2016 IEEE 57th Annual Symposium on Foundations of Computer Science

Zero-knowledge proof systems for QMA

(Extended Abstract)

Anne Broadbent[1] Zhengfeng Ji[2] Fang Song[3] John Watrous[4]

[1]*Department of Mathematics and Statistics, University of Ottawa, Canada*

[2]*Centre for Quantum Computation and Intelligent Systems, School of Software*
Faculty of Engineering and Information Technology, University of Technology Sydney, Australia; and
State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, China

[3]*Institute for Quantum Computing and Department of C&O, University of Waterloo, Canada; and*
Computer Science Department, Portland State University, U.S.A.

[4]*Institute for Quantum Computing and School of Computer Science, University of Waterloo, Canada; and*
Canadian Institute for Advanced Research, Toronto, Canada

***Abstract*—Prior work has established that all problems in NP admit classical zero-knowledge proof systems, and under reasonable hardness assumptions for quantum computations, these proof systems can be made secure against quantum attacks. We prove a result representing a further quantum generalization of this fact, which is that every problem in the complexity class QMA has a quantum zero-knowledge proof system. More specifically, assuming the existence of an unconditionally binding and quantum computationally concealing commitment scheme, we prove that every problem in the complexity class QMA has a quantum interactive proof system that is zero-knowledge with respect to efficient quantum computations. Our QMA proof system is sound against arbitrary quantum provers, but only requires an honest prover to perform polynomial-time quantum computations, provided that it holds a quantum witness for a given instance of the QMA problem under consideration.**

I. Introduction

Zero-knowledge proof systems, first introduced by Goldwasser, Micali and Rackoff [1], are interactive protocols that allow a prover to convince a verifier of the validity of a statement while revealing no additional information beyond the statement's validity. Although this notion may initially have seemed paradoxical, several problems that are not known to be efficiently computable, such as the Quadratic Residuosity and Graph Isomorphism problems (and their complements), were shown to admit zero-knowledge proof systems [1], [2]. Under reasonable intractability assumptions, Goldreich, Micali and Wigderson [2] gave a zero-knowledge protocol for the Graph 3-Coloring problem and, because of its NP-completeness, for all NP problems. This line of work was further extended in [3], which showed that all problems in IP have zero-knowledge proof systems.

Since the invention of this concept, zero-knowledge proof systems have become a cornerstone of modern theoretical cryptography. In addition to the conceptual innovation of formulating a complexity-theoretic notion of knowledge, zero-knowledge proof systems are essential building blocks in a host of cryptographic constructions, such as the design of secure two-party and multi-party computation protocols [4].

The extensive works on zero-knowledge largely reside in a classical world, but the development of quantum information science has urged another look at the landscape of zero-knowledge proof systems in a *quantum* world. Namely, both honest users and adversaries may potentially possess the capability to exchange and process quantum information. There are, of course, zero-knowledge protocols that immediately become insecure in the presence of quantum attacks due to efficient quantum algorithms that break the intractability assumptions upon which these protocols rely. For instance, Shor's quantum algorithms for factoring and computing discrete logarithms [5] invalidate the use of these problems, generally conjectured to be classically hard, as a basis for the security of zero-knowledge protocols against quantum attacks. Even with computational assumptions against quantum adversaries, however, it is still highly non-trivial to establish the security of classical zero-knowledge proof systems in the presence of malicious *quantum* verifiers because of a technical reason that we now briefly explain.

The zero-knowledge property of a proof system is concerned with the computations that may be realized through an interaction between a (possibly malicious) verifier and the prover. That is, the malicious verifier may take an arbitrary input (usually called the *auxiliary input* to distinguish it from the input string to the proof system under consideration), interact with the prover in any way it sees fit, and produce an output that is representative of what it has learned through the interaction. Roughly speaking, the prover is said to be *zero-knowledge* on a particular set of input strings if any computation of the sort just described can be efficiently approximated by a *simulator* operating entirely on its own—

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.13

IEEE computer society

meaning that it does not interact with the prover, and in the case of an NP problem it does not possess a witness for the fixed problem instance being considered. The proof system is then said to be zero-knowledge when this zero-knowledge property holds for the set of all yes-instances of the problem under consideration.

Classically speaking, the zero-knowledge property is most typically established through the *rewinding* technique. In essence, the simulator can store a copy of its auxiliary input, and it can make guesses and store intermediate states representing a hypothetical prover/verifier interaction—and if it makes a bad guess or otherwise experiences bad luck when simulating this hypothetical interaction, it simply reverts to an earlier stage (or possibly back to the beginning) of the simulation and tries again. Indeed, it is generally the simulator's freedom to disregard the temporal restrictions of the actual prover/verifier interaction in a way such as this that makes it possible to succeed.

However, rewinding a quantum simulation is more problematic; the *no-cloning theorem* [6] forbids one from copying quantum information, making it impossible to store copies of intermediate states, and measurements generally have an irreversible effect [7] that may partially destroy quantum information. Such difficulties were first observed by van de Graaf [8] and further studied in [9], [10]. Later, a *quantum rewinding* technique was found [11] to establish that several interactive proof systems, including the Goldreich-Micali-Wigderson Graph 3-Coloring proof system [2], remain zero-knowledge against malicious quantum verifiers (under appropriate quantum intractability assumptions in some cases). It follows that all NP problems have zero-knowledge proof systems even against quantum malicious verifiers, provided that a quantum analogue of the intractability assumption required by the Goldreich-Micali-Wigderson Graph 3-Coloring proof system are in place.

This work studies the quantum analogue of NP, known as QMA, in the context of zero-knowledge. These are problems with a succinct *quantum* witness satisfying similar completeness and soundness conditions to NP (or its randomized variant MA). Quantum witnesses and verifications are conjectured to be more powerful than their classical counterparts: there are problems that admit short quantum witnesses, whereas there is no known method for verification using a polynomial-sized classical witness. In other words, NP $\subseteq$ QMA holds trivially, and the containment is typically conjectured to be proper. The question we address in this paper is: *Does every problem in* QMA *have a zero-knowledge quantum interactive proof system?* That is, can one always devise a proof system that reveals nothing about a quantum witness beyond its validity?

