Quantum Arthur-Merlin games

Chris Marriott John Watrous

Department of Computer Science
University of Calgary
2500 University Drive NW
Calgary, Alberta, Canada T2N 1N4

{marriott, jwatrous}@cpsc.ucalgary.ca

Abstract

This paper studies quantum Arthur-Merlin games, which are a restricted form of quantum interactive proof system in which the verifier's messages are given by unbiased coin-flips. The following results are proved.

- *For one-message quantum Arthur-Merlin games, which correspond to the complexity class* QMA*, completeness and soundness errors can be reduced exponentially without increasing the length of Merlin's message. Previous constructions for reducing error required a polynomial increase in the length of Merlin's message. Applications of this fact include a proof that logarithmic length quantum certificates yield no increase in power over* BQP *and a simple proof that* $\mathrm{QMA} \subseteq \mathrm{PP}$*.*
- *In the case of three or more messages, quantum Arthur-Merlin games are equivalent in power to ordinary quantum interactive proof systems. In fact, for any language having a quantum interactive proof system there exists a three-message quantum Arthur-Merlin game in which Arthur's only message consists of just a single coin-flip that achieves perfect completeness and soundness error exponentially close to 1/2.*
- *Any language having a two-message quantum Arthur-Merlin game is contained in* $\mathrm{BP} \cdot \mathrm{PP}$*. This gives some suggestion that three messages are stronger than two in the quantum Arthur-Merlin setting.*

1. Introduction

This paper investigates the complexity-theoretic aspects of quantum Arthur-Merlin games, which are defined in an analogous way to classical Arthur-Merlin games [3, 4]. Specifically, quantum Arthur-Merlin games are quantum interactive proof systems [13, 23] in which Arthur (the verifier) only sends uniformly generated sequences of bits to Merlin (the prover) instead of arbitrary quantum information. Thus, Arthur may not send any quantum information at all to Merlin, and one may view all of Arthur's computations (quantum or classical) as taking place after all messages have been exchanged. Similar to the classical case, quantum Arthur-Merlin games give rise to complexity classes depending on the number of messages exchanged between Arthur and Merlin. In particular, we obtain three primary complexity classes corresponding to Arthur-Merlin games with one message, two messages, and three or more messages.

In the one-message case, Merlin sends a single message to Arthur, who checks it and makes a decision to accept or reject the input. The corresponding complexity class is denoted QMA, and has recently been considered in several papers [2, 8, 9, 15, 19, 21, 22]. Because there is really no interaction between Arthur and Merlin in this situation, Merlin's message to Arthur may be viewed as a quantum witness or certificate that Arthur checks in polynomial time with a quantum computer. To our knowledge, the idea of a quantum state playing the role of a certificate in this sense was first proposed by Knill [14], and the idea was later studied in greater depth by Kitaev [11]. Kitaev proved various fundamental properties of QMA, which are described in Kitaev, Shen, and Vyalyi [12] and Aharovov and Naveh [1].

One of the facts that Kitaev proved was that the completeness and soundness errors in a QMA protocol may be efficiently reduced, essentially by parallel repetition. Because quantum information cannot be copied, however, and Arthur's verification procedure is potentially destructive to Merlin's message, Arthur requires multiple copies of Merlin's message for this method to work. Consequently, this method requires a polynomial increase in the length of Merlin's message to Arthur in order to achieve exponentially decreasing error. In this paper, we prove that this increase in the length of Merlin's message is not required after all; us-

ing a different error reduction method, an exponential reduction in error is possible with no increase whatsoever in the length of Merlin's message to Arthur.

It is known that QMA is contained in PP, which can be proved using the GapP-based method of Fortnow and Rogers [6] together with some simple facts from matrix analysis. This fact was noted without proof in Ref. [13]. A proof does appear, however, in a recent paper of Vyalyi [21], who in fact strengthens this result to show that QMA is contained in a subclass $\mathrm{A_0PP}$ of PP. Based on our new amplification result, we give a simplified proof of this fact. We also use our amplification result to prove that quantum Merlin-Arthur games in which Merlin's message has logarithmic length give no increase in power over BQP.

