sigma_{X_j^L} \otimes \tau_{X_j^R}. \tag{76}$$

More generally, we will say that an operator C_X is *decomposable with respect to* the pair of commuting operators $A_{U\cup X}$ and $B_{W\cup X}$ if it has the same algebraic structure on $\mathcal{H}_X$, i.e. if

$$C_X = \sum_{j=1}^{d} c_{X_j^R} \otimes c_{X_j^R}. \tag{77}$$

for some factorization of $\mathcal{H}_X$, such that Eqs. (74) and (75) hold. Imposing the commutativity of $\rho^{(n)}_{U|X}$ and $\rho^{(n)}_{W|X}$, along with the decomposability of ρ_X with respect to $\rho^{(n)}_{U|X}$ and $\rho^{(n)}_{W|X}$ as additional constraints is enough to straightforwardly show that any of Eqs. (36)–(39) imply conditional independence for all values of n.

4. Graphical Models

In this section, quantum conditional independence is used to define quantum Graphical Models that generalize their classical counterparts. The main focus is on quantum Markov Networks and n-Bifactor Networks, since these allow for the simplest formulation of the Belief Propagation algorithms to be described in Section 5. Section 4.1 reviews the definition of classical Markov Networks and the Hammersley–Clifford theorem, which gives an explicit representation for the probability distributions associated with classical Markov Networks. Motivated by this, Section 4.2 defines the class of quantum n-Bifactor Networks, which are the most general class of networks on which our Belief Propagation algorithms operate. Section 4.3 reviews the theory of dependency models and graphoids, which is useful for proving theorems about Graphical Models, and shows that quantum conditional independence can be used to define a graphoid. Section 4.4 defines quantum Markov Networks and gives some partial characterization results for the associated quantum states, along similar lines to the Hammersley–Clifford theorem. Most of these definitions and characterization results are summarized in Fig. 10.

The remaining two subsections briefly outline two other quantum Graphical Models: Quantum Factor Graphs in Section 4.5.1 and Quantum Bayesian Networks in Section 4.5.2. These structures are equivalent from the point of view of the efficiency of Belief Propagation algorithms, since it is always possible to convert them into n-Bifactor Networks and vice versa with only a linear overhead in graph size. An explicit method for converting a quantum factor graph into a quantum 1-Bifactor Network is given because factor graphs are used in the application to quantum error correction developed in Section 7.1.

4.1. Classical Markov Networks

Let $G = (V, E)$ be an undirected graph and suppose that each vertex $v \in V$ is associated with a random variable, also denoted v. Let $P(V)$ be the joint distribution of the variables. $(G, P(V))$ is a *Classical Markov Network* if for all $U \subseteq V$, $H(U : V - (n(U) \cup U)|n(U)) = 0$, where $n(U)$ is the set of nearest neighbors of U in G (see Fig. 1). Further, if $P(V)$ is strictly positive for all possible valuations of the variables, then $(G, P(V))$ is called a *Positive Classical Markov Network*. For such positive networks there is a powerful characterization theorem [17,9].

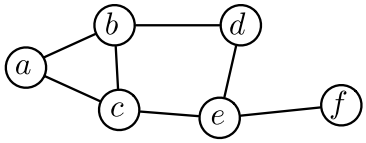

Fig. 1. The equalities $H(a:d\cup e\cup f|b\cup c)=0$, $H(f:a\cup b\cup c\cup d|e)=0$, and $H(a\cup b:e\cup f|c\cup d)=0$ are examples of constraints that are satisfied when $(G,P(V))$ is a Markov Network.

Theorem 4.1 (*Hammersley–Clifford* [18]). *$(G,P(V))$ is a positive classical Markov network iff it can be written as*

$$P(V)=\frac{1}{Z}\prod_{C\in\mathfrak{C}}\psi(C), \tag{78}$$

where $\mathfrak{C}$ is the set of cliques of G, $\psi(C)$ is a positive function defined on the random variables in C and Z is a normalization factor.

A set of vertices $C\subseteq V$ in a graph is a clique if $\forall u,v\in C,\ u\neq v\rightarrow(u,v)\in E$, i.e. every vertex in C is connected to every other vertex in C by an edge. Note that the decomposition in Eq. (78) is generally not unique, even up to normalization. A distribution of the form of Eq. (78) is said to factorize with respect to the graph G.

Markov chains are a special case of Markov Networks in which the graph is a chain. These are included in the slightly more general class of networks where the graph is a tree. For trees the only cliques are the individual vertices and the pairs of vertices that are connected by an edge, and the associated probability distributions have a representation in terms of marginal and mutual probability distributions of the form

$$P(V)=\prod_{v\in V}P(v)\prod_{(u,v)\in E}P(u:v), \tag{79}$$

which generalizes the decomposition for three variable Markov chain given in Eq. (24). For more general networks wherein the graph has cycles, there is no Hammersley–Clifford decomposition in which the functions $\psi(C)$ are marginal and mutual probability distributions.

The Hammersley–Clifford decomposition can be put in a form more familiar to physicists by introducing a positive constant β and defining the functions $H(C)=-\beta^{-1}\log\psi(C)$, which are always well defined since $\psi(C)$ is positive. Then Eq. (78) can be written as

$$P(V)=\frac{1}{Z}\exp\left(-\beta\sum_{C\in\mathfrak{C}}H(C)\right), \tag{80}$$

which is a Gibbs state for a system with a Hamiltonian $\sum_{C\in\mathfrak{C}}H(C)$ and partition function Z. This is a generalization of the lattice models studied in statistical physics to arbitrary graphs. If G is a cubic lattice for instance, then, as for trees, the only cliques are the individual vertices and pairs of vertices connected by an edge, so for such lattices the edges represent local nearest-neighbor interactions.

In many applications, such as in statistical physics, the functions $\psi(C)$ are often constants for cliques containing three or more vertices even in the case where the graph has cliques with more than two vertices. In this case, we again have that the only nontrivial functions are defined on the vertices and edges of the graph, so the state can be written as

$$P(V) = \frac{1}{Z}\prod_{v\in V}\psi(v)\prod_{(u,v)\in E}\psi(u:v). \tag{81}$$

Here, the edge functions are denoted $\psi(u:v)$ because of the close parallel with Eq. (79), but they are general positive functions rather than mutual distributions. We adopt the terminology *bifactor distribution* to describe distributions of the form of Eq. (81) and *Bifactor Network* for the pair $(G, P(V))$. For example, the distribution associated with a local nearest-neighbor model on an arbitrary graph, such as the spin glasses studied in statistical physics, would be a bifactor distribution.

4.2. Quantum Bifactor Networks

A proper generalization of Markov Networks to quantum theory involves the replacement of random variables with quantum systems and the replacement of classical conditional independence with its quantum counterpart. This theory is developed in the following sections, but it is convenient to first introduce a class of states that parallels the classical bifactor distributions of Eq. (81).

Let $G = (V, E)$ be a graph, let each vertex $v \in V$ be associated to a quantum system with Hilbert space $\mathcal{H}_v$. Let $\mathcal{H}_V = \otimes_{v\in V}\mathcal{H}_v$ and consider the class of states ρ_V that can be expressed as

$$\rho_V = \frac{1}{Z}\left(\bigotimes_{u\in V}\mu_u\right)\star^{(n)}\left(\left(\star^{(n)}\right)_{(v,w)\in E}\nu_{v:w}\right), \tag{82}$$

where Z is normalization constant, the μ_u's are operators on $\mathcal{H}_u$ and the $\nu_{v:w} = \nu_{w:v}$ are operators on $\mathcal{H}_v \otimes \mathcal{H}_w$. As stated, this expression is ambiguous because the $\star^{(n)}$ product is neither commutative or associative apart from in the limit $n \to \infty$. To avoid this ambiguity we impose the additional constraint that $[\nu_{u:v}, \nu_{w:x}] = 0$ for finite n, in which case the expression $(\star^{(n)})_{(v,w)\in E}\nu_{v:w}$ reduces to $\prod_{(v,w)\in E}\nu_{v:w}$. The state ρ_V is an *n-bifactor state* if it can be written as

$$\rho_V = \frac{1}{Z}\left(\bigotimes_{u\in V}\mu_u\right)\star^{(n)}\left(\prod_{(v,w)\in E}\nu_{v:w}\right), \tag{83}$$

with $[\nu_{u:v}, \nu_{w:x}] = 0$, and it is an ∞-*bifactor state* if it can be written as

$$\rho_V = \frac{1}{Z}\left(\bigotimes_{u\in V}\mu_u\right)\odot\left(\bigodot_{(v,w)\in E}\nu_{v:w}\right), \tag{84}$$

with no commutativity constraint on the $\nu_{v:w}$. The pair (G, ρ_V) is referred to as a quantum *n-Bifactor Network*, or ∞-*Bifactor Network*, respectively.

It turns out that not every quantum Bifactor Network is a quantum Markov Network, but the quantum generalizations of Belief Propagation algorithms to be developed in Section 5 can be formulated for any Bifactor Network. Therefore, readers who are mainly interested in algorithms and applications rather than proofs can skip to Section 5, perhaps pausing to read Section 4.5.1 on the way in order to understand the application to quantum error correction.

The next goal is to formulate the theory of quantum Markov Networks and provide characterization theorems analogous to the Hammersley–Clifford theorem. In order to

do so it is convenient to first introduce the theory of dependency models and graphoids, which is useful for proving theorems about Graphical Models.

4.3. Dependency models and graphoids

Graphs and conditional independence relations share a number of important properties that are responsible for the structure of Graphical Models. These properties are also shared by a number of other mathematical structures and they can be abstracted into structures known as dependency models and graphoids, which were introduced by Gieger, Verma, and Pearl [51,15]. Here, the theory is briefly reviewed and quantum conditional independence is shown to also give rise to a graphoid.

A *dependency model* M over a finite set V is a tripartite relation over disjoint subsets of V. The statement that $(U, W, X) \in M$ will be denoted $I(U, W|X)$, with a possible subscript on the I to denote the type of dependency model. $I(U, W|X)$ should be taken to mean that "U and W only interact via X", or that "U and W are independent given X".

Example 4.2. An *Undirected Graph Dependency Model* I_G is defined in terms of an undirected graph G. Let V be the set of vertices of G and then let $I_G(U, W|X)$ if every path from a vertex in U to a vertex in W passes through a vertex in X. I_G is often called the *Global Markov Property*.

Example 4.3. A *Probabilistic Dependency Model* I_P is defined in terms of a probability distribution $P(V)$ over a set V of random variables. $I_P(U, W|X)$ is true if U and W are conditionally independent given X.

Example 4.4. A *Quantum Dependency Model* I_ρ is defined in terms of a density operator ρ_V acting on the tensor product of Hilbert spaces labeled by elements of a set V. $I_\rho(U, W|X)$ is true if U and W are quantum conditionally independent given X.

A *graphoid* is a dependency model that for all disjoint $U, W, X, Y \subseteq V$ satisfies the following axioms:

$$\text{Symmetry:} \quad I(U, W|X) \Rightarrow I(W, U|X) \tag{85}$$

$$\text{Decomposition:} \quad I(U, W \cup Y|X) \Rightarrow I(U, W|X) \tag{86}$$

$$\text{Weak Union:} \quad I(U, W \cup Y|X) \Rightarrow I(U, W|X \cup Y) \tag{87}$$

$$\text{Contraction:} \quad I(U, W|X) \text{ and } I(U, Y|X \cup W) \Rightarrow I(U, W \cup Y|X). \tag{88}$$

A *positive graphoid* is a graphoid that also satisfies the additional axiom

$$\text{Intersection:} \quad I(U, W|X \cup Y) \text{ and } I(U, Y|W \cup X) \Rightarrow I(U, W \cup Y|X). \tag{89}$$

Theorem 4.5. *The quantum dependency model is a graphoid.*

Proof. Symmetry is immediate because $S(U : W|X)$ is invariant under exchange of U and W. Decomposition and Weak Union follow from the strong subadditivity inequality. Specifically, for $A, B, C \subseteq V$, strong subadditivity asserts that $S(A : B|C) \geqslant 0$, or in terms of von Neumann entropies

$$S(A \cup C) + S(B \cup C) - S(C) - S(A \cup B \cup C) \geqslant 0. \tag{90}$$

Decomposition asserts that if $S(U : W \cup Y|X) = 0$ then $S(U : W|X) = 0$. This is true if $S(U : W \cup Y|X) - S(U : W|X) \geqslant 0$, since $S(U : W|X)$ is guaranteed to be positive by strong subadditivity. Expanding $S(U : W \cup Y|X) - S(U : W|X)$ and canceling terms gives

$$\begin{aligned} &S(U : W \cup Y|X) - S(U : W|X) \\ &\quad = S(U \cup W \cup X) + S(W \cup X \cup Y) - S(W \cup X) - S(U \cup W \cup X \cup Y), \end{aligned} \tag{91}$$

but the right hand side is positive by Eq. (90) with $A = U, B = Y, C = W \cup X$.

Weak Union is proved via a similar argument applied to $S(U : W \cup Y|X) - S(U : W|X \cup Y)$. It follows from Eq. (90) by taking $A = U, B = Y, C = X$. Finally, contraction follows from noting that $S(U : W|X) + S(U : Y|X \cup W) = S(U : W \cup Y|X)$, which is straightforward to show by expanding in terms of von Neumann entropies. □

The well-known analogous result for classical probability distributions follows immediately because classical probability distributions can be represented by density matrices that are diagonal in an orthonormal product basis, and for such states the von Neumann entropies of subsystems are equal to the Shannon entropies of the corresponding marginal distributions. Additionally, if $P(V)$ is positive for all possible valuations of the variables then the associated dependency model is actually a positive graphoid. The analogous quantum property would be to require that ρ_V is a strictly positive operator, i.e. it is of full rank, but we have not been able to prove that this property implies intersection.

The undirected graph dependency model is also a positive graphoid. The proof is straightforward, so it is not given here. The following theorem is important for the theory of Markov networks.

Theorem 4.6 (Lauritzen [25]). *The undirected graph dependency model is equivalent to the dependency model obtained by setting $I(U, V - (U \cup n(U))|n(U))$ for all $U \subseteq V$, where $n(U)$ is the set of nearest neighbors of U, and demanding closure under the positive graphoid axioms.*

The condition $I(U, V - (U \cup n(U))|n(U))$ defines the *Local Markov Property* on a graph. Note that although its closure under the positive graphoid axioms is equivalent to the Global Markov Property, this is not the case for a graphoid that does not satisfy intersection [25].

4.4. Quantum Markov Networks

Using the terminology of the previous section, the definition of a classical Markov Network can be conveniently reformulated as a pair $(G, P(V))$, where $G = (V, E)$ is an undirected graph and $P(V)$ is a probability distribution over random variables represented by the vertices, such that the graphoid I_P satisfies the local Markov property with respect to the graph G. The definition of a quantum Markov network can now be obtained by replacing the probabilistic dependency model with a quantum dependency model.

Let $G = (V, E)$ be an undirected graph and suppose that each vertex $v \in V$ is associated with a quantum system, also denoted v, with Hilbert space $\mathcal{H}_v$. Let ρ_V be a state on $\mathcal{H}_V = \otimes_{v \in V} \mathcal{H}_v$. (G, ρ_V) is a *Quantum Markov Network* if the graphoid I_ρ satisfies the local Markov property with respect to the graph G. Further, if ρ_V is of full rank, then (G, ρ_V) is called a *Positive Quantum Markov Network*. Note that unlike in the classical case, we can-

not conclude that the global Markov property holds for positive quantum Markov networks because the intersection axiom has not been proved.

The remainder of this section provides some partial characterization results for quantum Markov networks, along the lines of the Hammersley–Clifford theorem. The most generally applicable of these results makes use of the $\odot$ product.

Theorem 4.7. *Let $G = (V, E)$ be an undirected graph and let $\mathfrak{C}$ be the set of cliques of G. If (G, ρ_V) is a positive quantum Markov network then there exist positive operators σ_C acting on the cliques of G, i.e. $C \in \mathfrak{C}$, such that*

$$\rho_V = \bigodot_{C \in \mathfrak{C}} \sigma_C. \tag{92}$$

This theorem is analogous to one direction of the Hammersley–Clifford theorem and the proof is very similar to a standard proof for the classical case [41], but is somewhat involved so it is given in Appendix B. However, unlike the classical case, the converse does not hold, i.e. there are states of the form Eq. (92) that do not satisfy the local Markov property as illustrated by the following example.