We answer this question positively by constructing a quantum interactive proof system, for any problem in QMA, that is zero-knowledge against polynomial-time quantum adversaries, under a reasonable intractability assumption.

Theorem 1. *Assuming the existence of an unconditionally binding and quantum computationally concealing bit commitment scheme, every problem in* QMA *has a quantum computational zero-knowledge proof system.*

A few of the desirable features of our proof system are as follows:

1. Our proof system has a simple structure, similar to the classical Goldreich-Micali-Wigderson Graph 3-Coloring proof system. It can be viewed as a three-phase process: the prover commits to a quantum witness, the verifier makes a random challenge, and finally the prover responds to the challenge by partial opening of the committed information that suffices to certify the validity.
2. All communications in our proof system are classical except for the first commitment message, and the verifier can measure the quantum message immediately upon its arrival (which has a strong technological appeal).
3. Our protocol is based on mild computational assumptions. The sort of bit commitment scheme it requires can be implemented, for instance, under the existence of injective one-way functions that are hard to invert in quantum polynomial time.
4. Our protocol is prover-efficient. It is sound against general quantum provers, but given a valid quantum witness, an honest prover only needs to perform efficient quantum computations. As has already been suggested, aside from the preparation of the first quantum message, all of the remaining computations performed by the honest prover are classical polynomial-time computations.

As a key ingredient of our zero-knowledge proof system, we introduce a new variant of the well-known k*-local Hamiltonian* problem. The k-local Hamiltonian problem asks if the minimum eigenvalue (or ground state energy, in physics parlance) of an n-qubit Hamiltonian $H = \sum_j H_j$, where each H_j is k-local (i.e., acts trivially on all but k of the n qubits), is below a particular threshold value. This problem was introduced and proved to be QMA-complete (for the case $k = 5$) by Kitaev [12]. We show that each H_j can be restricted to be realized by a *Clifford operation*, followed by a standard basis measurement, and the QMA-completeness is preserved. (As is explained in more detail in the full version of this paper, Clifford operations are a discrete subset of unitary quantum operations whose algebraic properties make them both useful and easy to analyze, albeit not universal for quantum computation on their own.) For an arbitrary problem $A \in$ QMA, we can reduce an instance of A efficiently to an instance of the k-local Clifford Hamiltonian problem, and a valid witness for A can also be transformed into a witness for the corresponding k-local Clifford Hamiltonian problem instance by an efficient quantum procedure. As a result, A has a zero-knowledge proof system by composing this reduction with our zero-knowledge proof system for the

k-local Clifford Hamiltonian problem.

Our proof system also employs a new encoding scheme for quantum states, which we construct by extending the *trap scheme* proposed in [13]. While our new scheme can be seen as a *quantum authentication scheme* (cf. [14]–[16]), it in addition allows performing arbitrary constant-qubit Clifford circuits and measuring in the computational basis directly on authenticated data without the need for auxiliary states. The only previously known scheme supporting this feature requires high-dimensional quantum systems (i.e., qudits rather than qubits) [15], which makes it inconvenient in our setting where all quantum operations are on qubits.

II. Overview and comparisons with prior work

We will now give an overview of our protocol and the techniques upon which it relies, and discuss its relationships to existing work. As this is only an extended abstract, and not a full paper, the formal analysis has been omitted—readers interested in this analysis are referred to the full paper [17].

A. Our protocol and techniques

A natural approach to constructing zero-knowledge proof systems for QMA is to consider a quantum analogue of the Goldreich-Micali-Wigderson proof system for Graph 3-Coloring (which we will hereafter refer to as the GMW 3-Coloring proof system). Let us focus in particular on the local Hamiltonian problem, and consider a proof system in which the prover holds a quantum witness state for an instance of this problem, commits to this witness, and receives the challenge from the verifier (which, let us say, is a random term of the local Hamiltonian). The prover might then open the commitments of the set of qubits on which the term acts non-trivially so that the verifier can measure the local energy for this term and accept or reject accordingly.

There is a major difficulty when one attempts to carry out such an approach. The zero-knowledge property of the GMW 3-Coloring proof system depends crucially on a structural property of the problem: the honest prover is free to randomize the three colors used in its coloring, and when the commitments to the colors of two neighboring vertices are revealed, the verifier will see just a uniform mixture over all pairs of different colors. This uniformity of the coloring marginals is important in achieving the zero-knowledge property of the proof system. Unlike the case of 3-Coloring, however, none of the known QMA-complete problems under Karp reductions has such desirable properties. For example, if we use local Hamiltonian problems directly in a GMW-type proof system, of the sort suggested above, information about the reduced state of the quantum witness will be leaked to the verifier, possibly violating the zero-knowledge requirement.

To overcome this difficulty, we employ several ideas that enable the prover to "partially" open the commitments, revealing only the fact that the committed state lives in certain subspaces, and nothing further. Our first technique simplifies the verification circuit for QMA-complete problems through the introduction of the local Clifford-Hamiltonian problem that was already described. Somewhat more specifically, our formulation of this problem requires every Hamiltonian term to take the form $C^*|0^k\rangle\langle 0^k|C$ for some Clifford operation C. Because the local Clifford-Hamiltonian problem remains QMA-complete, it implies a random Clifford verification procedure for problems in QMA: intuitively, the verification of a quantum witness has been simplified to a Clifford measurement followed by a classical verification.

The Clifford verification procedure works in harmony with the encryption of quantum data via the quantum one-time pad [18] and other derived hybrid schemes that are used by our proof system. This has the important effect of transforming statements about quantum states into those about the classical keys of the quantum one-time pad, which naturally leads to our second main idea: the use of zero-knowledge proofs for NP against quantum attacks to simplify the construction of zero-knowledge proofs for QMA. In our protocol, the verifier measures the encrypted quantum data and asks the prover to prove, using a zero-knowledge protocol for NP, that the decryption of this classical data is consistent with the verifier's accepting condition.