In the two-message case, Arthur flips some number of fair coins, sends the results of those coin-flips to Merlin, and Merlin responds with some quantum state. Arthur performs a polynomial-time quantum computation on the random bits together with Merlin's response, which determines whether Arthur accepts or rejects. The corresponding complexity class will be denoted QAM. Two facts about QAM are proved in this paper. The first is the very basic fact that parallel repetition reduces error exactly as in the classical case. (Parallel repetition is currently known to hold for general quantum interactive proof systems only in the case of perfect completeness.) The second fact is that $\mathrm{QAM} \subseteq \mathrm{BP} \cdot \mathrm{PP}$. This may be viewed as weak evidence that two-message quantum Arthur-Merlin games are not as powerful as PSPACE.

Finally, in the three-message case, Merlin sends Arthur a message consisting of some number of qubits, Arthur flips some number of fair coins and sends the results to Merlin, and then Merlin responds with a second collection of qubits. Arthur performs a polynomial-time quantum computation on all of the qubits sent by Merlin together with the values of his own coin-flips, and decides whether to accept or reject. The corresponding complexity class will be denoted QMAM. It is proved that any language having an ordinary quantum interactive proof system is contained in QMAM, implying $\mathrm{QMAM} = \mathrm{QIP}$. In principle this fact resembles the theorem of Goldwasser and Sipser [7] establishing that classical Arthur-Merlin games and interactive proof systems are equivalent in power. However, there is no similarity in the proofs of these facts. Indeed, our result is stronger than what is likely to hold classically. Specifically, we prove that any language having a quantum interactive proof system has a three-message quantum Arthur-Merlin game in which Arthur's only message to Merlin consists of just a single coin-flip (in order to achieve perfect completeness and soundness error exponentially close to 1/2). This is impossible classically unless interaction is useless in classical interactive proof systems; for if Arthur flips only one coin, Merlin may as well send his first message and the two possible second messages to Arthur in a single message. The reason why this strategy fails in the quantum case is that Merlin's first and second messages may need to be entangled in order to be convincing to Arthur, but it is not possible for Merlin to simultaneously entangle his two possible second messages with the first.

The remainder of this paper is organized in the following way. First, Section 2 discusses some notation and background information used in the paper. Section 3 discusses one-message quantum Arthur-Merlin games, Section 4 discusses the two-message case, and Section 5 discusses the case of three or more messages. These sections therefore correspond to the three complexity classes QMA, QAM, and QMAM, respectively. The paper concludes with Section 6, which mentions some open problems relating to quantum Arthur-Merlin games.

2. Preliminaries

By default, all strings and languages in this paper will be over the alphabet $\Sigma = \{0,1\}$. We denote by $poly$ the set of all functions $f : \mathbb{N} \rightarrow \mathbb{N}\backslash\{0\}$ (where $\mathbb{N} = \{0,1,2,\ldots\}$) for which there exists a polynomial-time deterministic Turing machine that outputs $1^{f(n)}$ on input 1^n.

We will assume that the reader has some familiarity with the mathematics of quantum information, which is discussed in detail in Kitaev, Shen, and Vyalyi [12] and Nielsen and Chuang [17]. It will be possible for us to restrict our attention to pure quantum states for much of the paper, although mixed quantum states will be used occasionally (in particular, in Sections 3.2, 5.1, and 5.2).

For simplicity we will define quantum Arthur-Merlin games in terms of quantum circuits composed of gates from the Shor basis: Toffoli gates, Hadamard gates, and i-shift gates (which induce the mapping $|0\rangle \mapsto |0\rangle$, $|1\rangle \mapsto i|1\rangle$). This is a universal set of gates (see Ref. [10]), so there is no loss of generality in restricting our attention to this set. Assume that a reasonable encoding scheme has been fixed that allows quantum circuits to be encoded as binary strings having length at least the size of the encoded circuit and at most some fixed polynomial in the circuit's size.

A collection $\{A_x : x \in \Sigma^*\}$ of quantum circuits is said to be *polynomial-time uniform* if there exists a polynomial-time deterministic Turing machine M that, on input $x \in \Sigma^*$, outputs an encoding of the circuit A_x. Although this is not the most conventional notion of circuit uniformity, as the family is parameterized by strings rather than string lengths, it is better suited to our needs. More generally, such a family may be parameterized by tuples of strings; for instance we will consider families of the form

$$\left\{A_{x,y} \,:\, x \in \Sigma^*,\, y \in \Sigma^{r(|x|)}\right\}$$

for r some function in $poly$ when two- and three-message quantum Arthur-Merlin games are discussed.