Example 4.8. Consider a chain of 3 qubits A, B, and C coupled through an anti-ferromagnetic Heisenberg interaction $H = \sigma_A^x \sigma_B^x I_C + \sigma_A^y \sigma_B^y I_C + \sigma_A^z \sigma_B^z I_C + I_A \sigma_B^x \sigma_C^x + I_A \sigma_B^y \sigma_C^y + I_A \sigma_B^z \sigma_C^z$ where σ^x, σ^y, and σ^z denote the Pauli operators

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \text{and } \sigma^y = \sigma^z \sigma^x. \tag{93}$$

The Gibbs state $\rho_{A \cup B \cup C}(\beta) = \frac{1}{Z(\beta)} \exp(-\beta H)$ has the form Eq. (92), but for any finite β it has a nonzero mutual information between A and C conditioned on B as shown in Fig. 2.

For trees, a decomposition into reduced and mutual density operators analogous to Eq. (79) is possible. For this, we need the following lemma.

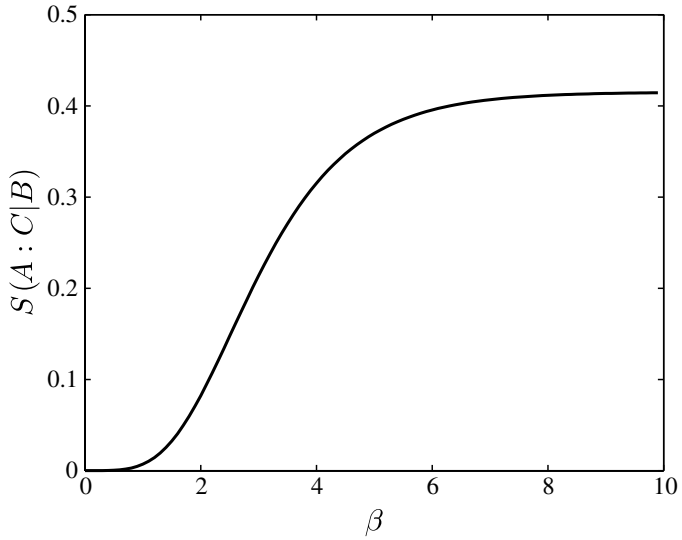

Fig. 2. Conditional mutual information for a three-vertex anti-ferromagnetic Heisenberg spin-$\frac{1}{2}$ chain as a function of inverse temperature β.

Lemma 4.9. *Let $G = (V, E)$ be a graph, let (G, ρ_V) be a quantum Markov network and let $u \in V$. Let $G' = (V', E')$ be the graph obtained by removing u from V and removing all edges that connect u to any other vertex from the graph. Let $G'' = (V', E'')$ be the graph obtained by adding to G' an edge between every pair of distinct neighbors of u in the original graph G. Let $\rho_{V'} = \mathrm{Tr}_u(\rho_{V'})$. Then $(G'', \rho_{V'})$ is a quantum Markov network.*

Proof. For $U \subset V$, let $U_u = U - u$ if $u \in U$ and $U_u = U$ otherwise, and denote $n_G(U_u)$ and $n_{G''}(U_u)$ the neighbors of U_u in the graphs G and G'', respectively. It must be shown that $I_{\rho_V}(U, V - (U \cup n_G(U))|n_G(U))$ for all $U \subset V$ implies $I_{\rho_{V'}}(U_u, V' - (U_u \cup n_{G''}(U_u))|n_{G''}(U_u))$ for every $U_u \subset V'$. By symmetry, we can assume without loss of generality that $u \in U$. There are two different cases to consider:

Case I: $n_G(u) \cap U \neq \emptyset$.
This implies that $n_{G''}(U_u) = n_G(U)$ and so $V' - (U_u \cup n_{G''}(U_u)) = V - (U \cup n_G(U))$. We conclude that $I_{\rho_{V'}}(U_u, V' - (U_u \cup n_{G''}(U_u))|n_{G''}(U_u))$ is equivalent to $I_{\rho_V}(U - u, V - (U \cup n_G(U))|n_G(U))$, and the result follows from decomposition.

Case II: $n_G(u) \cap U = \emptyset$.
This implies that $n_{G''}(U_u) = n_G(U_u)$. Consider the local Markov property on the original graph G applied to U_u: $I_{\rho_V}(U_u, V - (U_u \cup n_G(U_u))|n_G(U_u))$ which is equivalent to $I_{\rho_V}(U_u, u \cup V' - (U_u \cup n_{G''}(U_u))|n_{G''}(U_u))$, and the result follows from decomposition. □

Theorem 4.10. *Let $G = (V, E)$ be a tree. If (G, ρ_V) is a positive quantum Markov network then it can be written as*

$$\rho_V = \left(\bigotimes_{v \in V} \rho_v\right) \star^{(n)} \left(\prod_{(v,u) \in E} \rho^{(n)}_{v:u}\right). \tag{94}$$

Proof. The proof is by induction on the number of vertices in the tree. It is clearly true for a single vertex, so consider a tree $G = (V, E)$ with N vertices and choose a leaf vertex $u \in V$. Construct the quantum Markov network $(G'', \rho_{V'})$ as in Lemma 4.9. Since u is a leaf it only has one neighbor in G, denoted w, so the only difference between G and G'' is that u and the single edge connecting u to the rest of the graph have been removed. By the inductive assumption, $\rho_{V'}$ has a decomposition of the form

$$\rho_{V'} = \left(\bigotimes_{v \in V'} \rho_v\right) \star^{(n)} \left(\prod_{(v,x) \in E''} \rho^{(n)}_{v:u}\right). \tag{95}$$

Generally, $\rho_V = \rho_{V' \cup \{u\}} = \rho_{V'} \star^{(n)} \rho^{(n)}_{u|V'}$. The local Markov property implies that $I_\rho(u, V' - w|w)$, so that $\rho_{u|V'} = \rho_{u|w}$, which in turn can be written as $\rho_{u|w} = \rho_u \star^{(n)} \rho_{u:w}$, so

$$\rho_V = \rho_{V'} \star^{(n)} (\rho_u \star^{(n)} \rho_{u:w}). \tag{96}$$

Every term in Eq. (95) commutes with ρ_u, because they are defined on different tensor product factors. Also, $\rho_{u:w}$ commutes with all the other mutual density operators either because they act on different tensor product factors or because the fact that w is the only

neighbor of u implies that u is quantum conditionally independent of any other subsystem given w. □

In the classical case, the Hammersley–Clifford decomposition is not necessarily unique, and when the graph is a tree the decomposition into marginal and mutual distributions is only one possibility. Similarly, a state ρ_V might have a decomposition of the form of Eq. (94) but with more general operators in place of the mutual and marginal states. This provides another motivation for the definition of an n-bifactor state that was given in Eq. (83). As mentioned in Section 4.2, not all n-bifactor states are quantum Markov networks, but a subset of them are, as shown by the following theorem.

Theorem 4.11. *Let $G = (V, E)$ be a tree with each vertex $v \in V$ associated to a quantum system with Hilbert space $\mathcal{H}_v$. Let $\mathcal{H}_V = \otimes_{v\in V}\mathcal{H}_v$ and let ρ_V be an n-bifactor state on $\mathcal{H}_V$. If μ_v is decomposable with respect to all pairs $v_{u:v}$ and $v_{w:v}$, then (G, ρ_V) is a quantum Markov network.*

The notion of decomposability used in the statement of this theorem is defined at Eq. (77). The proof is straightforward and we leave it as an exercise.

4.5. Other Graphical Models

In this section, quantum generalizations of two other Graphical Models are described: Factor Graphs and Bayesian Networks. Generally, the choice of which model to use depends on the application and Belief Propagation algorithms have been developed for all of them in the classical case. For example, Factor Graphs arise naturally in the theory of error correcting codes, Bayesian Networks are commonly used to model causal reasoning in artificial intelligence, and Markov Networks are useful in statistical physics. However, it is now understood that the classical versions of these three models are interconvertable, and that upon such conversion the different Belief Propagation algorithms are all equivalent in complexity [6,59,22]. Some similar results also hold for the quantum case, as we illustrate by showing how a quantum factor graph can be converted into a 1-Bifactor Network. This construction is used in the application to quantum error correction described in Section 7.1.

4.5.1. Quantum factor graphs

A *quantum factor graph* consists of a pair (G, ρ_V), where $G = (U, E)$ is a bipartite graph and ρ_V is a quantum state. A bipartite graph is an undirected graph for which the set of vertices can be partitioned into two disjoint sets, V and F, such that $(v, f) \in E$ only if $v \in V$ and $f \in F$. The vertices in V are referred to as "variable nodes" and those in F as "function nodes". Each variable node v is associated with a quantum system, also labeled v, with a Hilbert space $\mathcal{H}_v$, and ρ_V is a state on $\otimes_{v\in V}\mathcal{H}_v$. The Hilbert space associated to a function node f is the tensor product of the Hilbert spaces of the adjacent variable nodes[2]: $\mathcal{H}_f = \otimes_{v\in n(f)}\mathcal{H}_v$. The state associated with a factor graph is of the form

$$\rho_V = \frac{1}{Z}\prod_{f\in F} X_f \star \bigotimes_{v\in V} \mu_v \tag{97}$$

where μ_v is an operator on $\mathcal{H}_v$, X_f is an operator on $\mathcal{H}_f$ and $[X_f, X_g] = 0$.

[2] The following equality is not just meant in the sense of an isomorphism, they are the same Hilbert spaces.

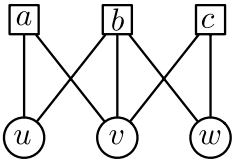

Fig. 3. Factor graph representation of the state $(|000\rangle + |111\rangle)_{uvw}$, with $\mu_u = \mu_v = \mu_w = I$ and $X_a = (I + \sigma_u^z \otimes \sigma_v^z)$, $X_b = (I + \sigma_u^x \otimes \sigma_v^x \otimes \sigma_w^x)$, and $X_c = (I + \sigma_v^z \otimes \sigma_w^z)$.

For example, such a state would be obtained after performing a sequence of projective von Neumann measurements on a product state of the variable nodes (see Fig. 3). More precisely, for each $f \in F$, let $\{P_f^j\}$ be a complete set of orthogonal projectors, and let $\otimes_{v\in V}\mu_v$ be the initial state of V. When the projective measurements $\{P_f^j\}$ are performed at each function node and commuting outcomes $P_f^j = X_f$ are obtained, the post-measurement state is of the form of Eq. (97). Similarly, factor graph states could be obtained from more general POVM measurements $\{E_f^j\}$, provided the state update rule $\rho_V \to \frac{(E_f^j)^{\frac{1}{2}}\rho_V(E_f^j)^{\frac{1}{2}}}{\mathrm{Tr}(E_f^j\rho_V)}$ is used. In that case, the X_f could be any positive operator rather than being restricted to projectors as in the case of a von Neumann measurement.

To convert a factor graph into a 1-Bifactor Network, we need to treat the function nodes as distinct quantum systems, and so endow them with their own Hilbert spaces $\mathcal{H}_f = \otimes_{v\in n(f)}\mathcal{H}_{R_v^f}$ where $\mathcal{H}_{R_v^f}$ is isomorphic to $\mathcal{H}_v$. The system R_v^f is called a reference system for v in f. Then, the state of the function nodes can be written on the graph $G = (U, E)$, where $U = V \cup F$, $\rho_V = \mathrm{Tr}_F(\rho_U)$ and

$$\rho_U = \frac{1}{Z}\bigotimes_{u\in U}\mu_u\star\prod_{(v,f)\in E}v_{v:f}, \tag{98}$$

where for $u \in F$, $\mu_u = X_u^T$, $v_{v:f} = d_v|\Phi\rangle\langle\Phi|_{v\cup R_v^f} \otimes I_{f-R_v^f}$ and $|\Phi\rangle_{v\cup R_v^f} = \frac{1}{\sqrt{d_v}}\sum_{j=1}^{d_v}|j\rangle_v|j\rangle_{R_v^f}$ denotes the maximally entangled state between v and its reference R_v^f.

4.5.2. Quantum Bayesian Networks

Apart from Markov Networks, there are other Graphical Models that make use of the theory of dependency models and graphoids. Bayesian Networks provide an example, and they are commonly applied in expert systems to model causal reasoning [36,37]. The basic idea is to replace the undirected graph of a Markov network with a Directed Acyclic Graph (DAG), wherein the directed edges represent direct cause-effect relationships. The quantum graphoid can be used to give a straightforward generalization of the classical networks, which we only treat briefly here. To describe the generalization, a few definitions and facts about DAGs are required.

For a vertex v in a DAG $G = (V, E)$, let $m(v)$ denote the parents of v, i.e. $m(v) = \{u \in V|(u, v) \in E\}$. The set of ancestors of v is denoted $a(v)$ and consists of those vertices u for which there exists a path in the graph starting at u and ending at v. Conversely, the set of descendants of v is denoted $d(v)$ and consists of those vertices u for which there exists a path in the graph starting at v and ending at u. The set of parents of a subset $U \subseteq V$ of vertices is defined as $m(U) = \cup_{u\in U}m(u) - U$ and similarly $a(U) = \cup_{u\in U}a(u) - U$ and $d(U) = \cup_{u\in U}d(u) - U$. The set of nondescendants of a subset $U \subseteq V$ of vertices is defined to be $nd(U) = V - (d(U) \cup U)$. Note that the vertices in U

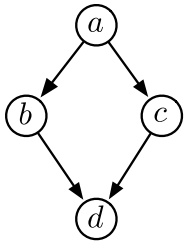

Fig. 4. This directed acyclic graph has two distinct ancestral orderings: (a, b, c, d) and (a, c, b, d). The equalities $S(d : a|b \cup c) = 0$ and $S(b : d \cup d|a) = 0$ are examples of constraints that are satisfied when (G, ρ_V) is a Quantum Bayesian Network.

are not considered to be nondescendants of U for technical convenience. Finally, every DAG has at least one ancestral ordering of its vertices $(v_1, v_2, \ldots, v_n)$, such that if $v_j \in a(v_k)$ then $j < k$ (see Fig. 4).

A *Quantum Bayesian Network* is a pair (G, ρ_V), where $G = (V, E)$ is a DAG, each vertex $v \in V$ is associated with a quantum system, also denoted v, with Hilbert space $\mathcal{H}_v$, and ρ_V is a quantum state on $\mathcal{H}_V = \otimes_{v \in V} \mathcal{H}_v$. The state ρ_V satisfies the conditional independence constraints $I_\rho(U, nd(U) - m(U)|m(U))$ for all subsets $U \subseteq V$ (see Fig. 4).

The definition of a classical Bayesian Network is obtained by replacing the quantum systems with classical random variables. It can be shown that $(G, P(V))$ is a classical Bayesian Network iff $P(V) = \prod_{v \in V} P(v|m(v))$, and a partial quantum generalization of this can be obtained using the conditional density operator.

Due to the nonassociativity of the $\star^{(n)}$ products, expressions like $A\star^{(n)}B\star^{(n)}C$ are ambiguous. It is convenient to adopt the convention that they are evaluated left-to-right, so that $A\star^{(n)}B\star^{(n)}C = (A\star^{(n)}B)\star^{(n)}C$. Similarly, we adopt the convention that

$$\left(\star^{(n)}\right)_{j=1}^{N} A_j = \left(\left(\left(A_1\star^{(n)}A_2\right)\star^{(n)}A_3\right)\ldots\right)\star^{(n)}A_N. \tag{99}$$

Theorem 4.12. *If (G, ρ_V) is a Quantum Bayesian Network and $(v_1, v_2, \ldots, v_N)$ is an ancestral ordering of V then*

$$\rho_V = \left(\star^{(n)}\right)_{j=1}^{N} \rho^{(n)}_{v_j|m(v_j)}. \tag{100}$$

Proof. For any ordering $(v_1, v_2, \ldots, v_N)$ of the vertices, an arbitrary state can always be written as

$$\rho_V = \left(\star^{(n)}\right)_{j=1}^{N} \rho^{(n)}_{v_j|v_{j-1}v_{j-2}\ldots v_1}. \tag{101}$$

This is a quantum generalization of the chain rule for conditional probabilities, which follows straightforwardly from the definition of conditional density operators. If $(v_1, v_2, \ldots, v_N)$ is in fact an ancestral ordering, then $\{v_{j-1}, v_{j-2}, \ldots, v_1\} \subseteq nd(v_j)$, so $I_\rho(v_j, nd(v_j)|m(v_j))$ implies that $\rho^{(n)}_{v_j|v_{j-1}v_{j-2}\ldots v_1} = \rho^{(n)}_{v_j|m(v_j)}$. □

5. Quantum Belief Propagation

In this section, we discuss algorithms for solving the inference problem that we started with in Section 2 for the case of n-Bifactor Networks. In fact, we start with the seemingly

simpler problem of computing the reduced density operators of the state on the vertices and on pairs of vertices connected by an edge, and then present a simple modification of the algorithm to solve the inference problem for local measurements.