In fact, if the verifier measures the quantum data according to the specifications of the protocol, the combination of the Clifford verification and the use of zero-knowledge proofs for NP suffices. A problem arises, however, if the verifier does not perform the honest measurement. Our third technique, inspired by work on quantum authentication [13], [15], [16], [19], employs a new scheme for encoding quantum states. Roughly speaking, if the prover encodes a witness state under our encoding scheme, then the verifier is essentially forced to perform the measurement honestly—any attempt to fake a "logically different" measurement result will succeed with negligible probability. In our proof system, we adapt the trap scheme proposed in [13] so that we can perform any constant-sized Clifford operations on authenticated quantum data followed by computational basis measurements, benefiting along the way from ideas concerning quantum computation on authenticated quantum data.

The resulting zero-knowledge proof system for QMA has a similar overall structure to the GMW 3-Coloring protocol: the prover encodes the quantum witness state using a quantum authentication scheme, and sends the encoded quantum data together with a commitment to the secret keys of the authentication to the verifier. The verifier randomly samples a term $C^*|0^k\rangle\langle 0^k|C$ in the local Clifford-Hamiltonian problem, applies the operation C transversally on the encoded quantum data and measures all qubits corresponding to the k qubits of the selected term in the computational basis, and sends the measurement outcomes to the prover. The prover and verifier then invoke a quantum-secure zero-

knowledge proof for the NP statement that the commitment correctly encodes an authentication key and, under this key, the verifier's measurement outcomes do not decode to 0^k.

B. Comparisons to related work

There has been other work on quantum complexity and theoretical cryptography, some of which is discussed below, that allows one to conclude statements having some similarity to our results. We will argue, however, that with respect to the problem of devising zero-knowledge quantum interactive proof systems for QMA, our main result is stronger in almost all respects. In addition, we believe that our proof system is appealing both because it is conceptually simple and represents a natural extension of well-known classical methods.

1. *Zero-knowledge proof systems for all of* IP. Hallgren, Kolla, Sen, and Zhang [20] proved that classical zero-knowledge proof systems for IP [3] can be made secure against malicious quantum verifiers under a certain technical condition. It appears that this condition holds assuming the existence of a quantum computationally hiding commitment scheme. Because QMA is contained in IP, this would imply a classical zero-knowledge protocol for QMA. However, this generic protocol would require a computationally *unbounded* prover to carry out the honest protocol, and it is unlikely that one can reduce the round complexity without causing unexpected consequences in complexity theory [21]–[23].
2. *Secure two-party computations.* Another approach to constructing zero-knowledge proofs for QMA is to apply the general tool of secure two-party quantum computation [15], [19], [24]. In particular, we may imagine two parties, a prover and a verifier, jointly evaluating the verification circuit of a QMA problem, with the prover holding a quantum witness as his/her private input. In principle, one can design a two-party computation protocol so that the verifier learns the validity of the statement but nothing more about the prover's private input. While we believe that a careful analysis could make this approach work, it comes at a steep cost. First, we need to make significantly stronger computational assumptions, as secure quantum two-party computation relies on (at least) secure computations of classical functions against quantum adversaries. The best-known quantum-secure protocols for classical two-party computation assume quantum-secure dense public-key encryption [25] or similar primitives [26], in contrast to the existence of a quantum computationally hiding commitment scheme. (Roughly speaking, this distinction is analogous to "cryptomania" versus "minicrypt," according to Impagliazzo's five-world paradigm [27].) Secondly, the protocol obtained this way is only an *argument* system. That is, the protocol is only sound against computationally bounded dishonest provers. Moreover, the generic quantum two-party computation protocol evaluates the verification circuit gate by gate, and in particular interactions are unavoidable for some (non-Clifford) gates. This causes the round complexity to grow in proportion to the size of the verification circuit. In addition, the communications are inherently quantum, which makes the protocol much more demanding from a technological viewpoint.

   On the positive side, through this approach, it is possible to achieve negligible soundness error using just one copy of witness state. In contrast, our proof system directly inherits the soundness error of the most natural and direct verification for the local Clifford-Hamiltonian problem (i.e., randomly select a Hamiltonian term and measure). If one reduces an arbitrary QMA-verification procedure to an instance of this problem, the resulting soundness guarantee could be significantly worse.
3. *Zero-knowledge proofs for Density Matrix Consistency.* It was pointed out by Liu [28] that the Density Matrix Consistency problem, which asks if there exists a global state of n qubits that is consistent with a collection of k-qubit density matrix marginals, should admit a simple zero-knowledge proof system following the GMW 3-Coloring approach. This fact was one of the inspirations for our work. While it approaches our main result, it does not necessarily admit a zero-knowledge proof system for all problems in QMA, as the Density Matrix Consistency problem is only known to be hard for QMA with respect to Cook reductions.
4. *Other results on Clifford verifications for* QMA. Clifford verification with classical post-processing of QMA was considered in [29] using so-called *magic states* as an ancillary resource. Our construction is arguably simpler, uses only constant-size Clifford operations, and most importantly does not require any resource states. This helps to avoid checking the correctness of resource states in the final zero-knowledge protocol. One byproduct of our Clifford-Hamiltonian reduction proof is an alternative proof of the single-qubit measurement verification for QMA proposed by [30].

III. The local Clifford-Hamiltonian problem

The k-local Hamiltonian problem (k-LH) is a well-known example of a complete promise problem for QMA, provided that certain assumptions are in place regarding the gap between the ground state energy (i.e., the smallest eigenvalue) of input Hamiltonians for yes- and no-inputs [12]. We introduce a restricted version of the local Hamiltonian in which each Hamiltonian term H_j must be given by a rank 1 projection operator of the form $C_j^*|0^k\rangle\langle 0^k|C_j$, for some choice of a k-qubit Clifford operation C_j. For brevity, we will refer to any such operator as a *k-local Clifford-Hamiltonian projection*. The precise statement of

our problem variant is as follows.

The k-local Clifford-Hamiltonian problem (k-LCH)

Input: A collection $H_1, \ldots, H_m$ of k-local Clifford-Hamiltonian projections, along with positive integers p and q expressed in unary notation (i.e., as strings 1^p and 1^q) and satisfying $2^p > q$.