It will sometimes be helpful when describing certain quantum Arthur-Merlin games to refer to *quantum registers*. These are simply collections of qubits to which we assign some name. The qubits of a given register may be entangled with other qubits, so it may not be possible to describe the state of a quantum register by a single vector. When we refer to the *reduced state* of a given register, we are referring to the mixed state obtained by tracing out all other registers beside the one to which we are referring.

The definitions for quantum Arthur-Merlin games and the resulting complexity classes QMA, QAM, and QMAM will appear in their respective sections.

3. QMA

A QMA verification procedure A is a polynomial-time uniform family $\{A_x : x \in \Sigma^*\}$ of quantum circuits together with a function $m \in poly$. The function m specifies the length of Merlin's message to Arthur, and it is assumed that each circuit A_x acts on $m(|x|) + k(|x|)$ qubits for some function k specifying the number of work (or *ancilla*) qubits used by the circuit. In order to simplify our notation, when the input x has been fixed or is implicit we will generally write m to mean $m(|x|)$, k to mean $k(|x|)$, and so forth. When we want to emphasize the length of Merlin's message, we will refer to A as an m-qubit QMA verification procedure.

Consider the following process for a string $x \in \Sigma^*$ and a quantum state $|\psi\rangle$ on m qubits:

1. Run the circuit A_x on the input state $|\psi\rangle|0^k\rangle$.
2. Measure the first qubit of the resulting state in the standard basis, interpreting the outcome 1 as *accept* and the outcome 0 as *reject*.

The probability associated with each of the two possible outcomes will be referred to as $\Pr[A_x \text{ accepts } |\psi\rangle]$ and $\Pr[A_x \text{ rejects } |\psi\rangle]$ accordingly.

Definition 3.1. The class $\mathrm{QMA}(a, b)$ consists of all languages $L \subseteq \Sigma^*$ for which there exists a QMA verification procedure $\{A_x : x \in \Sigma^*\}$ for which the following holds:

1. For all $x \in L$ there exists an m qubit quantum state $|\psi\rangle$ such that
$$\Pr[A_x \text{ accepts } |\psi\rangle] \geq a.$$
2. For all $x \notin L$ and all m qubit quantum states $|\psi\rangle$,
$$\Pr[A_x \text{ accepts } |\psi\rangle] \leq b.$$

For any $m \in poly$, the class $\mathrm{QMA}_m(a, b)$ consists of all languages $L \subseteq \Sigma^*$ for which there exists an m-qubit QMA verification procedure that satisfies the above properties.

One may consider the cases where a and b are constants or functions of the input length $n = |x|$ in this definition. If a and b are functions of the input length, it is assumed that $a(n)$ and $b(n)$ can be computed deterministically in time polynomial in n. When no reference is made to the probabilities a and b, it is assumed $a = 2/3$ and $b = 1/3$.

3.1. Amplification

It is known that QMA is robust with respect to error bounds in the following sense.

Theorem 3.2 (Kitaev). *Let $a, b : \mathbb{N} \to (0, 1)$ and $p \in poly$ satisfy*
$$a(n) - b(n) \geq \frac{1}{p(n)}$$
for all $n \in \mathbb{N}$. Then
$$\mathrm{QMA}(a, b) = \mathrm{QMA}(1 - 2^{-q}, 2^{-q})$$
for every $q \in poly$.

A proof of this theorem appears in Section 14.2 of Kitaev, Shen, and Vyalyi [12]. The idea of the proof is as follows. If we have a verification procedure A with completeness and soundness probabilities given by a and b, we construct a new verification procedure that independently runs A on some sufficiently large number of copies of the original certificate and accepts if the number of acceptances of A is appropriately large (above $(a + b)/2$, say). The difficulty in proving that this construction works lies in the fact that the new certificate cannot be assumed to consist of several copies of the original certificate, but may be an arbitrary (possibly highly entangled) quantum state. Intuitively, however, entanglement cannot help Merlin to cheat; under the assumption that $x \notin L$, the probability of acceptance for any particular execution of A is bounded above by b, and this is true regardless of whether one conditions on the outcomes of any of the other executions of A. This construction requires an increase in the length of Merlin's message to Arthur in order to reduce error.

The main result of this section is the following theorem, which states that one may decrease error without any increase in the length of Merlin's message.