Recall that n-bifactor states are of the form

$$\rho_V = \frac{1}{Z}\left(\bigotimes_{u\in V}\mu_u\right)\star^{(n)}\left(\prod_{(v,w)\in E} v_{v:w}\right), \tag{102}$$

and that the operators associated with vertices and edges do not have to be straightforwardly related to the reduced and mutual density operators. Therefore, it is not clear a priori that even the simpler task can be done efficiently. *Quantum Belief Propagation* (QBP) algorithms are designed to solve this problem by exploiting the special structure of n-bifactor states. Since the class of states under consideration is different for each value of n, there is not one but a family of algorithms. The algorithm that is designed to solve inference problems on n-Bifactor Networks is denoted $\mathrm{QBP}^{(n)}$.

To avoid cumbersome notation, focus will be given to n-bifactor states with $n < \infty$. Recall that the operators $v_{u:v}$ defining these states mutually commute. This is not true of ∞-bifactor states. Nevertheless, a Belief Propagation algorithm for ∞-bifactor states can be readily defined from the finite n one, by replacing *all* products appearing in Eqs. (103)–(105) by the $\odot$ product. Under this modification, the convergence Theorem 5.6 applies to ∞-Bifactor Networks, and its proof only requires straightforward modifications.

The remainder of this section is structured as follows. Section 5.1 gives a description of the QBP algorithms and Section 5.2 shows that $\mathrm{QBP}^{(n)}$ converges on trees if the n-Bifactor Network is also a quantum Markov Network and that $\mathrm{QBP}^{(1)}$ converges on trees in general. In both cases, the algorithm converges in a time that scales linearly with the diameter of the tree. Finally, Section 5.3 explains how to modify the algorithm to solve inference problems for local measurements.

5.1. Description of the algorithm

To describe the operation of the QBP algorithms, it is helpful to imagine that the graph G represents a network of computers with a processor situated at each vertex. The algorithm could equally well be implemented on a single processor, in which case the network is just a convenient fiction. Pairs of processors are connected by a communication channel if there is an edge between the corresponding vertices. The processor at vertex u has a memory that stores the value of μ_u as well as the value of $v_{u:v}$ for each vertex v that is adjacent to u in the graph. The task assigned to each processor is to compute the local reduced state ρ_u and the joint states $\rho_{u\cup v}$.[3] At each time step t, the processor at u updates its "beliefs" about ρ_u and $\rho_{u\cup v}$ via an iterative formula. These beliefs are denoted $b_u^{(n)}(t)$ and $b_{uv}^{(n)}(t)$, and are supposed to be approximations to the true reduced states ρ_v and $\rho_{u\cup v}$ based on the information available to the processor at time step t. Since the reduced states may depend on information stored at other vertices, the processors pass operator valued messages $m_{u\to v}^{(n)}(t)$ along the edges at each time step in order to help their neighbors.

[3] Of course, it would be sufficient to only have one processor compute $\rho_{u\cup v}$ for each edge.

The message $m^{(n)}_{u\to v}(t)$ is an operator on $\mathcal{H}_v$ and is initialized to the identity operator $m^{(n)}_{u\to v}(0) = I_v$ at $t = 0$. For $t > 0$ it is computed via the iterative formula

$$m^{(n)}_{u\to v}(t) = \frac{1}{Y}\mathrm{Tr}_u\left(\mu_u \star^{(n)} \left[\left\{\prod_{v'\in n(u)-v} m^{(n)}_{v'\to u}(t-1)\right\}\star^{(n)} \nu_{u:v}\right]\right). \tag{103}$$

Here, Y is an arbitrary normalization factor that should be chosen to prevent the the matrix elements of $m^{(n)}_{u\to v}(t)$ becoming increasingly small as the algorithm proceeds. It is convenient to choose Y such that $\mathrm{Tr}_v(m^{(n)}_{u\to v}(t)) = 1$.

The beliefs about the local density operator ρ_u at time t are given by the simple formula

$$b^{(n)}_u(t) = \frac{1}{Y'}\mu_u \star^{(n)} \prod_{v'\in n(u)} m^{(n)}_{v'\to u}(t), \tag{104}$$

where Y' is again a normalization factor that should be chosen to make $\mathrm{Tr}_u(b^{(n)}_u(t)) = 1$. On the other hand, the beliefs about $\rho_{u\cup v}$ also depend on the messages received by the processor at v, so we have to imagine that each vertex shares its messages with its neighbors. Having done so, the beliefs about $\rho_{u\cup v}$ are computed via (see Fig. 5)

$$b^{(n)}_{uv}(t) = \frac{1}{Y''}(\mu_u\mu_v)\star^{(n)}\left[\left\{\prod_{w\in n(u)-v} m^{(n)}_{w\to u}(t)\prod_{w'\in n(v)-u} m^{(n)}_{w'\to v}(t)\right\}\star^{(n)}\nu_{u:v}\right], \tag{105}$$

where Y'' is again a normalization factor.

The beliefs obtained from the $\mathrm{QBP}^{(n)}$ algorithm on input $\{\mu_u\}_{u\in V}$ and $\{\nu_{u:v}\}_{(u,v)\in E}$ after t time steps are denoted $[b^{(n)}_u(t), b^{(n)}_{uv}(t)] = \mathrm{QBP}^{(n)}_t(\mu_u, \nu_{u:v})$. The goal of the next section is to provide conditions under which the beliefs represent the exact solution to the inference problem, i.e. to find states and values of t such that $\mathrm{QBP}^{(n)}_t(\mu_u, \nu_{u:v}) = [\rho_u, \rho_{u\cup v}]$.

5.2. Convergence on trees

At time t, the beliefs $b^{(n)}_u(t)$ and $b^{(n)}_{uv}(t)$ represent estimates of the reduced states ρ_u and $\rho_{u\cup v}$ of the input n-bifactor state ρ_V. Note that when the μ_u and the $\nu_{u:v}$ all commute with one another and are diagonal in local basis, the $\mathrm{QBP}^{(n)}$ algorithms all coincide for different n (including $n = \infty$) and correspond to the well-known classical Belief Propagation algorithm. This algorithm always converges on trees in a time that scales like the diameter of the tree. Its convergence on general graphs is not fully understood and constitutes an active area of research [58,59]. In the quantum setting, the μ_u and the $\nu_{u:v}$ do not commute in general, but for finite n, the $\nu_{u:v}$ commute with each other by assumption. This has straightforward consequence that will be of use later.

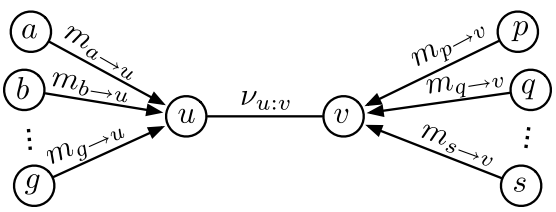

Fig. 5. Belief b_{uv} is a function of μ_u, μ_v, $\nu_{u:v}$, and the incoming messages at vertices u and v, except $m_{u\to v}$ and $m_{v\to u}$.

Proposition 5.1. *For all* $u, v \in V$, $x \in n(u)$, *and* $w \in n(v)$, *the following commutation relations hold* $[v_{u:v}, m^{(n)}_{x\to u}(t)] = 0$ *and* $[m^{(n)}_{w\to v}(t), m^{(n)}_{x\to u}(t)] = 0$.

Before proving the convergence of Quantum Belief Propagation, the following classical example can help build intuition of its workings, and also serves to outline the crucial steps in proving convergence.

Example 5.2. Consider the function P of N discrete variables $x_j \in \{1, 2, \ldots, d\}$

$$P(x_1, x_2, \ldots, x_N) = \psi(x_1, x_2)\psi(x_2, x_3)\ldots\psi(x_{N-1}, x_N) \tag{106}$$

which could be for instance a classical bifactor distribution on a chain with N sites. To evaluate the marginal function $P(x_N) = \sum_{x_1, x_2, \ldots, x_{N-1}} P(x_1, x_2, \ldots, x_N)$, one can proceed directly and carry the sum over d^N terms. A more efficient solution is obtained by invoking the distributive law to reorder the various sums and products into

$$P(x_N) = \sum_{x_{N-1}} \left(\psi(x_{N-1}, x_N) \left(\ldots \left(\sum_{x_2} \psi(x_2, x_3) \left(\sum_{x_1} \psi(x_1, x_2) \right) \right) \ldots \right) \right),$$

and performing the sums sequentially, starting with $\sum_{x_1}$, then $\sum_{x_2}$, and so on

$$\begin{aligned} P(x_N) &= \sum_{x_{N-1}} \left(\psi(x_{N-1}, x_N) \left(\ldots \left(\sum_{x_2} \psi(x_2, x_3) M_{1\to 2}(x_2) \right) \ldots \right) \right) \\ &= \sum_{x_{N-1}} (\psi(x_{N-1}, x_N)(\ldots M_{2\to 3}(x_3)\ldots)) \\ &\quad \vdots \\ &= \sum_{x_{N-1}} \psi(x_{N-1} : x_N) M_{N-2\to N-1}(x_{N-1}) \end{aligned}$$

where the "messages" are defined recursively $M_{j\to j+1}(x_{j+1}) = \sum_{x_j} \psi(x_j : x_{j+1}) M_{j-1\to j}(x_j)$, with $M_{1\to 2} = \sum_{x_1} \psi(x_1 : x_2)$. Each of these steps involves the sum of d^2 terms, so $P(x_N)$ can be computed with order Nd^2 operations.

This example differs from the Belief Propagation algorithm described in the previous section in three important aspects. Firstly, it relied on the distributive law, which does not hold in general for the $\star^{(n)}$ product, i.e. $\mathrm{Tr}_u(X_{uv}\star^{(n)} Y_{vw}) \neq \mathrm{Tr}_u(X_{uv})\star^{(n)} Y_{vw}$ in general. This will motivate Theorems 5.4 and 5.5, that establish necessary conditions for the validity of the distributive law. Secondly, the graph in that example is a chain, whereas Belief Propagation operates on any graph. However, Belief Propagation is only guaranteed to converge on trees, and the above example generalizes straightforwardly to such graphs. Thirdly, the messages in the example must be computed in a prescribed order: $M_{i-1\to i}$ is required to compute $M_{i\to i+1}$. This last point is important and deserves an extensive explanation.

Suppose that instead of computing the messages $M_{i\to i+1}$ sequentially, messages at each vertex were computed at every time step, following the rule $m_{i\to i\pm 1}(t, x_{i\pm 1}) = \sum_{x_i} m_{i\mp 1\to i}(t-1, x_i)\psi(x_i : x_{i\pm 1})$, as in Eq. (103), with the initialization $m_{i\pm 1\to i}(0, x_i) = 1$. Then, one can easily verify that for $t \geqslant i$, $m_{i\to i+1}(t, x_{i+1}) = M_{i\to i+1}(x_{i+1})$. In other words,

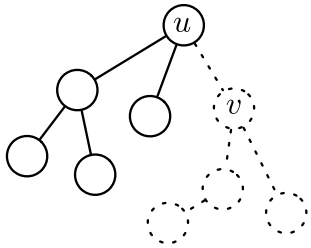

Fig. 6. For $(u,v) \in E$, the graph G_v^u is obtained from G by considering u as the root and removing the subtree associated to vertex v. In this example, $depth(G_v^u) = 2$.

the messages $m_{i\to i+1}$ become time independent after a time equal to the distance between vertex i the beginning of the chain. This observation can in fact be generalized as follows.

Lemma 5.3. *When G is a tree, the* QBP$^{(n)}$ *messages $m^{(n)}_{u\to v}(t)$ are time independent for $t > depth(G_v^u)$, where G_v^u is the tree obtained from G by choosing u as the root, and removing the subtree associated to v (see Fig. 6).*

Proof. The proof is by induction. If u is a leaf, it has a unique neighbor $n(u)$ and $m^{(n)}_{u\to n(u)}(t) = \mathrm{Tr}_u(\mu_u \star^{(n)} v_{u:n(u)})$ which is time independent. When u is not a leaf, then $m^{(n)}_{u\to v}(t)$ for $v \in n(u)$ is a function of $m^{(n)}_{v'\to u}(t-1)$ for $v' \in n(u) - v$. If those messages are time independent for $t \geqslant t^*$, then $m^{(n)}_{u\to v}(t)$ becomes time independent for $t \geqslant t^* + 1$, and the proof follows by induction. □

When operated on a tree, all beliefs computed by QBP algorithm converge to a steady state after a time equal to the diameter of the tree. Note that when the graph contains loops, the beliefs do not necessarily reach a steady state. It remains to be shown that on trees, this steady state is the correct solution. For this, we need a technical result that requires some new notation. Let U and W be two nonintersecting subsets of V. Define the two subsets of edges $E_U = \{(u,w) \in E : u,w \in U\}$ and $E_{U:W} = \{(u,w) \in E : u \in U \text{ and } w \in W\}$. Let $\Gamma_U = \otimes_{u\in U}\mu_u$ and for any $F \subset E$, let $\Lambda_F = \prod_{(u,w)\in F} v_{v:w}$.

Theorem 5.4. *Let (G,ρ_V) be an n-Bifactor Network with graph $G = (V,E)$. Let U,W,X be nonintersecting subsets of V such that $U \cup W \cup X = V$. When $S(U:X|W) = 0$, the following diagram is commutative.*

$$\begin{array}{ccc} \Gamma_{U\cup W}\star^{(n)}(\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}}) & \overset{\mathrm{Tr}_U}{\to} & \mathrm{Tr}_U(\Gamma_{U\cup W}\star^{(n)}(\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}})) \\ \downarrow \Gamma_X\star^{(n)}(\cdot\Lambda_{E_X}) & & \downarrow \Gamma_X\star^{(n)}(\cdot\Lambda_{E_X}) \\ \rho_V = \Gamma_V\star^{(n)}\Lambda_{E_V} & \overset{\mathrm{Tr}_U}{\to} & \mathrm{Tr}_U(\rho_V). \end{array} \tag{107}$$

Proof. The down-right path is the simplest. The first equality follows from the fact that Λ_{E_X} commutes with $\Gamma_{U\cup W}$ and all other Λ_E's, and the definition $\rho_V = (\Gamma_{U\cup W}\otimes\Gamma_X)\star^{(n)}(\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}}\Lambda_{E_X})$. The second equality is just a definition. The right-down path uses the representation of states that saturate strong subadditivity Eq. (27), which implies that ρ_V has a decomposition of the form $\rho_V = \sum_{j=1}^d p_j \sigma_{UW_j^{(1)}} \otimes \tau_{W_j^{(2)}X}$. First observe that

$$\Gamma_{U\cup W}\star^{(n)}(\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}}) = (\Gamma_X^{-1}\star^{(n)}\rho_V)\Lambda_{E_X}^{-1} \tag{108}$$

$$= \left(\Gamma_X^{-1}\star^{(n)}\sum_{j=1}^{d} p_j\sigma_{UW_j^{(1)}}\otimes\tau_{W_j^{(2)}X}\right)\Lambda_{E_X}^{-1} \tag{109}$$

$$= \sum_{j=1}^{d} p_j\sigma_{UW_j^{(1)}}\otimes\left[(\Gamma_X^{-1}\star^{(n)}\tau_{W_j^{(2)}X})\Lambda_{E_X}^{-1}\right]. \tag{110}$$

It follows that

$$\Gamma_X\star^{(n)}\left[\mathrm{Tr}_U(\Gamma_{U\cup W}\star^{(n)}(\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}}))\Lambda_{E_X}\right] = \sum_{j=1}^{d} p_j\sigma_{W_j^{(1)}}\otimes\tau_{W_j^{(2)}X}$$
$$= \mathrm{Tr}_U(\rho_V). \qquad \square$$

Specializing to the case $n = 1$ enables a stronger result to be derived that does not require independence assumptions.