Yes: It holds that $\langle \rho, H_1 + \cdots + H_m \rangle \leq 2^{-p}$ for some choice of an n-qubit state ρ.

No: It holds that $\langle \rho, H_1 + \cdots + H_m \rangle \geq 1/q$ for every n-qubit state ρ.

Theorem 2. *The 5-local Clifford-Hamiltonian problem is* QMA*-complete with respect to Karp reductions. Moreover, for any promise problem $A = (A_{yes}, A_{no}) \in$ QMA and a polynomially bounded function p, there exists a Karp reduction f from A to 5-LCH having the form*

$$f(x) = \left\langle H_1, \ldots, H_m, 1^{p(|x|)}, 1^q \right\rangle \tag{1}$$

for every $x \in A_{yes} \cup A_{no}$.

The proof is omitted in this extended abstract, and can be found in the complete version of the paper.

Remark 1. If one is given a witness to a given QMA problem A, it is possible to efficiently compute a witness to the corresponding k-local Hamiltonian problem instance through Kitaev's reduction. Our reduction also inherits this property.

Remark 2. States of the form $C|0^k\rangle$, for a Clifford operation C, are stabilizer states of k qubits. Theorem 2 therefore implies that there exists a QMA verification procedure in which the verifier randomly chooses a k-qubit stabilizer state and checks whether the quantum witness state is orthogonal to it.

IV. Description of the proof system

In this section we describe our zero-knowledge proof system for the local Clifford-Hamiltonian problem. The analysis of the proof system is discussed in the section following this one. As suggested previously, our proof system makes use of a bit commitment scheme, and in the interest of simplicity in explaining and analyzing the proof system we shall assume that this scheme is non-interactive. One could, however, replace this non-interactive commitment scheme by a different scheme (such as Naor's scheme with a 1-round commitment phase [31]).

Throughout this section it is to be assumed that an instance of the k-local Clifford-Hamiltonian problem has been selected. The instance describes Clifford-Hamiltonian projections $H_1, \ldots, H_m$, each given by $H_j = C_j^* |0^k\rangle\langle 0^k| C_j$ for k-qubit Clifford operations $C_1, \ldots, C_m$, along with a specification of which of the n qubits these projections act upon. The proof system does not refer to the parameters p and q in the description of the k-local Clifford Hamiltonian problem, as these parameters are only relevant to the performance of the proof system and not its implementation. It must be assumed, however, that the completeness parameter 2^{-p} is a negligible function of the entire problem instance size in order for the proof system to be zero-knowledge.

A. Prover's witness encoding

Suppose $\mathsf{X} = (\mathsf{X}_1, \ldots, \mathsf{X}_n)$ is an n-tuple of single-qubit registers. These qubits are assumed to initially be in the prover's possession, and store an n-qubit quantum state ρ representing a possible witness for the instance of the k-LCH problem under consideration.

The first step of the proof system requires the prover to encode the state of X, using a scheme that consists of four steps. Throughout the description of these steps it is to be assumed that N is a polynomially bounded function of the input size and is an even positive integer power of 7. In effect, N acts as a security parameter (for the zero-knowledge property of the proof system), and we take it to be an even power of 7 so that it may be viewed as a number of qubits that could arise from a concatenated Steane code allowing for a transversal application of Clifford operations (as is described in greater detail in the full version of the paper [17]). In particular, through an appropriate choice of N, one may guarantee that this code has any desired polynomial lower-bound for the minimum non-zero Hamming weight of its underlying classical code.

1. For each $i = 1, \ldots, n$, the qubit X_i is encoded into N qubits by means of the concatenated Steane code. This results in the N-tuples
$$\left(\mathsf{Y}_1^1, \ldots, \mathsf{Y}_N^1\right), \ldots, \left(\mathsf{Y}_1^n, \ldots, \mathsf{Y}_N^n\right). \tag{2}$$
2. To each of the N-tuples in (2), the prover concatenates an additional N *trap qubits*, with each trap qubit being initialized to one of the single qubit pure states $|0\rangle$, $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, or $|\circlearrowleft\rangle = (|0\rangle - i|1\rangle)/\sqrt{2}$, selected independently and uniformly at random. This results in a collection of qubit tuples
$$\mathsf{Y} = \left(\left(\mathsf{Y}_1^1, \ldots, \mathsf{Y}_{2N}^1\right), \ldots, \left(\mathsf{Y}_1^n, \ldots, \mathsf{Y}_{2N}^n\right)\right). \tag{3}$$
The prover stores the string $t = t_1 \cdots t_n$, for $t_1, \ldots, t_n \in \{0, +, \circlearrowleft\}^N$ representing the randomly chosen states of the trap qubits.
3. A random permutation $\pi \in S_{2N}$ is selected, and the qubits in each of the $2N$-tuples (3) are permuted according to π. (Note that it is a single permutation π that is selected and applied to all of the $2N$-tuples simultaneously.)
4. The quantum one-time pad is applied independently to each qubit in (3) (after they are permuted in step 3). That is, for $a_i, b_i \in \{0,1\}^{2N}$ chosen independently and uniformly at random, the unitary transformation $X^{a_i} Z^{b_i}$

is applied to $(\mathsf{Y}_1^i, \ldots, \mathsf{Y}_{2N}^i)$, and the strings a_i and b_i are stored by the prover, for each $i = 1, \ldots, n$.

The randomness required by these encoding steps may be described by a tuple (t, π, a, b), where t is the string representing the states of the trap qubits described in step 2, $\pi \in S_{2N}$ is the permutation applied in step 3, and $a = a_1 \cdots a_n$ and $b = b_1 \cdots b_n$ are binary strings representing the Pauli operators applied in the one-time pad in step 4. After performing the above encoding steps, the prover sends the resulting qubit tuples (3) along with a commitment

$$z = \mathsf{commit}((\pi, a, b), s) \tag{4}$$

to the tuple $\big(\pi, a, b\big)$, to the verifier. Here we assume that s is a random string chosen by the prover that allows for this commitment. (It is not necessary for the prover to commit to the selection of the trap qubit states indicated by t, although it would not affect the properties of the proof system if it were modified so that the prover also committed to the trap qubit state selections.)