Theorem 3.3. *Let $a, b : \mathbb{N} \to (0, 1)$ and $p \in poly$ satisfy*
$$a(n) - b(n) \geq \frac{1}{p(n)}$$
for all $n \in \mathbb{N}$. Then
$$\mathrm{QMA}_m(a, b) = \mathrm{QMA}_m(1 - 2^{-q}, 2^{-q})$$
for every $q, m \in poly$.

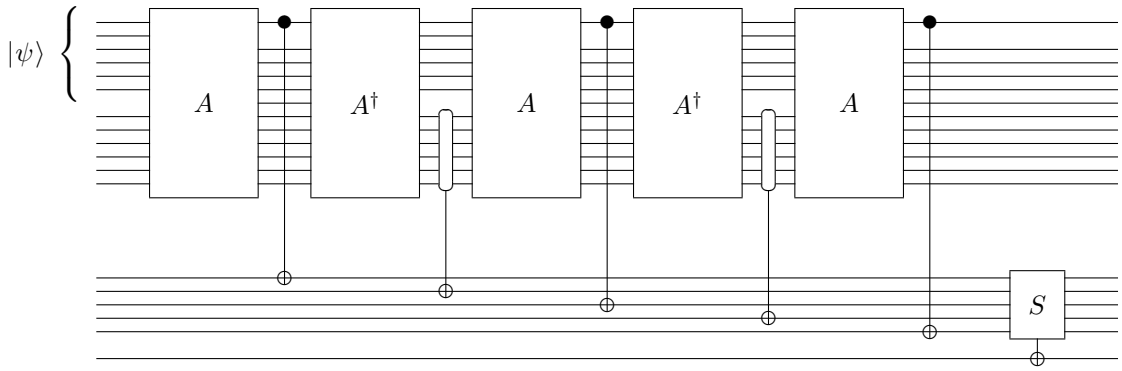

Figure 1. Example circuit diagram for verification procedure B.

Proof. Assume $L \in \mathrm{QMA}_m(a,b)$, and A is an m-qubit QMA verification procedure that witnesses this fact. We will describe a new m-qubit QMA verification procedure B with exponentially small completeness and soundness error. It will simplify matters to assume hereafter that the input x is fixed—it will be clear that the new verification procedure is polynomial-time uniform. As the input x is fixed, we will write m, k, etc., rather than $m(|x|)$, $k(|x|)$, etc., and we will write A and B to denote A_x and B_x, respectively.

It will be helpful to refer to the m message qubits along with the k work-space qubits of A as a single $m+k$ qubit quantum register R. Define projections acting on the possible states of the register R as follows:

$$\begin{array}{l} \Pi_1 = |1\rangle\langle 1| \otimes I_{m+k-1}, \;\; \Delta_1 = I_m \otimes |0^k\rangle\langle 0^k|, \\ \Pi_0 = |0\rangle\langle 0| \otimes I_{m+k-1}, \;\; \Delta_0 = I_{m+k} - \Delta_1. \end{array} \tag{1}$$

Here, and throughout the paper, we write I_l to denote the identity operator acting on l qubits. The measurement described by $\{\Pi_0, \Pi_1\}$ is just a measurement of the first qubit of R in the computational basis; this measurement determines whether Arthur accepts or rejects after the circuit A is applied. The measurement described by $\{\Delta_0, \Delta_1\}$ gives outcome 1 if the last k qubits of R, which correspond to Arthur's work-space qubits, are set to their initial all-zero state, and gives outcome 0 otherwise.

The procedure B operates as follows:

1. Assume the first m qubits of R contain Merlin's message $|\psi\rangle$ and the remaining k qubits are set to the state $|0^k\rangle$.
2. Set $r_0 \leftarrow 1$ and $i \leftarrow 1$.
3. Repeat:
   a. Apply A to R and measure R with respect to the measurement described by $\{\Pi_0, \Pi_1\}$. Let r_i denote the outcome, and set $i \leftarrow i+1$.
   b. Apply $A^\dagger$ to R and measure R with respect to the measurement described by $\{\Delta_0, \Delta_1\}$. Let r_i denote the outcome, and set $i \leftarrow i+1$.

   Until $i \geq N$ (where N is chosen depending on the desired error bound).
4. For each $i = 1, \ldots, N$ set

$$s_i \leftarrow \begin{cases} 1 & \text{if } r_i = r_{i-1} \\ 0 & \text{if } r_i \neq r_{i-1}. \end{cases}$$

   Accept if $\sum_{i=1}^N s_i \geq N \cdot \frac{a+b}{2}$ and reject otherwise.