Theorem 5.5. *Let (G,ρ_V) be a 1-Bifactor Network with graph $G = (V,E)$. Let U,W,X be nonintersecting subsets of V such that $U\cup W\cup X = V$. The following diagram is commutative.*

$$\begin{array}{ccc} \Gamma_U\star(\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}}) & \overset{\mathrm{Tr}_U}{\rightarrow} & \mathrm{Tr}_U(\Gamma_U\star(\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}})) \\ \downarrow \Gamma_{W\cup X}\star(\cdot\Lambda_{E_X}) & & \downarrow \Gamma_{W\cup X}\star(\cdot\Lambda_{E_X}) \\ \rho_V = \Gamma_V\star\Lambda_{E_V} & \overset{\mathrm{Tr}_U}{\rightarrow} & \mathrm{Tr}_U(\rho_V). \end{array} \tag{111}$$

Proof. The theorem follows simply from the cyclic property of the partial trace:

$$\mathrm{Tr}_U(\rho_V) = \mathrm{Tr}_U\left([\Gamma_U^{\frac{1}{2}}\otimes\Gamma_{W\cup X}^{\frac{1}{2}}]\Lambda_E[\Gamma_U^{\frac{1}{2}}\otimes\Gamma_{W\cup X}^{\frac{1}{2}}]\right) \tag{112}$$

$$= \Gamma_{W\cup X}^{\frac{1}{2}}\mathrm{Tr}_U(\Gamma_U\Lambda_E)\otimes\Gamma_{W\cup X}^{\frac{1}{2}} \tag{113}$$

$$= \Gamma_{W\cup X}^{\frac{1}{2}}\mathrm{Tr}_U(\Gamma_U\Lambda_{E_{U\cup W}}\Lambda_{E_{U\cup W:X}})\Lambda_{E_X}\otimes\Gamma_{W\cup X}^{\frac{1}{2}}. \qquad \square \tag{114}$$

We are now positioned to state and prove the main result of this section.

Theorem 5.6. *Let (G,ρ_V) be an n-Bifactor Network with graph $G = (V,E)$, and let $[b_u^{(n)}(t), b_{uv}^{(n)}(t)] = \mathrm{QBP}_t^{(n)}(\mu_u, \nu_{u:v})$. If (G,ρ_V) is a quantum Markov network and G is a tree, then for all $t \geqslant$ diameter(G), $b_u^{(n)}(t) = \rho_u$ and $b_{uv}^{(n)}(t) = \rho_{u\cup v}$.*

Proof. First, observe that by definition $b_u^{(n)}(t) = \mathrm{Tr}_v(b_{uv}^{(n)}(t))$, so it is sufficient to prove that $b_{uv}^{(n)}(t) = \rho_{u\cup v}$. Consider $u\cup v$ to be the root of the tree. We proceed by induction, repeatedly tracing out leaves from the bifactor state except u and v until we are left with only vertices u and v. Set $G(0) = G$ and let $G(t) = (V(t),E(t))$ be the tree left after t such rounds of removing leaves. Denote the leaves of $G(t)$ apart from u and v by $l(t)$, the children of x by $c(x)$, and the unique parent of x by $m(x)$. At $t = 0$, consider tracing out a leaf w of G

$$\mathrm{Tr}_w(\rho_V) = \mathrm{Tr}_u((\mu_w \otimes \Gamma_{V-w})\star^{(n)}(v_{w:m(w)}\Lambda_{E_{V-w}})) \tag{115}$$

$$= \Gamma_{V-w}\star^{(n)}\left[\mathrm{Tr}_w\left(\mu_w\star^{(n)}v_{w:m(w)}\right)\Lambda_{E_{V-w}}\right] \tag{116}$$

$$= \Gamma_{V-w}\star^{(n)}\left[m^{(n)}_{w\to m(w)}(1)\Lambda_{E_{V-w}}\right] \tag{117}$$

where we have used Theorem 5.4 going from the first to the second line. Since this holds for all leaves, we conclude that

$$\mathrm{Tr}_{l(0)}(\rho_V) = \Gamma_{V(1)}\star^{(n)}\left(\prod_{x\in l(0)}\prod_{y\in c(x)} m^{(n)}_{y\to x}(1)\Lambda_{V(1)}\right). \tag{118}$$

We thus make the inductive assumption that

$$\rho_{V(t)} = \Gamma_{V(t)}\star^{(n)}\left(\prod_{x\in l(t)}\prod_{y\in c(x)} m^{(n)}_{y\to x}(t)\Lambda_{V(t)}\right). \tag{119}$$

It follows that

$$\rho_{V(t+1)} = \mathrm{Tr}_{l(t)}(\rho_{V(t)}) \tag{120}$$

$$= \mathrm{Tr}_{l(t)}\left(\Gamma_{V(t)}\star^{(n)}\left[\prod_{x\in l(t)}\prod_{y\in c(x)} m^{(n)}_{y\to x}(t)\Lambda_{V(t)}\right]\right) \tag{121}$$

$$= \mathrm{Tr}_{l(t)}\left(\Gamma_{V(t+1)}\star^{(n)}\left[\prod_{x\in l(t)}\mu_x\star^{(n)}\left(\prod_{y\in c(x)} m^{(n)}_{y\to x}(t)v_{x:m(x)}\Lambda_{V(t+1)}\right)\right]\right) \tag{122}$$

$$= \Gamma_{V(t+1)}\star^{(n)}\left[\prod_{x\in l(t)}\mathrm{Tr}_x\left(\mu_x\star^{(n)}\left(\prod_{y\in c(x)} m^{(n)}_{y\to x}(t)v_{x:m(x)}\right)\right)\Lambda_{V(t+1)}\right] \tag{123}$$

$$= \Gamma_{V(t+1)}\star^{(n)}\left[\prod_{x\in l(t)} m^{(n)}_{x\to m(x)}(t+1)\Lambda_{V(t+1)}\right] \tag{124}$$

$$= \Gamma_{V(t+1)}\star^{(n)}\left[\prod_{x\in l(t+1)}\prod_{y\in c(x)} m^{(n)}_{y\to x}(t+1)\Lambda_{V(t+1)}\right] \tag{125}$$

also assumes the same form, so Eq. (119) follows by induction. We have again used Theorem 5.4 in going from the third to the fourth line. When $V(t)$ contains only u and v then this reduces to $\rho_{u\cup v} = b^{(n)}_{uv}(t)$, which is what we set out to prove. □

Once again, specializing to the case $n = 1$ enables a stronger result to be derived that does not rely on independence assumptions.

Corollary 5.7. *Let (G,ρ_V) be an 1-Bifactor Network with graph $G = (V,E)$, and let $[b_u(t), b_{uv}(t)] = \mathrm{QBP}^{(1)}_t(\mu_u, v_{u:v})$. If G is a tree, then for all $t \geqslant$ diameter(G), $b_u(t) = \rho_u$ and $b_{uv}(t) = \rho_{u\cup v}$.*

Proof. This Corollary is a consequence of Theorem 5.5 and the fact that the proof of Theorem 5.6 only relies on the commutativity of the diagram Eq. (107). □

This last result gives us additional information about the structure of correlations in 1-bifactor states that is captured by the following corollary.

Corollary 5.8. *Let (G, ρ_V) be an 1-Bifactor Network on graph $G = (V, E)$. If G is a tree, then the mutual density operators commute*: $[\rho_{u:v}, \rho_{w:x}] = 0$ *for all* (u, v) *and* $(w, x) \in E$.

Proof. The only nontrivial case is $[\rho_{u:v}, \rho_{v:w}]$ with $u \neq w$. Let $[b_u(t), b_{uv}(t)] = \mathrm{QBP}_t^{(1)}(\mu_u, \nu_{u:v})$ and denote

$$A_{u-v}(t) = \prod_{w \in n(u)-v} m_{w \to u}(t). \tag{126}$$

Observe that $A_{u-v}(t)$ is an operator on $\mathcal{H}_u$, and by Proposition 5.1, $[A_{u-v}(t), \nu_{u:w}] = 0$ for all u, v, and $w \in V$. From Theorem 5.6, we have for $t \geqslant \mathit{diameter}(G)$

$$[\rho_{u:v}, \rho_{v:w}] = [A_{u-v}(t)A_{v-u}(t)\nu_{u:v}, A_{v-w}(t)A_{w-v}(t)\nu_{v:w}] = 0. \qquad \square \tag{127}$$

Corollary 5.7 shows that for 1-bifactor states on trees, $\mathrm{QBP}^{(1)}$ enables an efficient evaluation of the one-vertex and two-vertex reduced density operators ρ_u for all $u \in V$ and $\rho_{u \cup v}$ for all $(u, v) \in E$. Can this result be generalized to arbitrary bifactor states? This question is of interest since, as we will detail in Section 7.2, the Gibbs states used in statistical physics are ∞-bifactor states. However, it is known that approximating the ground state energy of a two-local Hamiltonian on a chain is QMA-complete [5,21].[4] Knowledge of $\rho_{u \cup v}$ leads to an efficient evaluation of the energy. Therefore, without any independence assumptions, it is unlikely that an efficient QBP algorithm for n-Bifactor Networks will converge to the correct marginals for $n > 1$. This contrasts with classical BP that always converges to the exact solution on trees. However, Section 6.3 gives a QBP algorithm that solves the inference problem for any n-bifactor state on a tree in a time that scales exponentially with n.

5.3. Solving inference problems

We close this section with a discussion of how QBP algorithm can solve inference problems when local measurements are executed on a bifactor state. In other words, for an outcome of a local measurement on a subsystem U described by a POVM element $E_U^{(j)} = \otimes_{u \in U} E_u^{(j)}$, we are interested in evaluating the marginal states $\rho_{u|E_U^{(j)}}$ and $\rho_{u \cup v|E_U^{(j)}}$ conditioned on the outcome, where

$$\rho_{u|E_U^{(j)}} = \frac{1}{Y} \mathrm{Tr}_{V-u}\left((E_U^{(j)})^{\frac{1}{2}} \rho_V (E_U^{(j)})^{\frac{1}{2}} \right) \tag{128}$$

$$\rho_{u \cup v|E_U^{(j)}} = \frac{1}{Y} \mathrm{Tr}_{V-\{u,v\}}\left((E_U^{(j)})^{\frac{1}{2}} \rho_V (E_U^{(j)})^{\frac{1}{2}} \right), \tag{129}$$

and Y is a normalization factor. For $u, v \notin U$, this amounts to a local modification of the bifactor state that accounts for the action of the measurement, the QBP algorithm being otherwise unaltered. We focus on 1-Bifactor Networks and return to the general case at the end of this section.

Theorem 5.9. *Let (G, ρ_V) be a 1-Bifactor Network with $G = (V, E)$ a tree. For $U \subset V$, let $\{E_U^{(j)}\} = \{\otimes_{u \in U} E_u^{(j)}\}$ be a POVM on the subsystem U and let $W = V - U$. Define*

[4] QMA stands for Quantum Merlin and Arthur and it is the natural quantum generalization of the classical complexity class NP. So to the best of our knowledge, solving a QMA-complete problem would require an exponential amount of time even on a quantum computer.

$\mu_u^{(j)} = \mu_u \star E_u^{(j)}$ *for* $u \in U$ *and* $\mu_u^{(j)} = \mu_u$ *for* $u \in W$. *Let* $[b_{uv}(t), b_{uv}(t)] = \mathrm{QBP}^{(1)}(\mu_u^{(j)}, v_{u:v})$. *Then for all* $t \geqslant \mathrm{diameter}(G)$, $b_u(t) = \rho_{u|E_U^{(j)}}$ *for all* $u \in W$ *and* $b_{u\cup v}(t) = \rho_{uv|E_U^{(j)}}$ *for all* $(u,v) \in E_W$.

Proof. The reduced state on W conditioned on the measurement outcome $E_U^{(j)}$ is given by

$$\rho_{W|E_U^{(j)}} = \frac{1}{Y}\mathrm{Tr}_U\left(\left(E_U^{(j)}\right)^{\frac{1}{2}}\rho_V\left(E_U^{(j)}\right)^{\frac{1}{2}}\right) \tag{130}$$

$$= \frac{1}{Y}\prod_{\substack{v\in W\\ u\in U}}\prod_{(w,x)\in E}\mu_v^{\frac{1}{2}}\mathrm{Tr}_U\left(\left(E_u^{(j)}\right)^{\frac{1}{2}}\mu_u^{\frac{1}{2}}v_{w:x}\mu_u^{\frac{1}{2}}\left(E_u^{(j)}\right)^{\frac{1}{2}}\right)\mu_v^{\frac{1}{2}} \tag{131}$$

$$= \frac{1}{Y}\prod_{\substack{v\in W\\ u\in U}}\prod_{(w,x)\in E}\mu_v^{\frac{1}{2}}\mathrm{Tr}_U\left(v_{w:x}\mu_u^{\frac{1}{2}}E_u^{(j)}\mu_u^{\frac{1}{2}}\right)\mu_v^{\frac{1}{2}} \tag{132}$$

$$= \frac{1}{Y}\prod_{\substack{v\in W\\ u\in U}}\prod_{(w,x)\in E}\left(\mu_v^{(j)}\right)^{\frac{1}{2}}\mathrm{Tr}_U\left(\left(\mu_u^{(j)}\right)^{\frac{1}{2}}v_{w:x}\left(\mu_u^{(j)}\right)^{\frac{1}{2}}\right)\left(\mu_v^{(j)}\right)^{\frac{1}{2}}. \tag{133}$$

The result thus follows from Corollary 5.7. □

The result of Theorem 5.9 can easily be extended to compute the conditional marginal state $\rho_{u|E_U^{(j)}}$ and $\rho_{u\cup v|E_U^{(j)}}$ for any u and v, not just those in $W = V - U$. This is achieved by altering the beliefs as follows

$$b_u(t) = \frac{1}{Z}E_u^{(j)}\star\mu_u\star\prod_{v'\in n(u)} m_{v'\to u}(t) \tag{134}$$

for $u \in U$,

$$b_{uv}(t) = \frac{1}{Z}E_{uv}^{(j)}\star(\mu_u\mu_v)\star\left[\prod_{w\in n(u)-v} m_{w\to u}(t)\prod_{w'\in n(v)-u} m_{w'\to v}(t)\star v_{u:v}\right] \tag{135}$$

with $E_{uv}^{(j)} = E_u^{(j)}\otimes I_v$ when $u \in U$ and $v \in W$ and $E_{uv}^{(j)} = E_u^{(j)}\otimes E_u^{(j)}$ when $u, v \in U$. The proof is straightforward and we omit it.

Theorem 5.9 shows how QBP leads to an efficient algorithm for solving inference problems on 1-bifactor states on trees with local measurements. This immediately implies an efficient algorithm for general n-bifactor states when (G, ρ_V) is a quantum Markov network. Indeed, Theorem 5.6 demonstrates that in that case the $\mathrm{QBP}^{(n)}$ algorithm can be used to efficiently compute the marginal density operators $\rho_{u\cup v}$ for all $(u,v) \in E$. From these, one can straightforwardly obtain the marginal operators ρ_u for all $u \in V$ and mutual operators $\rho_{u:v}$ for all $(u,v) \in E$. Theorem 4.10 states that ρ_V can be represented as a 1-bifactor state in terms of its marginal and mutual operators. The inference problem can then be solved using the $\mathrm{QBP}^{(1)}$ algorithm as explained above.

6. Heuristic methods

The previous section provided conditions under which QBP algorithms give exact solutions to inference problems on n-Bifactor Networks. Namely, the underlying graph must be a tree, and the state must be either a quantum Markov network or a 1-bifactor state.

When these conditions are not met, QBP algorithms may still be used as heuristic methods to obtain approximate solutions to the inference problem, although in general these approximations will be uncontrolled.

To draw a parallel, classical Belief Propagation algorithms have found applications in numerous distinct scientific fields where they are sometimes known under different name: Gallager decoding, Viterbi's algorithm, sum-product, and iterative turbo decoding in information theory; cavity method and the Bethe-Peierls approximation in statistical physics; junction-tree and Shafer-Shenoy algorithm in machine learning to name a few. In many of these examples, BP algorithms exhibit good performance on graphs with loops, even though the algorithm does not converge to the exact solution on such graphs. In fact, "Loopy Belief Propagation" is often the best known heuristic method to find approximate solutions to hard problems. Important examples include the near-Shannon capacity achieving turbo-codes and low density parity check codes. On the other hand, there are known examples for which loopy BP fail to converge and their general realm of applicability is not yet fully understood.

As in the classical case, one can expect loopy QBP to give reasonable approximations in some circumstances, for instance when the size of typical loops is very large. Intuitively, one expects a local algorithm to be relatively insensitive to the large scale structure of the underlying graph. However, quantum inference problems also pose a new challenge. Quite apart from issues regarding the graph's topology, an n-bifactor state with $n > 1$ may not obey the independence conditions required to ensure the convergence of QBP. The goal of this section is to suggest three techniques that are expected to improve the performance of QBP in such circumstances.