B. Verifier's random challenge

Upon receiving the prover's encoded witness and commitment, the verifier issues a challenge: for a randomly selected index $j \in \{1, \ldots, m\}$, the verifier will check that the j-th Hamiltonian term

$$H_j = C_j^* |0^k\rangle\langle 0^k| C_j \tag{5}$$

is not violated. Generally speaking, the verifier's actions in issuing this challenge are as follows: for a certain collection of qubits, the verifier applies the Clifford operation C_j transversally to those qubits, performs a measurement with respect to the standard basis, sends the outcomes to the prover, and then expects the prover to demonstrate that the obtained outcomes are valid (in the sense to be described later).

The randomly selected Hamiltonian term is to be determined by a binary string r, of a fixed length $\lceil \log m \rceil$, that should be viewed as being chosen uniformly at random. (In a moment we will discuss the random choice of r, which will be given by the output of a coin flipping protocol that happens to be uniform for honest participants.) It is not important exactly how the binary strings of length $\lceil \log m \rceil$ are mapped to the indices $\{1, \ldots, m\}$, so long as every index is represented by at least one string—so that for a uniformly chosen string r, each Hamiltonian term j is selected with a nonnegligible probability. We will write H_r and C_r in place of H_j and C_j, and refer to the Hamiltonian term determined by r, when it is convenient to do this.

It would be natural to allow the verifier to randomly determine which Hamiltonian term is to be tested—but, as suggested above, we will assume that the challenge is determined through a *coin flipping protocol* rather than leaving the choice to the verifier. More specifically, it should be assumed that the random choice of the string r that determines which challenge is issued is the result of independent iterations of Blum's coin-flipping protocol [32] (i.e., the honest prover commits to a random $y_i \in \{0, 1\}$, the honest verifier selects $z_i \in \{0, 1\}$ at random, the prover reveals y_i, and the two participants agree that the i-th random bit of r is $r_i = y_i \oplus z_i$). This guarantees (assuming the security of the commitment protocol) that the choices are truly random, and greatly simplifies the analysis of the zero-knowledge property of the proof system. Damgård and Lunemann [33] proved that Blum's coin-flipping protocol is quantum-secure, assuming a quantum-secure commitment scheme. The same analysis actually implies that we can run the basic Blum protocol in parallel to simultaneously generate *logarithmically* many coins, which we use in our proof system. The use of such a subroutine might not actually be necessary for the security of the proof system, but we leave the investigation of whether it is necessary to future work.

Now, let $(i_1, \ldots, i_k)$ denote the indices of the qubits upon which the Hamiltonian term determined by the random string r acts nontrivially. The verifier applies the Clifford operation C_r independently to each of the k-qubit tuples

$$\big(\mathsf{Y}_1^{i_1}, \ldots, \mathsf{Y}_1^{i_k}\big), \ldots, \big(\mathsf{Y}_{2N}^{i_1}, \ldots, \mathsf{Y}_{2N}^{i_k}\big), \tag{6}$$

which is equivalent to saying that C_r is applied transversally to the tuples

$$\big(\mathsf{Y}_1^{i_1}, \ldots, \mathsf{Y}_{2N}^{i_1}\big), \ \ldots, \ \big(\mathsf{Y}_1^{i_k}, \ldots, \mathsf{Y}_{2N}^{i_k}\big) \tag{7}$$

that encode the qubits on which the Hamiltonian term H_r acts nontrivially. The qubits (7) are then measured with respect to the standard basis, and the results are sent to the prover. We will let $u_{i_1}, \ldots, u_{i_k}$ denote the binary strings representing the verifier's standard basis measurement outcomes (or claimed outcomes) corresponding to the measurements of the tuples (7).

C. Prover's check and response

Upon receiving the verifier's claimed measurement outcomes corresponding to the randomly selected Hamiltonian term, the prover first checks to see that these outcomes could indeed have come from the measurements specified above, and then tries to convince the verifier that these measurement outcomes are consistent with the selected term.

In more detail, suppose that the Hamiltonian term determined by r has been challenged. As above, we assume that this term acts nontrivially on the k qubits indexed by the k-tuple $(i_1, \ldots, i_k)$, and we will write

$$u = u_{i_1} \cdots u_{i_k} \in \{0, 1\}^{2kN} \tag{8}$$

to denote the verifier's claimed standard basis measurement outcomes.

To define the prover's check for this string, it will be helpful to first define a predicate R_r, which is a function

of t, π, and u, and essentially represents the prover's check *after* it has made an adjustment to the verifier's response to account for the one-time pad. For each $i \in \{i_1, \ldots, i_k\}$, define strings $y_i, z_i \in \{0,1\}^N$ so that $\pi(y_i z_i) = u_i$. The predicate R_r takes the value 1 if and only if these two conditions are met:

1. For each $i \in \{i_1, \ldots, i_k\}$, the string y_i is a valid code word, with respect to the Hamming codes upon which concatenated Steane codes are based, with at least one of these code words corresponding to a nonzero binary value.
2. No trap qubit measurement outcomes contradict their settings: $\langle z_{i_1} \cdots z_{i_k} \,|\, C_r^{\otimes N} \,|\, t_{i_1} \cdots t_{i_k} \rangle \neq 0$.

Here we have written $|t_{i_1} \cdots t_{i_k}\rangle$ to denote the pure state of kN qubits obtained by tensoring the states $|0\rangle$, $|+\rangle$, and $|\circlearrowleft\rangle$ in this most natural way.