Although the description of this procedure refers to various measurements, it is possible to simulate these measurements with unitary gates in the standard way, which allows the entire procedure to be implemented by a unitary quantum circuit. Figure 1 illustrates a quantum circuit implementing this procedure for the case $N = 5$. In this figure, S represents the computation described in the last step of B (performed reversibly), and the last qubit rather than the first represents the output qubit to simplify the picture.

Suppose first that Merlin's message $|\psi\rangle$ is an eigenvector of $(I_m \otimes \langle 0^k|)A^\dagger \Pi_1 A (I_m \otimes |0^k\rangle)$, which is equivalent to $|\phi\rangle = |\psi\rangle|0^k\rangle$ being an eigenvector of $\Delta_1 A^\dagger \Pi_1 A \Delta_1$. Let the corresponding eigenvalue be p, which implies the verification procedure A would accept $|\psi\rangle$ with probability p. It will be shown that the verification procedure B would accept $|\psi\rangle$ with probability

$$\sum_{N \cdot \frac{a+b}{2} \leq j \leq N} \binom{N}{j} p^j (1-p)^{N-j}.$$

This will follow from the fact that the procedure B obtains each possible sequence $(s_1, \ldots, s_N)$ with probability $p^{w(s)}(1-p)^{N-w(s)}$ for $w(s) = \sum_{i=1}^N s_i$. This is straightforward if $p = 0$ or $p = 1$, so assume $0 < p < 1$.

Define vectors $|\gamma_0\rangle$, $|\gamma_1\rangle$, $|\delta_0\rangle$, and $|\delta_1\rangle$ as follows:

$$|\gamma_0\rangle = \frac{\Pi_0 A \Delta_1 |\phi\rangle}{\sqrt{1-p}}, \qquad |\delta_0\rangle = \frac{\Delta_0 A^\dagger \Pi_1 |\gamma_1\rangle}{\sqrt{1-p}},$$

$$|\gamma_1\rangle = \frac{\Pi_1 A \Delta_1 |\phi\rangle}{\sqrt{p}}, \qquad |\delta_1\rangle = \frac{\Delta_1 A^\dagger \Pi_1 |\gamma_1\rangle}{\sqrt{p}}.$$

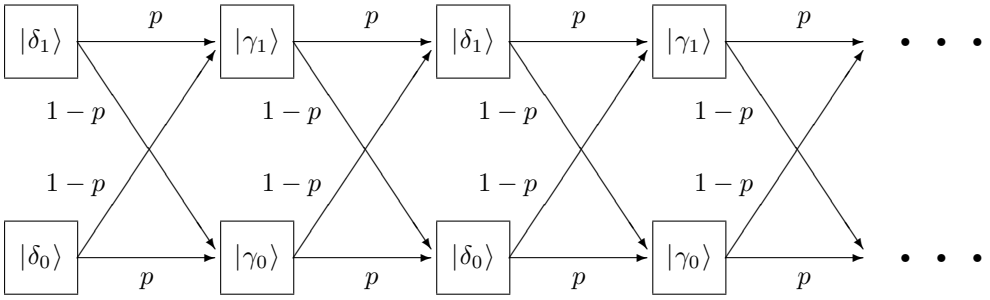

Figure 2. Transition probabilities for verification procedure M.

The fact that $|\phi\rangle$ is an eigenvector of $\Delta_1 A^\dagger \Pi_1 A \Delta_1$ with eigenvalue p implies that each of the vectors $|\gamma_0\rangle$, $|\gamma_1\rangle$, $|\delta_0\rangle$, $|\delta_1\rangle$ is a unit vector, and that $|\delta_1\rangle = |\phi\rangle$. We have

$$\begin{aligned} A|\delta_0\rangle &= -\sqrt{p}\,|\gamma_0\rangle + \sqrt{1-p}\,|\gamma_1\rangle \\ A|\delta_1\rangle &= \sqrt{1-p}\,|\gamma_0\rangle + \sqrt{p}\,|\gamma_1\rangle. \end{aligned} \tag{2}$$

The second equality follows from $|\delta_1\rangle = |\phi\rangle$, and the first follows from the second along with the observation that

$$A\left(\sqrt{1-p}|\delta_0\rangle + \sqrt{p}|\delta_1\rangle\right) = |\gamma_1\rangle.$$