6.1. Coarse graining

By definition, a quantum Markov network has the property that the correlations from one-vertex to the rest of the graph are screened off by its neighbors. When this property fails, QBP will not in general produce the correct solution to an inference problem. Coarse graining is a simple way of modifying a graph in such a way that the state may be a closer to forming a quantum Markov network with respect to the new graph than it was with respect to the original graph.

A coarse graining of a graph $G = (V, E)$ is a graph $\widetilde{G} = (\widetilde{V}, \widetilde{E})$, where $\widetilde{V}$ is a partition of V into disjoint subsets of and $(U, W) \in \widetilde{E}$ if there is an edge connecting a vertex in U to a vertex in W in G. The coarse grainings that are of most interest are those that partition V into connected sets of vertices (see Fig. 7 for example). It is an elementary exercise to show that if (G, ρ_V) is an n-Bifactor Network, then $(\widetilde{G}, \rho_{\widetilde{V}})$ is an n-Bifactor Network for any coarse graining $\widetilde{G}$. The intuition for why coarse graining might get us closer to a Markov network is that it effectively "thickens" the neighborhood of each vertex, which may then be more efficient at screening off correlations. This intuition is illustrated in Fig. 7 and is supported by the fact that Markov networks are fixed points of the coarse graining procedure, i.e. if $\widetilde{G}$ is a coarse graining of G, then $(\widetilde{G}, \rho_{\widetilde{V}})$ is a quantum Markov network whenever (G, ρ_V) is a Markov network.

Also note that every graph G can be turned into a tree by a suitable coarse graining. When the obtained Bifactor Network is a Markov Network or when $n = 1$, QBP is then guaranteed to converge to the exact solution. The Hilbert space dimension at the vertices

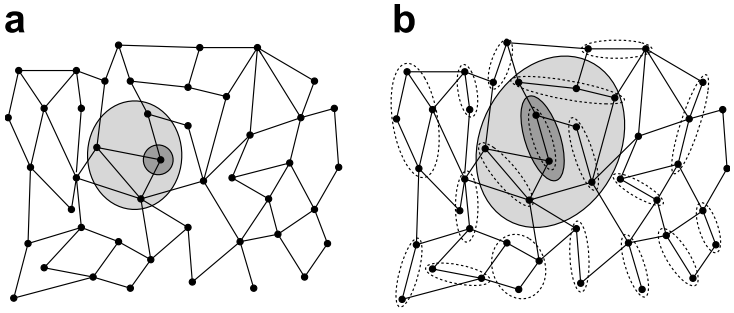

Fig. 7. Example of a coarse-grained graph. (a) Shows in light gray the neighborhood of the darkened vertex in the original graph. In (b) the dashed ellipses represent coarse-grained vertices. The neighborhood of the darkened coarse-grained vertex is represented by the light gray set.

of the coarse-grained graph is bounded by an exponential in the tree-width of G, so this technique is efficient only for graph of $O(\log(N))$ tree-width.

6.2. Sliding window QBP

Sliding window QBP is similar in spirit to coarse graining but is mainly suitable for chains (although the idea is easily generalized to arbitrary trees of low degree). Consider an n-bifactor state ρ_V on a one-dimensional lattice $G = (V, E)$ with $V = \{v_1, v_2, \ldots, v_N\}$ and $E = \{(v_j, v_{j+1})\}_{j=1,\ldots,N-1}$. When (G, ρ_V) is not a quantum Markov Network, the diagram of Eq. (107) will generally fail to be commutative. The commutativity of this diagram is essential for the success of QBP, as for instance it implies

$$\mathrm{Tr}_{v_1}\left((\mu_{v_1} \otimes \mu_{v_2})\star^{(n)}(v_{v_1:v_2} v_{v_2:v_3})\right) = \mu_{v_2}\star^{(n)}\left[\mathrm{Tr}_{v_1}(\mu_{v_1}\star^{(n)} v_{v_1:v_2})\star^{(n)} v_{v_2:v_3}\right] \tag{136}$$

$$= \mu_{v_2}\star^{(n)}[m_{v_1\to v_2}\star^{(n)} v_{v_2:v_3}]. \tag{137}$$

Thus, the Hilbert space of vertex v_1 is traced out before operators on vertex v_3 are brought into the picture. This enables the algorithm to progress along the lattice by evaluating a cumulative operator of constant dimension (i.e. the messages), much in the spirit of the transfer matrix of statistical physics. Without the Markov property, this is generally not possible.

However, when vertices separated by a distance ℓ are conditionally independent given the vertices between them, sliding window QBP can be operated efficiently to produce the exact solution of the inference problem. This works by defining new message operators

$$\tilde{m}_{v_{j+\ell-1}\to v_{j+\ell}} = \mathrm{Tr}_{\{v_1,v_2,\ldots,v_j\}}\left(\left[\bigotimes_{k=1}^{\ell+j-1}\mu_{v_k}\right]\star^{(n)}\left[\prod_{k=1}^{\ell+j-1} v_{v_k:v_{k+1}}\right]\right) \tag{138}$$

which act on $\mathcal{H}_{v_{j+1}} \otimes \mathcal{H}_{v_{j+2}} \otimes \cdots \mathcal{H}_{v_{j+\ell}}$. When

$$S(v_j : v_{j+\ell}|\{v_{j+1}, v_{j+2}, \ldots, v_{j+\ell-1}\}) = 0 \tag{139}$$

for all $v_j \in V$, we have the equality

$$\tilde{m}_{v_{j+\ell}\to v_{j+\ell+1}} = \mathrm{Tr}_{v_{j+1}}\left(\mu_{v_{j+\ell}}\star^{(n)}\left[\tilde{m}_{v_{j+\ell-1}\to v_{j+\ell}}\star^{(n)} v_{v_{j+\ell}:v_{j+\ell+1}}\right]\right), \tag{140}$$

so inference problems can be solved exactly with operators whose dimension grow exponentially with the ℓ rather than the lattice size N. In particular, this method can be applied

to spin systems that have a finite correlation length because then Eq. (139) can be expected to hold approximately for some finite ℓ.

6.3. Replica

The replica trick maps n-bifactor states to 1-bifactor states on which QBP$^{(1)}$ can be implemented without concerns for independence. This is achieved by replacing the systems v on each vertex of the graph G by n replicas, so that the Hilbert space associated to vertex v becomes $\mathcal{H}_v^{\otimes n}$. Thus, instead of connecting two quantum systems, edges of the graph now connect two sets of n replicas of the original systems: edges acquire a new dimension and become ribbons, see Fig. 8. As a consequence, the algorithm suffers an overhead exponential in n. The name "replica" is borrowed from the analogous technique used in the study of classical quenched disordered systems. The validity of this technique is based on the following observation.

Proposition 6.1. *Let $\{\mathcal{H}_j\}_{j=1,\ldots,n}$ be isomorphic Hilbert spaces. Let $T^{(n)}$ be the operator that cyclicly permutes these n systems. Let A_1 be an arbitrary operator on $\mathcal{H}_1$, and define $A_j = (T^{(n)})^{j-1} A_1 (T^{(n)\dagger})^{j-1}$ to be the corresponding operators on $\mathcal{H}_j$. Then for any set of operators $\{A_1^{(k)}\}$ on $\mathcal{H}_1$, the following equality holds*

$$A_1^{(1)} A_1^{(2)} \cdots A_1^{(n)} = \mathrm{Tr}_{2,3,\ldots,n}\left(\left[A_1^{(1)} \otimes A_2^{(2)} \otimes \cdots \otimes A_n^{(n)}\right] T^{(n)}\right). \tag{141}$$

We are now in a position to formalize the method of replicas.

Theorem 6.2. *Let (G, ρ_V) be an n-Bifactor Network, with operators μ_u and $v_{u:v}$. Then, ρ_V is locally isomorphic to a 1-bifactor state with Hilbert spaces comprising n replicas of the original system $\mathcal{H}'_u = \mathcal{H}_{u_1} \otimes \mathcal{H}_{u_2} \otimes \cdots \otimes \mathcal{H}_{u_n}$ for all $u \in V$. The partial isomorphism at vertex u is given by $\mathrm{Tr}_{u_2,u_3,\ldots,u_n}\left((T_u^{(n)\dagger})^{\frac{1}{2}} \cdot (T_u^{(n)})^{\frac{1}{2}}\right)$. More precisely, we claim that*

$$\rho_V = \mathrm{Tr}_{\{u_2,u_3,\ldots,u_n\}_{u\in V}}\left(U^\dagger\left(\bigotimes_{u\in V} \tilde{\mu}_u\right) \star \left(\prod_{(v,w)\in E} \tilde{v}_{v:w}\right) U\right) \tag{142}$$

where

$$\tilde{\mu}_u = \left(\mu_u^{\frac{1}{n}}\right)^{\otimes n} (T_u^{(n)}) \tag{143}$$

$$\tilde{v}_{u:v} = \left(v_{u:v}^{\frac{1}{n}}\right)^{\otimes n} \tag{144}$$

$$U = \bigotimes_{u\in V} \left(T_u^{(n)}\right)^{\frac{1}{2}} \tag{145}$$

are operators on $\mathcal{H}'_u$.

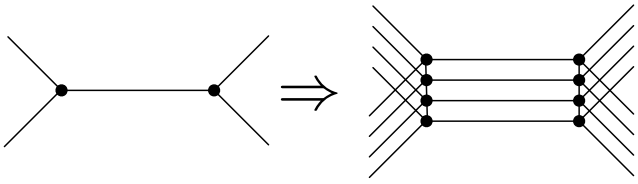

Fig. 8. Converting a n-bifactor state to a 1-bifactor state requires the introduction of n replicas of the original system.

Proof. First, note that $T_u^{(n)}$ commutes with $(\mu_u^{\frac{1}{n}})^{\otimes n}$, so $\tilde{\mu}_u^{\frac{1}{2}} = (\mu_u^{\frac{1}{2n}})^{\otimes n}(T_u^{(n)})^{\frac{1}{2}} = (T_u^{(n)})^{\frac{1}{2}}(\mu_u^{\frac{1}{2n}})^{\otimes n}$. Thus

$$\mathrm{Tr}_{\{u_2,u_3,\ldots,u_n\}_{u\in V}}\left(U^\dagger\left(\bigotimes_{u\in V}\tilde{\mu}_u\right)\star\left(\prod_{(v,w)\in E}\tilde{v}_{v:w}\right)U\right) \tag{146}$$

$$= \mathrm{Tr}_{\{u_2,u_3,\ldots,u_n\}_{u\in V}}\left(U^\dagger\left(\bigotimes_{u\in V}T_u^{(n)}\left(\mu_u^{\frac{1}{n}}\right)^{\otimes n}\right)\star\left(\prod_{(v,w)\in E}\left(v_{u:v}^{\frac{1}{n}}\right)^{\otimes n}\right)U\right) \tag{147}$$

$$= \mathrm{Tr}_{\{u_2,u_3,\ldots,u_n\}_{u\in V}}\left(\left(\bigotimes_{u\in V}\left(\mu_u^{\frac{1}{n}}\right)^{\otimes n}\right)\star\left(\prod_{(v,w)\in E}\left(v_{u:v}^{\frac{1}{n}}\right)^{\otimes n}\right)\bigotimes_{u\in V}T_u^{(n)}\right) \tag{148}$$

$$= \mathrm{Tr}_{\{u_2,u_3,\ldots,u_n\}_{u\in V}}\left(\left[\left(\bigotimes_{u\in V}\mu_u^{\frac{1}{n}}\right)\star\left(\prod_{(v,w)\in E}v_{u:v}^{\frac{1}{n}}\right)\right]^{\otimes n}\bigotimes_{u\in V}T_u^{(n)}\right) \tag{149}$$

$$= \left[\left(\bigotimes_{u\in V}\mu_u^{\frac{1}{n}}\right)\star\left(\prod_{(v,w)\in E}v_{u:v}^{\frac{1}{n}}\right)\right]^n = \rho_V \tag{150}$$

where we used Proposition 6.1 to obtain the last line. □

Since the dimension of the Hilbert at each vertex grows exponentially with n, the QBP$^{(1)}$ algorithm used to solve the corresponding inference problem suffers an exponential overhead. One can make a "replica symmetry ansatz", assuming that the state is symmetric under exchange of replica systems at any given vertex. Since the symmetric subspace of $\mathcal{H}_v^{\otimes n}$ grows polynomially[5] with n, QBP algorithm can be executed efficiently. The validity of this ansatz cannot be verified in general, but it may serve as a good heuristic method.

7. Applications

This section explains in some detail how QBP can be used as a heuristic algorithm to find approximate solutions to important problems in quantum error correction and the simulation of many-body quantum systems. The focus will be on the reduction of well-established problems to inference problems on n-Bifactor Networks. One can make use of the techniques discussed in the previous section whenever the resulting Graphical Model does not meet the requirements to ensure convergence of QBP, or when these conditions cannot be verified efficiently.

7.1. Quantum error correction

Maximum-likelihood decoding is an important task in quantum error correction (QEC). As in classical error correction, this problem reduces to the evaluation of marginals on a factor graph, also called Tanner graph in this context. More precisely, for independent error models, the quantum channel conditioned on error syndrome is a 1-bifactor state. As a consequence, qubit-wise maximum-likelihood decoding of a QEC stabilizer code reduces to an inference problem on a 1-Bifactor Network. Thus, there is

[5] More precisely, it grows as $\binom{n+d-1}{n} \approx n^{d-1}$.

no independence condition that needs to be verified, although the graph will generally contain loops. Before demonstrating this reduction, a brief summary of stabilizer QEC is in order, see [16] for more details. For details on the use of Belief Propagation for the decoding of classical error correction codes, the reader is referred to the text of MacKay [31] and forthcoming book of Richardson and Urbanke [44].

Consider a collection of N two-dimensional quantum systems (qubits) $V = \{u\}_{u=1,\ldots,N}$ with $\mathcal{H}_u = \mathbb{C}^2$. A QEC code is a subspace $\mathcal{C} \in \mathcal{H}_V$ that is the $+1$ eigensubspace of a collection of commuting operators S_j, $j = 1, \ldots, N-K$, called stabilizer generators. Each stabilizer generator is a tensor product of Pauli operators on a subset U_j of V:

$$S_j = \bigotimes_{u \in U_j} \sigma_u^{\alpha_j^u} \tag{151}$$

where $\alpha_j^u \in \{x, y, z\}$. When the stabilizer generators are multiplicatively independent, the code encodes K qubits, i.e. $\mathcal{C}$ has dimension 2^K. For each $j = 1, \ldots, N-K$, define the two projectors $P_j^{\pm} = (I \pm S_j)/2$. The code space is therefore defined as $\mathcal{C} = (\prod_j P_j^+)\mathcal{H}_V$.

Error correction consists of three steps. First, the system V is prepared in a code state ρ_V supported on $\mathcal{C}$, in such a way that $P_j^+ \rho_V P_j^+ = \rho_V$ for all j. The state is then subjected to the channel $\rho_V \to \mathcal{E}_{V|V}(\rho_V)$. Second, each stabilizer generator S_j is measured, yielding an outcome $s_j = \pm$ with probability $\mathrm{Tr}(P_j^{\pm} \mathcal{E}_{V|V}(\rho_V))$. The collection of all $N-K$ measurement outcomes s_j, called the error syndrome, is denoted $\mathbf{s} = (s_1, s_2, \ldots s_{N-K}) \in \{-,+\}^{N-K}$. Third, the channel $\mathcal{E}_{V|V}$ is updated conditioned the error syndrome $\mathbf{s}$. Based on this updated channel, the optimal recovery is computed and implemented.

The computationally difficult step in the above protocol consists in conditioning the channel on the error syndrome. To understand this problem, it is useful to express the channel in a Kraus form $\mathcal{E}_{V|V}(\rho_V) = \sum_k M_{V|V}^{(k)} \rho_V M_{V|V}^{(k)\dagger}$ where $\{M^{(k)}\}$ are operators on $\mathcal{H}_V$. When $s_j = +$, we learn that the error that has affected the state commutes with S_j, while $s_j = -$ indicates that the error anti-commutes with S_j. To update the channel conditioned on the error syndrome $s_j = +$ say, we first decompose each Kraus operator $M_{V|V}^{(k)}$ as the sum of an operator that commutes with S_j and an operator that does not commute with S_j: $M_{V|V}^{(k)} = M_{V|V}^{(k)+} + M_{V|V}^{(k)'}$ where $M_{V|V}^{(k)+} = P_j^+ M_{V|V}^{(k)} P_j^+$ and $M_{V|V}^{(k)'} = M_{V|V}^{(k)} - M_{V|V}^{(k)+}$. The updated channel is obtained by throwing away the primed component $M_{V|V}^{(k)'}$ of each Kraus operator, and renormalizing.