Next, we will define a predicate Q_r, which is a function of the variables t, π, a, b, and u, where t, π, and u are as above and $a, b \in \{0,1\}^{2nN}$ refer to the strings used for the one-time pad. The predicate Q_r represents the prover's actual check, in the case that the Hamiltonian term determined by r has been selected, including an adjustment to account for the one-time pad. Let $c_1, \ldots, c_n, d_1, \ldots, d_n \in \{0,1\}^{2N}$ be the unique strings for which the equation

$$\begin{aligned} &C_r^{\otimes 2N}\big(X^{a_1} Z^{b_1} \otimes \cdots \otimes X^{a_n} Z^{b_n}\big) \\ &\quad = \alpha\big(X^{c_1} Z^{d_1} \otimes \cdots \otimes X^{c_n} Z^{d_n}\big) C_r^{\otimes 2N} \end{aligned} \tag{9}$$

holds for some choice of $\alpha \in \{1, i, -1, -i\}$. The Clifford operation C_r acts trivially on those qubits indexed by strings outside of the set $\{i_1, \ldots, i_k\}$, so it must be the case that $c_i = a_i$ and $d_i = b_i$ for $i \notin \{i_1, \ldots, i_k\}$, but for those indices $i \in \{i_1, \ldots, i_k\}$ it may be the case that $c_i \neq a_i$ and $d_i \neq b_i$. We will also write $c = c_1 \cdots c_n$ and $d = d_1 \cdots d_n$ for the sake of convenience. Given a description of the Clifford operation C_r it is possible to efficiently compute c and d from a and b. Having defined c and d, we may now express the predicate Q_r as follows:

$$Q_r(t, \pi, u, a, b) = R_r\big(t, \pi, u \oplus c_{i_1} \cdots c_{i_k}\big). \tag{10}$$

In essence, the predicate Q_r checks the validity of the verifier's claimed measurement results by first adjusting for the one-time pad, then referring to R_r.

The prover evaluates the predicate Q_r, and aborts the proof system if the predicate evaluates to 0 (as this is indicative of a dishonest verifier). Otherwise, the prover aims to convince the verifier that the measurement outcomes u are consistent with the prover's encoding, and also that they are not in violation of the Hamiltonian term H_r. It does this specifically by engaging in a classical zero-knowledge proof system for the following NP statement: there exists a random string s and an encoding key (t, π, a, b) such that (i) $\mathsf{commit}((\pi, a, b), s)$ matches the prover's initial commitment z, and (ii) $Q_r(t, \pi, u, a, b) = 1$.

V. Analysis of the proof system

As was stated previously, the analysis of the protocol described in the previous section is not included in this extended abstract, but can be found in the full version of the paper [17]. We will, however, discuss some aspects of the analysis in intuitive terms.

First, it is evident that the proof system described in the previous section is complete. For a given instance of the local Clifford Hamiltonian problem, if the prover and verifier both behave honestly, as suggested in the description of the proof system, the verifier will accept with precisely the same probability that would be obtained by randomly selecting a Hamiltonian term, measuring the original n-qubit witness state against the corresponding projection, and accepting or rejecting accordingly. For a positive problem instance, this acceptance probability is at least $1 - 2^{-p}$ (for every choice of a random string r).

The soundness of the proof system requires more work, and makes use of specific properties of concatenated Steane codes. One of the main sources of difficulty is that the prover is, of course, not required to follow the encoding scheme described in the protocol. Intuitively speaking, our soundness proof demonstrates that, for an arbitrary state sent by a malicious prover in its message, one can essentially decode from that state (with respect to a highly simplified variant of the encoding scheme, after peeling off the quantum one-time pad and discarding the trap qubits) a state that passes the Hamiltonian term test with at least the same probability as the verifier's acceptance probability in the proof system. Because this probability must be bounded away from 1 on average for any no-instance of the problem, we obtain a soundness guarantee for the proof system.

The most involved part of the analysis concerns the zero-knowledge property of the proof system. Figure 1 illustrates an arbitrary cheating verifier interacting with the honest prover. Observe that a cheating verifier may take a quantum register as input, store quantum information in between its actions, and output a quantum register. The goal of the proof is to demonstrate that, for any such cheating verifier, there exists an efficient simulator that implements a channel from Z_0 to Z_3 that is computationally indistinguishable from the channel implemented by the cheating verifier and prover interaction. In particular, the simulator does not have access to the witness state ρ. This is done through multiple steps.

Step 1: simulating the coin flipping protocol. By the results of [33], there must exist an efficient simulator S_1 for the interaction of V_1' with P_1. To be more precise, for S_1 being given an input of the same form as V_1', along with a uniformly chosen random string r of the length required by our proof system, the resulting action is quantum computationally indistinguishable from V_1' interacting with P_1.

Step 2: simulating the classical zero-knowledge protocol. The interaction between a cheating verifier V_3' and

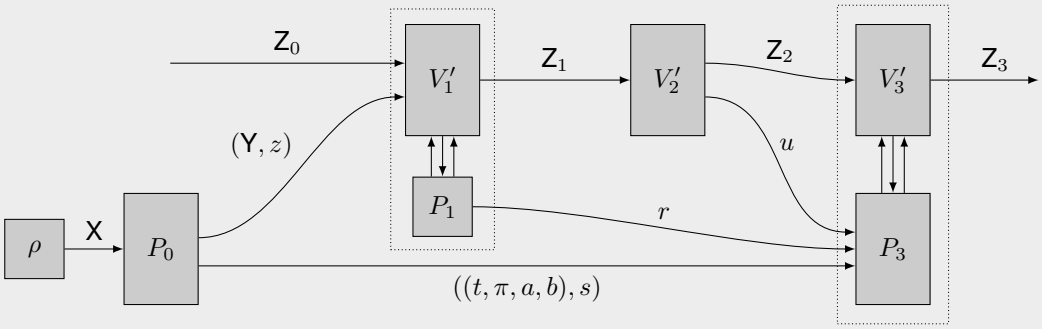

Figure 1. The interaction between the prover and an arbitrary verifier. The prover's quantum witness ρ is encoded into Y together with the encoding key (t, π, a, b) by the prover's action P_0. The string z represents the prover's commitment to (π, a, b) and the string s represents random bits used by the prover to implement this commitment. The string r represents the random bits generated by the coin flipping protocol, which is depicted within the dotted rectangle on the left. The string u represents the verifier's standard basis measurements for a subset of the qubits of Y determined by the challenge corresponding to the random string r. The classical zero-knowledge protocol is depicted within the dotted rectangle on the right. A potentially dishonest verifier takes an auxiliary quantum register Z_0 as input, may store quantum information (represented by registers Z_1 and Z_2), and outputs quantum information stored in register Z_3.

the prover P_3 in the classical zero-knowledge protocol is replaced by an efficient simulation.