With the above equations (2) in hand, it is not difficult to determine the probability associated with each sequence of measurement outcomes. We begin in state $|\phi\rangle = |\delta_1\rangle$ and apply A. After the measurement described by $\{\Pi_0, \Pi_1\}$ the (renormalized) state of register R becomes $|\gamma_0\rangle$ or $|\gamma_1\rangle$ according to whether the outcome is 0 or 1, with associated probabilities $1-p$ and p, respectively. If instead we were to start in state $|\delta_0\rangle$, the renormalized states after measurement would be the same, but the probabilities are reversed; we have probability p associated with outcome 0 and probability $1-p$ with outcome 1. For the second step of the loop the situation is similar. If the register R is in state $|\gamma_1\rangle$, the transformation $A^\dagger$ is applied, and the state is measured via the measurement $\{\Delta_0, \Delta_1\}$, the renormalized state after measurement will be either $|\delta_1\rangle$ or $|\delta_0\rangle$, with associated probabilities p and $1-p$. If instead the initial state is $|\gamma_0\rangle$ rather than $|\gamma_1\rangle$, the renormalized states after the measurement are again the same, but the probabilities are reversed. These transition probabilities are illustrated in Figure 2. In all cases we see that the probability of obtaining the same outcome as for the previous measurement is p, and the probability of the opposite outcome is $1-p$. The probability associated with a given sequence $s = (s_1, \ldots, s_N)$ is therefore $p^{w(s)}(1-p)^{N-w(s)}$ as claimed, as each s_i is 1 if the measurement outcomes r_{i-1} and r_i are equal, and is 0 otherwise. (Setting $r_0 = 1$ includes the first measurement outcome in this pattern.)

In general Merlin might not provide Arthur with an eigenvector of $(I_m \otimes \langle 0^k|)A^\dagger \Pi_1 A (I_m \otimes |0^k\rangle)$, but the above analysis makes it straightforward to determine the maximum probability with which Merlin can cause the procedure B to accept. Assume B uses l work-space qubits in addition to the k work-space qubits used by A. For any sequence $r = (r_1, \ldots, r_N)$ of measurement outcomes let $f(r)$ be 1 or 0 depending on whether the sequence would be accepted or rejected by B, respectively. Also define Λ_r to be a projection operator that projects onto states for which the work-space qubits of B record measurement outcomes $r = (r_1, \ldots, r_N)$. Then for

$$Q = \sum_{r : f(r)=1} (I_m \otimes \langle 0^{k+l}|)B^\dagger \Lambda_r B (I_m \otimes |0^{k+l}\rangle)$$

we have that the probability that B accepts message $|\psi\rangle$ is $\langle \psi | Q | \psi \rangle$. The largest probability with which Merlin can cause the procedure B to accept is given by the largest eigenvalue of Q.

Let $\{|\psi_1\rangle, \ldots, |\psi_{2^m}\rangle\}$ be a complete orthonormal collection of eigenvectors of

$$(I_m \otimes \langle 0^k|)A^\dagger \Pi_1 A (I_m \otimes |0^k\rangle),$$

with associated eigenvalues $p_1, \ldots, p_{2^m}$. By the above analysis we may conclude that for each r this set is also a set of eigenvectors of

$$(I_m \otimes \langle 0^{k+l}|)B^\dagger \Lambda_r B (I_m \otimes |0^{k+l}\rangle),$$

and therefore is a set of eigenvectors of the sum Q. The associated eigenvalues for Q are therefore given by

$$\sum_{N \cdot \frac{a+b}{2} \leq j \leq N} \binom{N}{j} p_i^j (1-p_i)^{N-j}.$$

Because each p_i is bounded above by the maximum acceptance probability for the procedure A, we have at this point that the theorem follows by a suitable choice for N along with standard Chernoff-type bounds. ∎

3.2. Applications

Two applications of Theorem 3.3 will be given in this section. The first is a simplified proof that QMA is contained in PP.

Theorem 3.4. $\mathrm{QMA} \subseteq \mathrm{PP}$.