In what follows, we demonstrate how the conditional channel can be expressed as a factor graph. This is most easily done using the Jamiołkowski representation of quantum channels. For each quantum system v, let R_v denote a reference for v, with Hilbert space $\mathcal{H}_{R_v} \simeq \mathcal{H}_v$. Define the maximally entangled state between system v and its reference by $|\Phi\rangle_{vR_v} = \frac{1}{\sqrt{d}} \sum_j |j\rangle_v |j\rangle_{R_v}$. Then, the Jamiołkowski representation of a channel $\mathcal{E}_{V|V}$ is a density operator $\rho_{\overline{V}}$ on $\mathcal{H}_{\overline{V}} = \mathcal{H}_V \otimes \mathcal{H}_{R_V}$ given by $\rho_{\overline{V}} = (\mathcal{E}_{V|V} \otimes \mathcal{I}_{R_V|R_V})(|\Phi\rangle\langle\Phi|_{VR_V})$, where $\mathcal{I}$ denotes the identity channel. For independent error models considered here, $\rho_{\overline{V}} = \otimes_{u \in V} \rho_{\overline{u}}$.

For each stabilizer generator S_j, denote $\overline{S}_j = \otimes_{u \in U_j} \sigma_u^{\alpha_j^u} \otimes \sigma_{R_u}^{\alpha_j^u}$, and construct the associated projectors $\overline{P}_j^{\pm} = (I \pm \overline{S}_j)/2$. An important property of these operator is that they fix the maximally entangled state $\overline{S}_j |\Phi\rangle_{VR_V} = \overline{P}_j^+ |\Phi\rangle_{VR_V} = |\Phi\rangle_{VR_V}$. Let E be an operator on V. If E commutes with S_j, we have $\overline{P}_j^+ (E \otimes I_{R_V}) |\Phi\rangle_{VR_V} = (E \otimes I_{R_V}) |\Phi\rangle_{VR_V}$ and $\overline{P}_j^- (E \otimes I_{R_V}) |\Phi\rangle_{VR_V} = 0$, while if E anti-commutes with S_j, the same identities hold with $\overline{P}_j^+$ and $\overline{P}_j^-$ exchanged. It follows from this observation that conditioned on the error syndrome $\mathbf{s}$, the channel is described by the Jamiołkowski matrix

$$\rho_{\overline{V}|\mathbf{s}} = \frac{1}{Z} \prod_j \overline{P}_j^{s_j} \star \bigotimes_{v \in V} \rho_{\overline{v}}, \tag{152}$$

that is a quantum factor graph.

There are a number of relevant quantities that can be evaluated from this factor graph. For instance, one can efficiently evaluate the conditional channel on any constant size set of qubits $W \subset V$ via partial trace. This is useful in iterative decoding schemes such as those used for quantum turbo-codes [39] and low density parity check codes [10]. In those cases, the conditional channel on W can only be evaluated approximately since it requires loopy QBP. The factor graph also enables exact evaluation of the logical error in a concatenated block coding scheme [42] such as used in fault-tolerant protocols.

7.2. Simulation of many-body quantum systems

In statistical physics, the state of a many-body quantum system V is a Gibbs state $\rho_V = \frac{1}{Z}\exp(-\beta H)$ for some Hamiltonian H, where $\beta = 1/T$ is the inverse temperature. Typically, H is the sum of single and two-body interactions $H = \sum_{u \in V} H_u + \sum_{(u,w) \in E} H_{uw}$ on some graph $G = (V, E)$. Understanding the correlations present in these states is a great challenge in theoretical physics. In this section, we describe how QBP can serve as an heuristic method to accomplish this task approximately. For an account of the use of Belief Propagation in classical statistical mechanical systems, we refer the reader to the text of Mézard and Montanari [32].

Defining $\mu_u = \exp(-\beta H_u)$ and $v_{v:w} = \exp(-\beta H_{vw})$ gives an expression for ρ_V of the form of Eq. (84):

$$\rho_V = \left(\bigotimes_{v \in V} \mu_v \right) \odot \left(\bigodot_{(v,w) \in E} v_{v:w} \right) \tag{153}$$

Thus, ρ_V is an ∞-bifactor state. As mentioned in Section 5, a QBP$^{(\infty)}$ algorithm can easily be formulated for this type of bifactor state, and still converge to the exact solutions of the corresponding inference problem when ρ_V is a quantum Markov network and G is a tree. This requires replacing all matrix products $\prod$ by the commutative product $\odot$ in the defining equations of QBP$^{(\infty)}$ Eqs. (103)–(105). The proof of convergence Theorem 5.6 under these more general conditions follows essentially the same reasoning.

Unfortunately, the convergence of the QBP algorithm in this case requires the state to be a quantum Markov network, which cannot be tested directly in general. As we will now explain, it is often possible to reasonably approximate a Gibbs state by an n-bifactor with finite n, and sometimes even $n = 1$.

First, note that it usually possible to coarse grain G such that $[v_{u:v}, v_{w:x}] = 0$ on the new graph. Consider for instance a one-dimensional chain $G = (V, E)$ with $V = \{u\}_{u=1,\ldots,N}$ and $E = \{(u, u+1)\}_{u=1,\ldots,N-1}$. We can construct a coarse-grained graph $\widetilde{G}$ by identifying all vertices $2u-1$ and $2u$ for $u = 1, \ldots, \lfloor \frac{N}{2} \rfloor$. The state ρ_V is then an ∞-bifactor state on $\widetilde{G}$, with operators

$$\tilde{\mu}_u = \mu_{2u-1} \odot \mu_{2u} \odot v_{2u-1:2u} \tag{154}$$

$$\tilde{v}_{u:u+1} = v_{2u:2u+1}, \tag{155}$$

satisfying $[\tilde{v}_{u:u+1}, \tilde{v}_{v:v+1}] = 0$.

A simple way to approximate a Gibbs states with n-bifactor state is to substitute $\odot$ by $\star^{(n)}$ for some large value of n. In the context of many-body physics, this is called a Trotter–Suzuki decomposition of the Gibbs state, and becomes more accurate as the ratio β/n decreases. The QBP$^{(n)}$ algorithm can then be operated on this n-bifactor state, but its convergence again requires some independence condition that cannot be verified systematically. Alternatively, one can use the method of replicas described in Section 6.3 and solve the inference problem exactly with QBP$^{(1)}$, but with an increase in complexity exponential in n. The method of replicas is then reminiscent of the well-known correspondence between quantum statistical mechanics in d dimensions and classical statistical mechanics in $d+1$ dimensions, where the extra dimension represents inverse temperature.

The 1-bifactor states also capture the correlations of some nontrivial quantum many-body systems. *Valence bond solid* (VBS) states were introduced in Refs. [4,3] as exact ground states (i.e. $T = 0$ Gibbs states) of spin systems with interesting properties. Recent work has generalized these constructions to *matrix product states* (MPS) in one-dimension [13,54,55], and *projected entangled-pair states* (PEPS) for higher dimensions [52,46]. These form an important class of states for the description of quantum many-body systems. For instance, *density matrix renormalization group* (DMRG) [57]—one of the most successful method for the numerical study of spin chains—is now understood as a variational method over MPS [40,12,53]. All these states are instances of 1-bifactor states.

For sake of simplicity, we will demonstrate this claim for one-dimensional MPS, but the same argument holds for higher dimensions. The MPS $|\Psi\rangle$ is a pure state of a collection of N d-dimensional quantum systems displayed on a one-dimensional lattice. Each vertex u is assigned two "virtual particles" L_u and R_u, where L and R stand for left and right (see Fig. 9 for a illustration of this construction in two-dimensions). Each of these particles are associated a Hilbert space $\mathcal{H}_{L_u} = \mathcal{H}_{R_u} = \mathbb{C}^D$. Initially, the right particle of vertex u is in a maximally entangled state with the left particle of vertex $u+1$; $|\Phi\rangle_{R_u \cup L_{u+1}} = \frac{1}{\sqrt{D}} \sum_{\alpha=1}^{D} |\alpha\rangle_{R_u} |\alpha\rangle_{L_{u+1}}$ where $|\alpha\rangle$ are orthogonal basis vectors for $\mathbb{C}^D$. (The lattice can be closed to form a circle, in which case we identify $N+1 = 1$.) The initial state is therefore $|\Phi_0\rangle = \otimes_u |\Phi\rangle_{R_u \cup L_{u+1}}$.

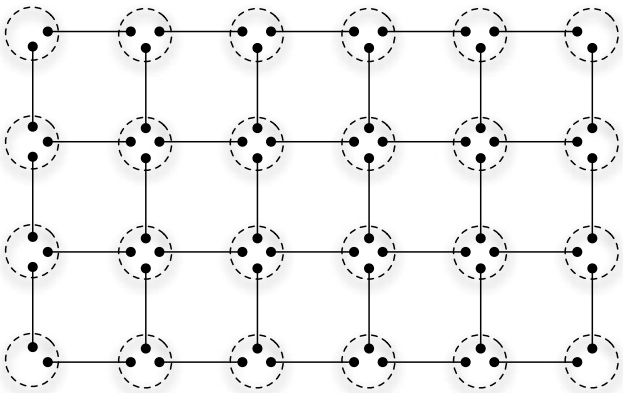

Fig. 9. Projected entangled-pair state on a two-dimensional square lattice. The vertices are associated to dashed circles. Each •—• represents a maximally entangled state of D-dimension shared between neighboring vertices. A partial isometry $A_u : (\mathbb{C}^D)^{c_u} \to \mathbb{C}^d$ is applied at each vertex, where c_u is the degree of vertex u.

To obtain the MPS, apply an operator $A_u : \mathcal{H}_{L_u} \otimes \mathcal{H}_{R_u} \to \mathbb{C}^d$

$$A_u = \sum_{j=1}^{d} \sum_{\alpha,\beta=1}^{D} A_u^{j,\alpha,\beta} |j\rangle\langle\alpha,\beta| \tag{156}$$

to each vertex of the lattice. The vectors $|j\rangle$ form an orthogonal basis for $\mathbb{C}^d$. The resulting state is

$$|\Psi\rangle = \bigotimes_{u=1}^{N} A_u |\Phi_0\rangle \propto \sum_{j_1,j_2,\ldots,j_N=1}^{d} \mathrm{Tr}(B_1^{j_1} B_2^{j_2} \ldots B_N^{j_N}) |j_1, j_2, \ldots, j_N\rangle \tag{157}$$

where the matrices B_u^j are the submatrices of A_u with matrix elements $(B_u^j)_{(\alpha,\beta)} = A_u^{j,\alpha,\beta}$.

For the corresponding 1-bifactor state, the underlying graph $G = (V, E)$ is also a one-dimensional lattice $V = \{1, 2, \ldots, N\}$ and $E = \{(1,2), (2,3), \ldots, (N-1, N)\}$. The Hilbert space associated to vertex u is $\mathcal{H}_u = \mathbb{C}^D \otimes \mathbb{C}^D$. As above, it is convenient to imagine that each vertex u is composed of two D-dimensional subsystems L_u and R_u. Then, up to a local isometry, the MPS of Eq. (157) can be expressed as a 1-bifactor state Eq. (83) with

$$\mu_u = A_u^\dagger A_u \text{ and } v_{u:v} = |\Phi\rangle\langle\Phi|_{R_u \cup L_v}. \tag{158}$$

Moreover, the operators $v_{u:v}$ mutually commute. To see the relation with Eq. (157), note that the operators A^u can be polar decomposed $A_u = U_u \sqrt{A_u^\dagger A_u} = U_u \mu_u^{\frac{1}{2}}$.[6] The matrix U_u is a partial isometry $\mathcal{H}_u \to \mathbb{C}^d$ and

$$|\Psi\rangle\langle\Psi| = \frac{1}{Z} \left(\prod_{u\in V} A_u \right) |\Phi_0\rangle\langle\Phi_0| \left(\prod_{u\in V} A_u^\dagger \right) \tag{159}$$

$$= \frac{1}{Z} \left(\prod_{u\in V} U_u \mu_u^{\frac{1}{2}} \right) \left(\prod_{(v,w)\in E} v_{u:v} \right) \left(\prod_{u\in V} \mu_u^{\frac{1}{2}} U_u^\dagger \right) \tag{160}$$

$$= \frac{1}{Z} \left(\prod_{u\in V} U_u \right) \left(\bigotimes_{u\in V} \mu_u \right) \star \left(\prod_{(v,w)\in E} v_{u:v} \right) \left(\prod_{u\in V} U_u^\dagger \right) \tag{161}$$

as claimed.

Bifactor states are thus relevant to the description of quantum many-body systems. QBP can sometimes be used to efficiently compute correlation functions, but in general for spatial dimension larger than one, its convergence is not guaranteed. This is mainly due to the presence of small loops in the underlying graph. Partial solutions have been proposed to overcome this difficulty [52], and it is conceivable that techniques from loopy Belief Propagation and its generalizations [59] will improve these algorithms. As in the classical case however, QBP may be more appropriate for the study of quantum systems on irregular sparse graphs, such as those encountered in classical spin glasses.

Finally, it should be noted that the Markov conditions required to certify the convergence of QBP—or the associated coarse-grained Markov conditions as explained in the previous section—are weaker than those typically studied in statistical physics, namely the vanishing of connected correlation functions beyond some length scale. For pure quantum states, the two notions coincide and are equivalent to the absence of long-range

[6] Note that μ_u has rank $\leqslant d$. This can be seen straightforwardly by writing $\mu_u = \sum_{j=1}^{d} A_u^{*j,\alpha,\beta} A_u^{j,\Gamma,\delta} = \sum_{j=1}^{d} |h_u^j\rangle\langle h_u^j|$ where $|h_u^j\rangle = \sum_{\alpha,\beta} A_u^{j,\alpha,\beta} |\alpha,\beta\rangle \in \mathcal{H}_u$.

entanglement. At finite temperature however, the state is mixed and the vanishing of mutual information between vertices u and $u+\ell$ conditioned on vertices $u+1,\ldots,u+\ell-1$ Eq. (139) does not imply the absence of connected correlations $\langle A_u A_{u+\ell}\rangle = \mathrm{Tr}(\rho_V A_u A_{u+\ell}) - \mathrm{Tr}(\rho_V A_u)\mathrm{Tr}(\rho_V A_{u+\ell})$.

8. Related work

In this section, our approach to quantum Graphical Models and Belief Propagation is compared to other proposals that have appeared in the literature. Firstly, Tucci has developed an approach to quantum Bayesian Networks [47], Markov Networks [49], and Belief Propagation [48] based on a different analogy between quantum theory and classical probability, namely the idea that probabilities should be replaced by complex valued amplitudes. Tucci's models require that these amplitudes should factorize according to conditions similar to those used in classical Graphical Models. One disadvantage of this is that the definition requires a fixed basis to be chosen for the system at each vertex of the graph, and the factorization condition for Bayesian Networks is not preserved under changes of this basis. In contrast, our definition of quantum conditional independence is based on an explicitly basis independent quantity, so it does not have this problem. Another difficulty with using amplitudes is that they are only well-defined for pure states, so that mixed states have to be represented as purifications on larger networks. In our approach, density operators are taken as primary, so mixed states can be represented without purification. On the other hand, the Tucci's definitions can easily accommodate unitary time evolution, whereas we do not have a general treatment of dynamics in our approach at the present time. A related definition of quantum Markov Networks, also based on amplitudes but without a development of the corresponding Belief Propagation algorithm, has been proposed by La Mura and Swiatczak [35], to which similar comments apply.

There has also been work on Quantum Markov networks within the quantum probability literature [26,2,1], although Belief Propagation has not been investigated in this literature. This is closer to the spirit of the present work, in the sense that it is based on the generalization of classical probability to a noncommutative, operator-valued probability theory. These works are primarily concerned with defining the Markov condition in such a way that it can be applied to systems with an infinite number of degrees of freedom, and hence an operator algebraic formalism is used. This is important for applications to statistical physics because the thermodynamic limit can be formally defined as the limit of an infinite number of systems, but it is not so important for numerical simulations, since these necessarily operate with a finite number of discretized degrees of freedom. Also conditional independence is defined in a different way via quantum conditional expectations, rather than the approach based on conditional mutual information and conditional density operators used in the present work. Nevertheless, it seems likely that there are connections to our approach that should to be investigated in future work.