Step 3: eliminating the commitment. Using the assumption that the commitment scheme is quantum computationally concealing, it may be argued that the entire process is computationally indistinguishable from one in which the commitment is made to a *fixed* choice of a tuple (π_0, a_0, b_0), independent of the prover's encoding key. One may then consider the commitment to this fixed tuple (π_0, a_0, b_0), together with the simulator S_1 and the cheating verifier action V_2', to form a single, efficiently implementable action V'. The interaction between this new action V' and the prover's encoding, the random string generator, and the predicate Q, is then considered.

Step 4: simulating an attack on the encoding scheme. The final step is to consider a simulation of an arbitrary efficiently implementable action V' of the form described in the previous step (where, of course, the simulator is not provided with a valid witness ρ). Such an action can, in fact, be efficiently simulated with statistical accuracy, not just in a computationally indistinguishable sense. This analysis makes use of specific properties of the encoding scheme and its connection to Clifford operations.

VI. Conclusion

We have described a zero-knowledge proof system for any problem in QMA, assuming the existence of a quantum computationally concealing and unconditionally binding commitment scheme. Such a commitment scheme can be obtained assuming quantum-secure one-way permutations [34] (or injections more generally) or a quantum-secure pseudo-random generator [31] that could potentially be based on one-way functions that are hard to invert for any quantum polynomial time algorithm [35]–[37]. We conclude with a few open questions and directions for future work.

1. Our proof system inherits the soundness error of the most straightforward verification procedure for the local Clifford-Hamiltonian problem, which is to randomly select a Hamiltonian term and perform a measurement corresponding to it. When an arbitrary QMA problem is reduced to the local Hamiltonian problem, the resulting soundness error may potentially be large (polynomially bounded away from 1). Can one obtain a zero-knowledge proof system for any QMA problem with small soundness error while remaining constant round (and maintaining the other features of our proof system)?

   We note that if a prover has polynomially many copies of a valid quantum witness, then one can essentially repeat our proof system in parallel to get a constant round zero-knowledge proof system having small soundness error for any QMA problem, assuming there is a constant-round zero-knowledge proofs for NP with negligibly small soundness error against *quantum* adversaries. However, existing constant-round zero-knowledge proofs for NP (e.g., [38] and [39]) involve sophisticated rewinding arguments, and it is unclear if they remain secure against quantum malicious verifiers.

2. Are there natural formalizations of *proofs of quantum knowledge*? Roughly speaking, one would expect such a notion to require that whenever a prover is able to prove the validity of a statement, one could construct a knowledge extractor that can extract a quantum witness given access to such a prover. It seems plausible that our proof system could be adapted to such a notion, although we have not investigated this notion in depth.

3. We have considered an encoding scheme for quantum states that ensures the secrecy of the state and allows for the transversal application of constant-size Clifford operations and measurement in the computational basis. It is an interesting open question to extend our encoding scheme, or to design a new one, so that it can support transversally applying a larger family of quantum operations.

4. Finally, we make one further remark on an abstract view of our proof system. Classically speaking, one can imagine a "commit-and-open" primitive where a sender commits to a message m, and later opens sufficient information so that a receiver can test a property $\mathcal{P}(\cdot)$ on m, and nothing more. For example, $\mathcal{P}$ can be an NP-relation $R(x, \cdot)$ that checks if message m is a valid witness. This can be implemented easily by a standard commitment scheme and during the opening phase, the sender and receiver run a zero-knowledge proof of $R(x, m) = 1$ instead of the standard opening. Our proof system, which combines a commitment scheme and classical zero-knowledge proofs for NP, can be viewed as a quantum analogue. Namely, we commit to a witness state and open just enough information to verify that some reduced density of the witness state falls into a specific subspace. We can only deal with properties of a very special form, and it is an interesting direction for future work to generalize and find applications of this sort of primitive.

Acknowledgments

We thank Michael Beverland, Sevag Gharibian, David Gosset, Yi-Kai Liu, and Bei Zeng for helpful conversations. A. B. and J. W. are supported in part by Canada's NSERC. F. S. did this work while at the University of Waterloo, and he was supported in part by CryptoWorks21, Canada's NSERC and CIFAR.

REFERENCES

[1] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," *SIAM Journal on Computing*, vol. 18, no. 1, pp. 186–208, 1989.

[2] O. Goldreich, S. Micali, and A. Wigderson, "Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems," *Journal of the ACM*, vol. 38, no. 3, pp. 690–728, 1991.

[3] M. Ben-Or, O. Goldreich, S. Goldwasser, J. Håstad, J. Kilian, S. Micali, and P. Rogaway, "Everything provable is provable in zero-knowledge," in *Advances in Cryptology – CRYPTO 1988*, ser. Lecture Notes in Computer Science, vol. 403. Springer-Verlag, 1990, pp. 37–56.

[4] O. Goldreich, S. Micali, and A. Wigderson, "How to play ANY mental game," in *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, 1987, pp. 218–229.

[5] P. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1484–1509, 1997.

[6] W. Wootters and W. Zurek, "A single quantum cannot be cloned," *Nature*, vol. 299, pp. 802–803, 1982.

[7] C. Fuchs and A. Peres, "Quantum-state disturbance versus information gain: Uncertainty relations for quantum information," *Physical Review A*, vol. 53, no. 4, p. 2038, 1996.

[8] J. van de Graaf, "Towards a formal definition of security for quantum protocols," Ph.D. dissertation, Université de Montréal, 1997.

[9] J. Watrous, "Limits on the power of quantum statistical zero-knowledge," in *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002, pp. 459–468.

[10] I. Damgård, S. Fehr, and L. Salvail, "Zero-knowledge proofs and string commitments withstanding quantum attacks," in *Advances in Cryptology – CRYPTO 2004*, ser. Lecture Notes in Computer Science, vol. 3152. Springer, 2004, pp. 254–272.