Proof. Let $L \subseteq \Sigma^*$ be a language in QMA. By Theorem 3.3 there exists a function $m \in \mathit{poly}$ such that

$$L \in \mathrm{QMA}_m \left(1 - 2^{-(m+2)}, 2^{-(m+2)}\right).$$

Let A be a verification procedure that witnesses this fact. Specifically, each circuit A_x acts on $k+m$ qubits, for some $k \in \mathit{poly}$, and satisfies the following. If $x \in L$, then there exists an m qubit state $|\psi\rangle$ such that

$$\Pr[A_x \text{ accepts } |\psi\rangle] \geq 1 - 2^{-m-2},$$

while if $x \notin L$, then

$$\Pr[A_x \text{ accepts } |\psi\rangle] \leq 2^{-m-2}$$

for every m qubit state $|\psi\rangle$.

For each $x \in \Sigma^*$, define a $2^m \times 2^m$ matrix Q_x as

$$Q_x = \left(I_m \otimes \langle 0^k|\right) A_x^\dagger \Pi_1 A_x \left(I_m \otimes |0^k\rangle\right).$$

Each Q_x is positive semidefinite, and

$$\langle\psi|Q_x|\psi\rangle = \Pr[A_x \text{ accepts } |\psi\rangle]$$

for any unit vector $|\psi\rangle$ on m qubits. The maximum probability with which A_x can be made to accept is therefore the largest eigenvalue of Q_x. Because the trace of a matrix is equal to the sum of its eigenvalues and all eigenvalues of Q_x are nonnegative, it follows that if $x \in L$, then

$$\mathrm{tr}(Q_x) \geq 1 - 2^{-m-2} \geq 3/4,$$

while if $x \notin L$, then

$$\mathrm{tr}(Q_x) \leq 2^m 2^{-m-2} \leq 1/4.$$

Now, based on a straightforward modification of Fortnow and Rogers [6] based on our choice of the Shor basis, we have that there exist a polynomially-bounded FP function t and GapP functions f and g such that

$$\Re(Q_x[i,j]) = \frac{f(x,i,j)}{2^{t(x)}} \quad \text{and} \quad \Im(Q_x[i,j]) = \frac{g(x,i,j)}{2^{t(x)}}$$

for $0 \leq i,j < 2^m$. Define

$$h(x) = \sum_{i=0}^{2^m-1} f(x,i,i).$$

By closure properties of GapP functions, $h \in \mathrm{GapP}$. Moreover, we have $h(x) = 2^{t(x)}\,\mathrm{tr}(Q_x)$, and therefore

$$x \in L \Rightarrow\ h(x) \geq \frac{3}{4} 2^{t(x)}$$

$$x \notin L \Rightarrow\ h(x) \leq \frac{1}{4} 2^{t(x)}.$$

Because $2^{t(x)}$ is an FP function, it follows that $2h(x) - 2^{t(x)}$ is a GapP function that is positive if $x \in L$ and negative if $x \notin L$. Thus, $L \in \mathrm{PP}$ as required. ∎

Remark. A simple modification of the above proof yields $\mathrm{QMA} \subseteq \mathrm{A_0PP}$, where $\mathrm{A_0PP}$ is defined in Vyalyi [21].

The second application concerns quantum Merlin-Arthur games where Merlin sends only a logarithmic number of qubits to Arthur.

Classical Merlin-Arthur games with logarithmic-length messages from Merlin to Arthur are obviously equivalent in power to BPP, because Arthur could simply search through all possible messages in polynomial time in lieu of interacting with Merlin. In the quantum case, however, this argument does not work, as one may construct exponentially large sets of pairwise nearly-orthogonal quantum states on a logarithmic number of qubits, such as those used in quantum fingerprinting [5]. Nevertheless, logarithmic length quantum messages can be shown to be useless in the context of QMA using a different method, based on the strong amplification property of QMA proved above.

For $a, b : \mathbb{N} \to [0,1]$ define $\mathrm{QMA}_{\log}(a,b)$ to be the class of all languages contained in $\mathrm{QMA}_m(a,b)$ for some $m(n) \in O(\log n)$, and let $\mathrm{QMA}_{\log} = \mathrm{QMA}_{\log}(2/3, 1/3)$. The choice of the constants 2/3 and 1/3 is arbitrary, which follows from Theorem 3.3.

Theorem 3.5. $\mathrm{QMA}_{\log} = \mathrm{BQP}$.