Lastly, during the final stage of preparation of this manuscript, two related papers have appeared on the physics archive. An article by Laumann, Scardicchio, and Sondhi [23] proposed a quantum generalization of the cavity method to study disordered quantum systems on trees. The cavity method is a special instance of Belief Propagation algorithm with applications in statistical physics. The quantum generalization was made using what we have called here the method off replicas. Hastings [19] proposed a QBP algorithm for

the simulation of quantum many-body systems based on ideas similar to the ones presented here. The connection between the two approaches, and in particular the application of the Lieb–Robinson bound [30] to conditional mutual information, is worthy of further investigation.

9. Conclusion

In this paper, we have presented quantum Graphical Models and Belief Propagation based on the idea that quantum theory is a noncommutative, operator-valued, generalization of probability theory. Our main results are summarized in Fig. 10. We expect these methods to have significant applications in quantum error correction and the simulation of many-body quantum systems. We are currently in the process of implementing these algorithm numerically in both of these contexts. Belief Propagation based decoding of several types of quantum error correction codes has already been implemented quite successfully, e.g. on concatenated block codes [42], turbo-codes [39], and sparse codes [10]. However, for the noise models considered there, the corresponding bifactor states only involve commuting operators and thus the corresponding inference problem could be solved by means of a classical Belief Propagation algorithm. We conclude with several open questions suggested by this work.

In the context of many-body physics, it would be interesting to relate the class of solutions obtained by QBP to other approximation schemes used in statistical physics, much in the spirit of the work of Yedidia [58] in the classical setting. A related problem would be to understand how the different classes of bifactor states relate to each other. We suspect that when the Hilbert space dimension at each vertex of the graph is held fixed, the n-bifactor states on that graph form a subset of the m-bifactor states when $n < m$. If that conjecture were true, it might lead to a family of approximation schemes converging to the correct solution. It would also reveal an interesting discrepancy between the classical and quantum settings. Classically, the problem of computing correlation functions in a disordered

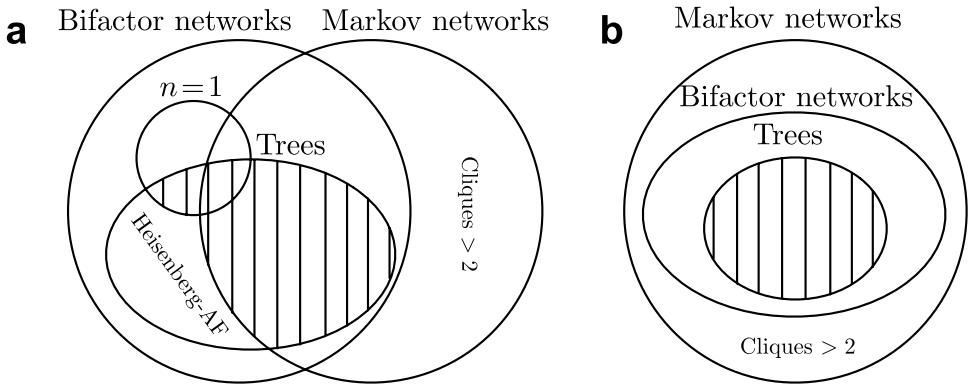

Fig. 10. Relation between Markov Networks, Bifactor Networks, and 1-Bifactor Networks in (a) quantum theory and (b) classical probability theory. The hashed regions indicate the domain of convergence of the associated Belief Propagation algorithms. (a) Convergence of Belief Propagation on trees for Markov Networks is Theorem 5.6 and for 1-Bifactor states is Corollary 5.7. That all Markov Networks on trees are Bifactor states is Theorem 4.10. The existence of Bifactor Networks on trees that are not Markov Networks is given by Example 3.7 for $n < \infty$ and the Heisenberg anti-ferromagnetic spin chain of Example 4.8 for $n = \infty$. Markov Networks on trees with cliques of size > 2 are generally not Bifactor Networks, c.f. Theorem 4.7. (b) That all classical Bifactor Networks are Markov Networks is the Hammersley–Clifford Theorem 4.1, and convergence of Belief Propagation on trees follows from Theorem 5.6.

many-body system and the problem of decoding an error correction code are equivalent. If our conjecture holds true, in the quantum case the latter is simpler than the former.

Whilst our definition of a quantum Markov Network is well motivated as a direct analog of a classical Markov Network, it does not seem to represent the most general class of states to which our Belief Propagation algorithms are applicable. In particular, in Section 5.2 it was shown that QBP converges on trees for arbitrary bifactor states defined with respect to the $\star$ product. One reason for this discrepancy might be that the quantum conditional independence condition, $I_\rho(U, W|X)$, only allows classical correlations to be mediated between U and W via X, i.e. $\rho_{U\cup W}$ is always separable, whereas the classical condition $I_P(U, W|X)$ is compatible with an arbitrary distribution $P(U \cup W)$. This suggests that quantum conditional independence could be relaxed to a condition that allows quantum correlations, i.e. entanglement, to be mediated by X, whilst still preserving the validity of Belief Propagation. It would be interesting to find a condition like this that also satisfies the graphoid axioms, so that it could naturally be represented on a graph.

Nevertheless, quite apart from their application in Belief Propagation algorithms, the mathematical structures investigated in this work should be of interest in other areas of quantum information and computation. Firstly, the characterizations of quantum conditional independence in terms of conditional density operators given in Section 3.3 should be useful, and indeed are currently being applied to the problem of pooling quantum states [29]. Another interesting area of investigation would be the computational complexity of inference on quantum Markov Networks. In the classical case, it is fairly straightforward to find families of Markov Networks that encode instances of NP complete problems, such as satisfiability or graph colorability. Therefore, one would expect to be able to encode problems that are similarly hard for quantum computers, i.e. complete for the complexity class QMA, as inference problems on quantum Markov Networks. This should be closely related to the quantum marginals problem, which has recently be proved to be QMA-complete [5,21].

Finally, this work leaves open the question of fully characterizing quantum Markov Networks. The most generally applicable result given here is Theorem 4.7, which is a direct analog of one direction of the classical Hammersley–Clifford theorem using the $\odot$ product. A full characterization would provide a converse to this theorem, i.e. a set of conditions on the operators in Eq. (92), satisfied by the construction used in the proof, such that all states of this form are guaranteed to satisfy the Markov condition. Analogous theorems for the $\star^{(n)}$ products would also be useful. This work also leaves open the question of intersection for quantum conditional mutual information, i.e. whether $S(U : W|X \cup Y) = 0$ and $S(U : Y|W \cup X) = 0$ imply $S(U : W \cup Y|X) = 0$ for strictly positive states. This result would imply that positive quantum Markov networks obey the global Markov property.

Acknowledgments

M.L. thanks Rob Spekkens for useful discussions about quantum conditional independence. D.P. is grateful to Harold Ollivier for many stimulating discussions on Belief Propagation.

At IQC, M.L. was supported in part by MITACS and ORDCF. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI. M.L. was also supported in part by Grant RFP1-06-006 from The Foundational Questions Institute (fqxi.org).

D.P. is supported in part by the Gordon and Betty Moore Foundation through Caltech's Center for the Physics of Information, by the National Science Foundation under Grant No. PHY-0456720, and by the Natural Sciences and Engineering Research Council of Canada.

Appendix A. A useful notation for probability distributions and density matrices

A.1. Probability distributions

In standard Kolmogorov probability theory for finite sample spaces, probabilities are given by a measure μ on a sample space $(\Omega, 2^{\Omega})$, where Ω is a set of elementary events and 2^{Ω} is the power set, i.e. the set of all subsets of Ω. Specifically, $\mu : 2^{\Omega} \to [0,1]$ and satisfies the axioms

$$\forall \Lambda \in 2^{\Omega}, \quad 0 \leqslant \mu(\Lambda) \leqslant 1 \tag{162}$$

$$\mu(\Omega) = 1 \tag{163}$$

$$\text{If } \Lambda_1, \Lambda_2, \ldots, \Lambda_d \text{ are disjoint sets in } 2^{\Omega} \text{ then } \mu(\cup_{j=1}^{d} \Lambda_j) = \sum_{j=1}^{d} \mu(\Lambda_j). \tag{164}$$

In particular, this implies that $\mu(\emptyset) = 0$ and $\forall \Lambda_1, \Lambda_2 \in 2^{\Omega}$,

$$\mu(\Lambda_1 \cup \Lambda_2) \geqslant \mu(\Lambda_1) \tag{165}$$

$$\mu(\Lambda_1 \cap \Lambda_2) \leqslant \mu(\Lambda_1) \tag{166}$$

$$\text{If } \Lambda_1 \subseteq \Lambda_2 \text{ then } \mu(\Lambda_1) \leqslant \mu(\Lambda_2). \tag{167}$$

The conditional probability of Λ_2, given Λ_1 is defined to be

$$\text{Prob}\,(\Lambda_2|\Lambda_1) = \frac{\mu(\Lambda_1 \cap \Lambda_2)}{\mu(\Lambda_1)} \tag{168}$$

provided $\mu(\Lambda_1) \neq 0$ and is undefined otherwise. In particular, for any $\Lambda \in 2^{\Omega}$, this means that $\text{Prob}(\Lambda|\emptyset)$ is always undefined and that $\text{Prob}(\Lambda|\Omega) = \mu(\Lambda)$.

Our notation for probability distributions over random variables works in an almost exactly opposite way to the Kolmogorov conventions, but is very convenient for the discussion of Graphical Models. For a random variable v that takes a finite number of possible values, write $P(v)$ for the probability distribution of v. For definiteness, suppose that v takes integer values $\{1, 2, \ldots, d\}$. Then, a sample space can be associated with v by setting $\Omega_v = \{v = 1, v = 2, \ldots, v = n\}$, and a measure $\mu : 2^{\Omega_v} \to [0,1]$ can be defined on this space. The notation $P(v)$ is a stand in for $\mu(v = j)$ when j is an arbitrary unspecified value. To give some precise examples of how this works, let f be a function with domain $\{1, 2, \ldots, d\}$ and let g be a function with domain $[0,1]$. Then, the expression $g(P(v)) = f(v)$ is interpreted as $\forall j, g(\mu(v = j)) = f(j)$, and the expression $\sum_v g(P(v))f(v)$ is interpreted as $\sum_j g(\mu(v = j))f(j)$. It is straightforward to see how this generalizes to more complicated examples.

Now consider the case of two random variables v, w for which we can set up sample spaces Ω_v and Ω_w as above. Joint probabilities are given by a measure μ on the sample space $(\Omega_v \times \Omega_w, 2^{\Omega_v \times \Omega_w})$. The notation $P(v,w)$ stands for $\mu(v = j \times w = k)$, where both j and k are arbitrary unspecified values. Note that

$$\mu(v = j \times w = k) = \mu((v = j \times \Omega_w) \cap (\Omega_v \times w = k)). \tag{169}$$

The notation $P(v, w)$ can be made precise in the same way as the examples given above for a single variable, but two additional definitions are worthy of note. Firstly, the marginal probability of v is written as $P(v) = \sum_w P(v, w)$ and this corresponds to the equation

$$\mu(v = j \times \Omega_w) = \sum_k \mu(v = j \times w = k). \tag{170}$$

Secondly, the conditional probability of w given v is written as $P(w|v) = \frac{P(v,w)}{P(v)}$, which corresponds to

$$\begin{aligned} \text{Prob}\ (\Omega_v \times w = k | v = j \times \Omega_w) &= \frac{\mu((v = j \times \Omega_w) \cap (\Omega_v \times w = k))}{\mu(v = j \times \Omega_w)} \\ &= \frac{\mu(v = j, w = k)}{\mu(v = j \times \Omega_w)}. \end{aligned} \tag{171}$$

The generalization of this to arbitrary numbers of random variables is straightforward.

The present notation can be extended to a set of random variables $V = \{v_1, v_2, \ldots, v_N\}$, where v_j is a random variable taking values in $\{1, 2, \ldots, d_j\}$. Consider the joint probability distribution of an arbitrary subset $U \subseteq V$. Let $I = \{i_1, i_2, \ldots, i_M\}$ be the index set of U, i.e. the subset of $\{1, 2, \ldots, N\}$ consisting of the indices of the v_j's that are contained in U. Then define $P(U) = P(v_{i_1}, v_{i_2}, \ldots, v_{i_M})$. This implies that $P(\emptyset) = 1$, which is opposite to the Kolmogorov convention for events, but recall that here $\emptyset$ is an empty set of random variables rather than an event in a sample space. To see this, note that the expression $P(U)$ may be read as meaning that the variables in U are constrained to take particular values, whilst the variables in $V - U$, the relative complement of V in V, may take any value. Thus $P(\emptyset)$ is the probability of the event corresponding to no constraints, i.e. the entire sample space. More precisely, if we define $K = \{k_1, k_2, \ldots, k_{N-M}\}$ to be the index set of $V - U$ and let $j_1, j_2, \ldots, j_M$ be particular instantiations of $v_{i_1}, v_{i_2}, \ldots, v_{i_M}$, then $P(U)$ corresponds to $\mu(v_{i_1} = j_1 \times v_{i_2} = j_2 \times \cdots \times v_{i_M} = j_M \times \Omega_{v_{k_1}} \times \Omega_{v_{k_2}} \times \cdots \times \Omega_{k_{N-M}})$. Thus, for $U = \emptyset$ we have $P(\emptyset) = \mu(\Omega_{v_1} \times \Omega_{v_2} \times \cdots \times \Omega_{v_N}) = 1$ via the standard Kolmogorov axioms.

All the usual set theoretic notions can be applied at the level of random variables, and it is straightforward to verify that the following relations hold for all $U, W \subseteq V$

$$P(U \cup W) \leqslant P(U) \tag{172}$$

$$P(U \cap W) \geqslant P(U) \tag{173}$$

$$\text{If } U \subseteq W \text{ then } P(U) \leqslant P(W). \tag{174}$$

Conditional probabilities $P(W|U)$ are only well defined for disjoint subsets, so $P(W|V)$ is always undefined and $P(W|\emptyset) = P(W)$.

Finally, note that this notation introduces an ambiguity for singleton sets $\{v\}$, since $P(v)$ and $P(\{v\})$ denote the same object. These are used interchangeably and set theoretic operations like $U \cup \{v\}$ are denoted $U \cup v$ when this does not cause ambiguity.

A.2. Density matrices

For quantum theory, the corresponding notation is obtained by replacing random variables v with finite-dimensional Hilbert spaces $\mathcal{H}_v$ and $P(v)$ with a density matrix ρ_v acting on $\mathcal{H}_v$. The density matrix ρ_v is referred to as the state of system v, with the fact that it is defined on a corresponding Hilbert space $\mathcal{H}_v$ left implicit. If we have a set V of N quantum

systems $V = \{v_1, v_2, \ldots, v_N\}$, then the state ρ_V is defined on the Hilbert space $\mathcal{H}_{v_1} \otimes \mathcal{H}_{v_2} \otimes \cdots \otimes \mathcal{H}_{v_N}$. For an arbitrary subset $U \subseteq V$, the state ρ_U is defined to be the partial trace of ρ_V over all the systems in $V - U$. With this convention, $\emptyset$ is associated with the trivial Hilbert space $\mathbb{C}$, so that $\rho_\emptyset = 1$. It is convenient to suppress tensor products with identity operators in order to equate operators acting on different subsets of V. Explicitly, if $U, W \subseteq V$ and A_U and B_W are operators acting on $\mathcal{H}_U$ and $\mathcal{H}_W$, respectively, then $A_U = B_W$ is defined to mean $A_U \otimes I_{W-(U\cap W)} = B_W \otimes I_{U-(U\cap W)}$. Generally, identity operators are omitted in this way unless their presence is required to clarify an argument.