[11] J. Watrous, "Zero-knowledge against quantum attacks," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 25–58, 2009.

[12] A. Kitaev, A. Shen, and M. Vyalyi, *Classical and Quantum Computation*, ser. Graduate Studies in Mathematics. American Mathematical Society, 2002, vol. 47.

[13] A. Broadbent, G. Gutoski, and D. Stebila, "Quantum one-time programs," in *Advances in Cryptology – CRYPTO 2013*, ser. Lecture Notes in Computer Science, vol. 8043. Springer, 2013, pp. 344–360.

[14] H. Barnum, C. Crépeau, D. Gottesman, A. Smith, and A. Tapp, "Authentication of quantum messages," in *Proceedings of the 43th Annual IEEE Symposium on Foundations of Computer Science*, 2002, pp. 449–458.

[15] M. Ben-Or, C. Crépeau, D. Gottesman, A. Hassidim, and A. Smith, "Secure multiparty quantum computation with (only) a strict honest majority," in *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006, pp. 249–260.

[16] D. Aharonov, M. Ben-Or, and E. Eban, "Interactive proofs for quantum computations," in *Innovations in Computer Science*, 2010, pp. 453–469.

[17] A. Broadbent, Z. Ji, F. Song, and J. Watrous, "Zero-knowledge proof systems for QMA," Available as arXiv:1604.02804, 2016.

[18] A. Ambainis, M. Mosca, A. Tapp, and R. de Wolf, "Private quantum channels," in *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, 2000, pp. 547–553.

[19] F. Dupuis, J. Nielsen, and L. Salvail, "Actively secure two-party evaluation of any quantum operation," in *Advances in Cryptology – CRYPTO 2012*, ser. Lecture Notes in Computer Science, vol. 7417. Springer, 2012, pp. 794–811.

[20] S. Hallgren, A. Kolla, P. Sen, and S. Zhang, "Making classical honest verifier zero knowledge protocols secure against quantum attacks," in *Proceedings of the 35th International Colloquium on Automata, Languages and Programming, Part II*, ser. Lecture Notes in Computer Science, vol. 5126. Springer-Verlag, 2008, pp. 592–603.

[21] S. Goldwasser and M. Sipser, "Private coins versus public coins in interactive proof systems," in *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, 1986, pp. 59–68.

[22] J. Watrous, "PSPACE has constant-round quantum interactive proof systems," *Theoretical Computer Science*, vol. 292, no. 3, pp. 575–588, 2003.

[23] O. Goldreich and Y. Oren, "Definitions and properties of zero-knowledge proof systems," *Journal of Cryptology*, vol. 7, no. 1, pp. 1–32, 1994.

[24] F. Dupuis, J. Nielsen, and L. Salvail, "Secure two-party quantum evaluation of unitaries against specious adversaries," in *Advances in Cryptology – CRYPTO 2010*, ser. Lecture Notes in Computer Science, vol. 6223. Springer, 2010, pp. 685–706.

[25] S. Hallgren, A. Smith, and F. Song, "Classical cryptographic protocols in a quantum world," *International Journal of Quantum Information*, vol. 13, no. 04, p. 1550028, 2015.

[26] C. Lunemann and J. Nielsen, "Fully simulatable quantum-secure coin-flipping and applications," in *Progress in Cryptology – AFRICACRYPT 2011*, ser. Lecture Notes in Computer Science, vol. 6737. Springer-Verlag, 2011, pp. 21–40.

[27] R. Impagliazzo, "A personal view of average-case complexity," in *Proceedings of 10th Annual IEEE Structure in Complexity Theory Conference*, 1995, pp. 134–147.

[28] Y.-K. Liu, "Consistency of local density matrices is QMA-complete," in *Proceedings of the 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2006 and 10th International Workshop on Randomization and Computation, RANDOM 2006*, ser. Lecture Notes in Computer Science. Springer-Verlag, 2006, vol. 4110, pp. 438–449.

[29] T. Morimae, M. Hayashi, H. Nishimura, and K. Fujii, "Quantum Merlin-Arthur with Clifford Arthur," *Quantum Information and Computation*, vol. 15, pp. 1420–1430, 2015.

[30] T. Morimae, D. Nagaj, and N. Schuch, "Quantum proofs can be verified using only single-qubit measurements," *Physical Review A*, vol. 93, no. 2, p. 022326, 2016.

[31] M. Naor, "Bit commitment using pseudorandomness," *Journal of Cryptology*, vol. 4, no. 2, pp. 151–158, 1991.

[32] M. Blum, "Coin flipping by telephone a protocol for solving impossible problems," *ACM SIGACT News*, vol. 15, no. 1, pp. 23–27, 1983.

[33] I. Damgård and C. Lunemann, "Quantum-secure coin-flipping and applications," in *Advances in Cryptology – ASIACRYPT 2009*, ser. Lecture Notes in Computer Science, vol. 5912. Springer, 2009, pp. 52–69.

[34] M. Adcock and R. Cleve, "A quantum Goldreich-Levin theorem with cryptographic applications," in *Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science*, ser. Lecture Notes in Computer Science. Springer-Verlag, 2002, vol. 2285, pp. 323–334.

[35] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby, "A pseudorandom generator from any one-way function," *SIAM Journal on Computing*, vol. 28, no. 4, pp. 1364–1396, 1999.

[36] M. Zhandry, "How to construct quantum random functions," in *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, 2012, pp. 679–687.

[37] F. Song, "A note on quantum security for post-quantum cryptography," in *Proceedings of the 6th International Workshop on Post-Quantum Cryptography*, ser. Lecture Notes in Computer Science. Springer, 2014, vol. 8772, pp. 246–265.

[38] O. Goldreich and A. Kahan, "How to construct constant-round zero-knowledge proof systems for NP," *Journal of Cryptology*, vol. 9, no. 3, pp. 167–189, 1996.

[39] U. Feige and A. Shamir, "Zero knowledge proofs of knowledge in two rounds," in *Advances in Cryptology – CRYPTO 1989*, ser. Lecture Notes in Computer Science, vol. 435. Springer-Verlag, 1990, pp. 526–544.