Appendix B. Proof of Theorem 4.7

Lemma B.1. *Let V be a collection of quantum systems with Hilbert space $\mathcal{H}_V = \otimes_{v\in V}\mathcal{H}_v$ and let H_V be an operator on $\mathcal{H}_V$. Let $|\alpha\rangle_v \in \mathcal{H}_v$ be a set of pure states, where $|\alpha\rangle_v$ may be a different state for each v, and for $U \subseteq V$ define $|\alpha\rangle_U = \otimes_{v\in U}|\alpha_v\rangle$. For all $U \subseteq V$ define*

$$J_U = \langle\alpha|_{V-U} H_V |\alpha\rangle_{V-U} \otimes I_{V-U}, \tag{175}$$

where $V - U$ denotes the relative complement of U in V, and

$$K_U = \sum_{W\subseteq U} (-1)^{|U-W|} J_W, \tag{176}$$

where $|\cdot|$ denotes the order, i.e. number of elements contained in, a set. Then,

$$H_V = \sum_{U\subseteq V} K_U. \tag{177}$$

Proof. Consider the double sum expression obtained by substituting Eq. (176) into the right hand side of Eq. (177).

$$\sum_{U\subseteq V}\sum_{W\subseteq U} (-1)^{|U-W|} J_W. \tag{178}$$

Note that the coefficient of J_W in this expression is

$$\sum_{\{U:W\subseteq U\subseteq V\}} (-1)^{|U-W|} = \sum_{X\subseteq(V-W)} (-1)^{|X|}. \tag{179}$$

If $W = V$ then $\emptyset$ is the only subset of $V - W$, so the last sum reduces to $(-1)^0 = 1$. The corresponding term in Eq. (178) is just H_V, so it just remains to prove that all the other terms sum to 0. For $W \neq V$, choose an arbitrary element $v \in (V - W)$. Let $\mathfrak{X} = \{X \subseteq (V - W) | v \notin X\}$ and let $\widetilde{\mathfrak{X}} = \{X \subseteq (V - W) | v \in X\}$. For each $X \in \mathfrak{X}$, define $\widetilde{X} \in \widetilde{\mathfrak{X}}$ via $\widetilde{X} = X \cup \{v\}$. This correspondence is a bijection, so exactly half of the subsets of $V - W$ contain v and the other half do not contain v. Further, if $X \in \mathfrak{X}$ has even order then $\widetilde{X}$ has odd order, and if $X \in \mathfrak{X}$ has odd order then $\widetilde{X}$ has even order. Thus, there are an equal number of odd and even order subsets of $V - W$, so the right hand side of Eq. (179) is zero. □

Lemma B.2. *Let V be a collection of quantum systems with Hilbert space $\mathcal{H}_V = \otimes_{v\in V}\mathcal{H}_v$ and let H_V be an operator on $\mathcal{H}_V$. Let $|\alpha\rangle_v \in \mathcal{H}_v$ be a set of pure states. For nonempty $U \subseteq V$ define K_U as in Eq. (176) and let $u \in U$. Then*

$$\langle\alpha|_u K_U |\alpha\rangle_u = 0 \tag{180}$$

Proof. Let $W \subseteq U$. If $u \notin W$ then $\langle\alpha|_u J_W |\alpha\rangle_u = \langle\alpha|_{V-W} H_V |\alpha\rangle_{V-W} I_{V-(W\cup\{u\})}$. Also,

$$\langle\alpha|_u J_{W\cup\{u\}} |\alpha\rangle_u = \langle\alpha|_{V-(W\cup\{u\})} \langle\alpha|_u H_V |\alpha\rangle_{V-(W\cup\{u\})} |\alpha\rangle_u I_{V-(W\cup\{u\})} \tag{181}$$

$$= \langle\alpha|_{V-W} H_V |\alpha\rangle_{V-W} \otimes I_{V-(W\cup\{u\})} \tag{182}$$

$$= \langle\alpha|_u J_W |\alpha\rangle_u. \tag{183}$$

From the same argument that was used in Lemma B.1, the element u divides the subsets of U into pairs, i.e. those that do not contain u and those obtained by adding u to such a set. As shown above, the operator obtained by projecting the J operator onto $|\alpha\rangle_u$ is the same for each such pair of subsets, but they enter into Eq. (176) with opposite sign and so the corresponding terms in $\langle\alpha|_u K_U |\alpha\rangle_u$ cancel. □

Proof (*Proof of Theorem 4.7*)**.** Apply Lemma B.1 with $H_V = \log \rho_V$ and set $\sigma_U = \exp(K_U)$ for all $U \subseteq V$. Rewriting Eq. (177) in terms of these operators gives

$$\rho_V = \bigodot_{U\subseteq V} \sigma_U. \tag{184}$$

It remains to show that σ_U is the identity whenever $U \notin \mathfrak{C}$, which is equivalent to proving that $K_U = 0$.For any $U \notin \mathfrak{C}$, we can find two vertices $u, t \in U$ that are not connected by an edge. In particular, this means that $t \notin n(u)$. Then, the Markov condition, $I(\{u\} : V - (\{u\} \cup -n(u))|n(u))$, implies that

$$\log \rho_{u|V-\{u\}} = \log \rho_{u|n(u)} \otimes I_{V-(\{u\}\cup n(u))} \tag{185}$$

$$= \log \rho_{u|n(u)} \otimes I_{V-(\{u\}\cup\{t\}\cup n(u))} \otimes I_t. \tag{186}$$

Now, let $\mathfrak{U} = \{W \subseteq U | u \notin W\}$ and let $\widetilde{\mathfrak{U}} = \{W \subseteq U | u \in W\}$. As before, every $W \in \mathfrak{U}$ is in one-to-one correspondence with a $\widetilde{W} \in \widetilde{\mathfrak{U}}$ defined by $\widetilde{W} = W \cup \{u\}$, and so Eq. (176) may be rewritten as

$$K_U = \sum_{W\in\mathfrak{U}} (-1)^{|U-W|} (J_W - J_{\widetilde{W}}). \tag{187}$$

Next, consider a particular W and the corresponding term $J_W - J_{\widetilde{W}}$. Using the standard rules of conditional density operators,

$$J_W = \langle\alpha|_{V-W} \log \rho_V |\alpha\rangle_{V-W} \otimes I_{V-W} \tag{188}$$

$$= \langle\alpha|_{V-W} \log \rho_{u|V-u} |\alpha\rangle_{V-W} \otimes I_{V-W} + \langle\alpha|_{V-W} \log \rho_{V-u} \otimes I_u |\alpha\rangle_{V-W} \otimes I_{V-W} \tag{189}$$

$$= \langle\alpha|_{V-W} \log \rho_{u|V-u} |\alpha\rangle_{V-W} \otimes I_{V-W} + \langle\alpha|_{V-(W\cup\{u\})} \log \rho_{V-u} |\alpha\rangle_{V(W\cup\{u\})} \otimes I_{V-(W\cup\{u\})} \otimes I_u. \tag{190}$$

Similarly, $J_{\widetilde{W}}$ may be written as

$$J_{\widetilde{W}} = \langle\alpha|_{V-\widetilde{W}} \log \rho_V |\alpha\rangle_{V-\widetilde{W}} \otimes I_{V-\widetilde{W}} \tag{191}$$

$$= \langle\alpha|_{V-(W\cup\{u\})} \log \rho_V |\alpha\rangle_{V-(W\cup\{u\})} \otimes I_{V-(W\cup\{u\})} \tag{192}$$

$$= \langle\alpha|_{V-(W\cup\{u\})} \log \rho_{u|V-u} |\alpha\rangle_{V-(W\cup\{u\})} \otimes I_{V-(W\cup\{u\})} + \langle\alpha|_{V-(W\cup\{u\})} \log \rho_{V-u} \otimes I_u |\alpha\rangle_{V-(W\cup\{u\})} \otimes I_{V-(W\cup\{u\})} \tag{193}$$

$$= \langle\alpha|_{V-(W\cup\{u\})} \log \rho_{u|V-u} |\alpha\rangle_{V-(W\cup\{u\})} \otimes I_{V-(W\cup\{u\})} + \langle\alpha|_{V-(W\cup\{u\})} \log \rho_{V-u} |\alpha\rangle_{V-(W\cup\{u\})} \otimes I_{V-(W\cup\{u\})} \otimes I_u. \tag{194}$$

The last terms (190) and (194) are identical, so they cancel in $J_W - J_{\widetilde{W}}$. Therefore, $J_W - J_{\widetilde{W}}$ is just the difference of (190) and (194). The remainder of the proof show that $\langle\alpha|_t J_W - J_{\widetilde{W}}|\alpha\rangle_t \otimes I_t = J_W - J_{\widetilde{W}}$. From this it follows that $\langle\alpha|_t K_U|\alpha\rangle_t \otimes I_t = K_U$, but Lemma B.2 shows that $\langle\alpha|_t K_U|\alpha\rangle_t = 0$, so this is enough to complete the proof.

There are two cases to deal with, either $t \notin W$ or $t \in W$. When $t \notin W$, both $V - W$ and $V - (W \cup \{u\})$ contain t. The effect of projecting out $|\alpha\rangle_t$ on terms (190) and (194) is to replace I_{V-W} and $I_{V-(W\cup\{u\})}$ with $I_{V-(W\cup\{t\})}$ and $I_{V-(W\cup\{u\}\cup\{t\})}$, respectively, but then tensoring with I_t restores the original identity operator so both terms are unaffected. In the case where $t \in W$, we make use of the Markov condition in the form of Eq. (186). The important point is that $\rho_{u|V-u}$ is of the form $\tau_{V-t} \otimes I_t$, so projecting out $|\alpha_t\rangle$ and retensoring with I_t again has no effect on the terms (190) and (194). □

References

[1] L. Accardi, F. Fidaleo, Markov Property—Recent Developments on the Quantum Markov Property, World Scientific, 2003.
[2] L. Accardi, F. Fidaleo, J. Func. Anal. 200 (2003) 324.
[3] I. Affleck, T. Kennedy, E. Lieb, H. Tasaki, Commun. Math. Phys. 115 (1988) 477.
[4] I. Affleck, T. Kennedy, E.H. Lieb, H. Tasaki, Phys. Rev. Lett. 59 (1987) 799.
[5] D. Aharonov, D. Gottesman, J. Kempe, The power of quantum systems on a line, 2007. Available from: <arXiv:0705.4077>.
[6] S. Aji, R. McEliece, IEEE Trans. Info. Theor. 46 (2000) 325.
[7] M. Asorey, A. Kossakowski, G. Marmo, E.C.G. Sudarshan, Open Sys. Info. Dyn. 12 (2006) 319. Available from: <quant-ph/0602228>.
[8] C. Berrou, A. Glavieux, P. Thitimajshima, Near shannon limit error-correcting coding and decoding, in ICC'93, Genève, Switzerland, May 1993, pp. 1064–1070.
[9] J. Besag, J. R. Stat. Soc. B 36 (1974) 192.
[10] T. Camara, H. Ollivier, J.-P. Tillich, Constructions and performance of classes of quantum ldpc codes, 2005. Available from: <quant-ph/0502086>.
[11] N.J. Cerf, C. Adami, Phys. Rev. Lett. 79 (1997) 5194–5197.
[12] J. Dukelsky, M. Martín-Delgado, T. Nishino, G. Sierra, Europhys. Lett. 43 (1998) 457.
[13] M. Fannes, B. Nachtergaele, R. Werner, Commun. Math. Phys. 144 (1992) 443.
[14] R.G. Gallager, Low Density Parity Check Codes, M.I.T. Press, Cambridge, Massachusetts, 1963.
[15] D. Geiger, T. Verma, J. Pearl, Networks 20 (1990) 507.
[16] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, California Institute of Technology, Pasadena, CA, 1997. Available from: <quant-ph/9705052>.
[17] G. Grimmet, Bull. Lond. Math. Soc. 5 (1973) 81.
[18] J. Hammersley, P. Clifford, Markov fields on finite graphs and lattices. 1971.
[19] M. Hastings, Quantum belief propagation, 2007. Available from: <arXiv:0706.4094>.
[20] P. Hayden, R. Jozsa, D. Petz, A. Winter, Commun. Math. Phys. 246 (2004) 359. Available from: <quant-ph/0304007>.

[21] S. Irani, The complexity of quantum systems on a one-dimensional chain, 2007. Available from: <arXiv:0705.4067>.
[22] F.-R. Kschischang, B.J. Frey, H.-A. Loeliger, IEEE Trans. Info. Theor. 47 (2001) 498–519.
[23] C. Laumann, A. Scardicchio, S. Sondhi, Cavity method for quantum spin glasses on the Bethe lattice, 2007. Available from: <arXiv:0706.4391>.
[24] S. Lauritzen, Graphical Models, Oxford University Press, 1996.
[25] S. Lauritzen, Fundamentals of graphical models. Lecture notes. Available from: <http://www.stats.ox.ac.uk/steffen/stflour/>, 2006.
[26] V. Leibscher, Markovianity of quantum random fields in the $\mathcal{B}(\mathcal{H})$ case in: W. Freudenberg (Ed.), Proceedings of the Conference on Quantum Probability and Infinite-Dimensional Analysis, Burg, Germany, 15–20 March 2001. Quantum Probability and White Noise Analysis, vol. 15, World Scientific, Singapore, 2003, pp. 151–159.
[27] M.S. Leifer, Conditional density operators and the subjectivity of quantum operations, in: G. Adenier (Ed.), Foundations of Probability and Physics-4, vol. 899, AIP Conference Proceedings, 2006, p. 172. Available from: <quant-ph/0611233>.
[28] M.S. Leifer, Phys. Rev. A 74 (2006) 042310.
[29] M.S. Leifer, R. Spekkens, in preparation.
[30] E. Lieb, D. Robinson, Commun. Math. Phys. 28 (1972) 251.
[31] D.J.C. MacKay, Information Theory, Inference and Learning Algorithms, Cambridge University Press, Cambridge, UK, 2003.
[32] M. Mézard, A. Montanari, Constraint Satisfaction Networks in Physics and Computation, Clarendon Press, 2007.
[33] M. Mézard, G. Parisi, Eur. Phys. J. B 20 (2001) 217.
[34] M. Mézard, G. Parisi, R. Zecchina, Science 297 (2002) 812.
[35] P.L. Mura, L. Swiatczak, Markovian entangled networks, 2007. Available from: <quant-ph/0702072>.
[36] R. Neapolitan, Probabilistic Reasoning in Expert Systems: Theory and Algorithms, Wiley, 1990.
[37] R. Neapolitan, Learning Bayesian Networks, Pearson Prentice Hall, 2004.
[38] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, Cambridge, UK, 2000.
[39] H. Ollivier, D. Poulin, J.-P. Tillich, Quantum turbo codes, in preparation.
[40] S. Östlund, S. Rommer, Phys. Rev. Lett. 75 (1995) 3537–3540.
[41] D. Pollard, Stochastic processes. Lecture notes. Available from: <http://www.stat.yale.edu/pollard/251.spring04/>, 2004.
[42] D. Poulin, Phys. Rev. A 74 (2006) 052333. Available from: <quant-ph/0606126>.
[43] J. Preskill, Lecture notes for physics 229: Quantum information and computation, 1999.
[44] T. Richardson, R. Urbanke, Modern Coding Theory, 2005. Available from: <http://lthcwww.epfl.ch/mct/index.php>.
[45] M.B. Ruskai, J. Math. Phys. 43 (2002) 4358.
[46] Y.-Y. Shi, L.-M. Duan, G. Vidal, Phys. Rev. A 74 (2006) 022320.
[47] R.R. Tucci, Int. J. Mod. Phys. B 9 (1995) 295. Available from: quant-ph/9706039.
[48] R.R. Tucci, How to compile a quantum Bayesian net, 1998. Available from: <quant-ph/9805016>.
[49] R.R. Tucci, Factorization of quantum density matrices according to bayesian and markov networks, 2007. Available from: <quant-ph/0701201>.
[50] A. Uhlmann, Commun. Math. Phys. 54 (1977) 21.
[51] T. Verma, J. Pearl, Causal networks: semantics and expressiveness, in: R.D. Shachter (Ed.), Uncertainty in Artificial Intelligence, vol. 4, Elsevier, 1990, p. 69.
[52] F. Verstraete, J.I. Cirac, Renormalization algorithms for quantum-many body systems in two and higher dimensions, 2004. Available from: <cond-mat/0407066>.
[53] F. Verstraete, D. Porras, J.I. Cirac, Phys. Rev. Lett. 93 (2004) 227205.
[54] G. Vidal, Phys. Rev. Lett. 93 (2004) 040502.
[55] G. Vidal, Phys. Rev. Lett. 98 (2006) 070201.
[56] M. Warmuth, Bayes rule for density matricesAdvances in Neural Information Processing Systems, vol. 18, MIT Press, 2005.
[57] S.R. White, Phys. Rev. Lett. 69 (1992) 2863.
[58] J.S. Yedidia, Advanced mean field methods: theory and practice, in: An Idiosyncratic Journey Beyond Mean Field Theory, MIT Press, 2001, p. 21.
[59] J.S. Yedidia, W.T. Freeman, Y. Weiss, Understanding belief propagation and its generalizations, Technical Report TR-2001-22, Mitsubishi Electric Reasearch Laboratories, 2002.