Proof: Assume $L \in \mathrm{QMA}_m$ for m logarithmic, and assume A is a QMA verification procedure that witnesses this fact and has completeness and soundness error less than $2^{-(m+2)}$. Let

$$Q_x = \left(I_m \otimes \langle 0^k|\right) A_x^\dagger \Pi_1 A_x \left(I_m \otimes |0^k\rangle\right).$$

Similar to the proof of Theorem 3.4, we have

$$x \in L \ \Rightarrow\ \mathrm{tr}(Q_x) \geq 3/4, \qquad x \notin L \ \Rightarrow\ \mathrm{tr}(Q_x) \leq 1/4.$$

We will describe a polynomial-time quantum algorithm B that decides L with bounded error. The algorithm B simply constructs a totally mixed state over m qubits and runs the verification procedure A using this state in place of Merlin's message. Running the verification procedure on the totally mixed state is equivalent to running the verification procedure on m qubits initialized to some uniformly generated standard basis state. This is easily simulated using

Hadamard transforms and controlled not gates on $2m$ qubits appropriately. The totally mixed state on m qubits corresponds to the density matrix $2^{-m} I_m$, from which it follows that the probability of acceptance of B is given by

$$\Pr[B \text{ accepts } x] = \operatorname{tr}\left(Q_x \, 2^{-m} I_m\right) = 2^{-m} \operatorname{tr}(Q_x).$$

Given that m is logarithmic in $|x|$, we have that the probabilities with which B accepts inputs $x \in L$ and inputs $x \notin L$ are bounded away from one another by the reciprocal of some polynomial. This difference can be amplified by standard methods, implying that $L \in \mathrm{BQP}$. ∎

4. QAM

A QAM verification procedure A consists of functions $m, r \in \mathit{poly}$ and a polynomial-time uniform family

$$\left\{ A_{x,y} \,:\, x \in \Sigma^*,\, y \in \Sigma^{r(|x|)} \right\}$$

of quantum circuits. As in the case of QMA verification procedures, each circuit $A_{x,y}$ acts on two collections of qubits: $m(|x|)$ qubits sent by Merlin and $k(|x|)$ qubits corresponding to Arthur's workspace. The notion of a circuit $A_{x,y}$ accepting a message $|\psi\rangle$ is defined in the same way as for QMA. In the present case, the string y corresponds to a sequence of coin-flips sent by Arthur to Merlin, on which Merlin's message may depend.

Definition 4.1. The class $\mathrm{QAM}(a, b)$ consists of all languages $L \subseteq \Sigma^*$ for which there exists a QAM verification procedure A satisfying the following conditions.

1. If $x \in L$ then there exists a collection of states $\{|\psi_y\rangle\}$ on m qubits such that
$$\frac{1}{2^r} \sum_{y \in \Sigma^r} \Pr[A_{x,y} \text{ accepts } |\psi_y\rangle] \;\geq\; a.$$
2. If $x \notin L$ then for every collection of states $\{|\psi_y\rangle\}$ on m qubits it holds that
$$\frac{1}{2^r} \sum_{y \in \Sigma^r} \Pr[A_{x,y} \text{ accepts } |\psi_y\rangle] \;\leq\; b.$$

As for QMA, one may consider the cases where a and b are constants or functions of $n = |x|$, and in the case that a and b are functions of the input length it is assumed that $a(n)$ and $b(n)$ can be computed deterministically in time polynomial in n. Also as before, let $\mathrm{QAM} = \mathrm{QAM}(2/3, 1/3)$.

4.1. Error reduction for QAM

The first fact about QAM that we prove is that completeness and soundness errors may be reduced by running many copies of the protocol in parallel. The proof is similar in principle to the proof of Lemma 14.1 in [12], which corresponds to our Theorem 3.2.

Proposition 4.2. *Let $a, b : \mathbb{N} \to (0, 1)$ satisfy*

$$a(n) - b(n) \geq \frac{1}{p(n)}$$

for all $n \geq \mathbb{N}$ for some $p \in \mathit{poly}$. Then for any $q \in \mathit{poly}$,

$$\mathrm{QAM}(a, b) = \mathrm{QAM}(1 - 2^{-q}, 2^{-q}).$$

Proof. Let $L \in \mathrm{QAM}(a, b)$, and let A be a QAM verification procedure witnessing this fact. We consider a new QAM verification procedure that corresponds to playing the game described by $\{A_{x,y}\}$ in parallel N times. The new procedure accepts if and only if the number of acceptances of the original game is at least $N \cdot \frac{a+b}{2}$. Although Merlin is not required to play the repetitions independently, we will show that playing the repetitions independently in fact gives him an optimal strategy. The proposition then follows by choosing an appropriately large value of N and applying a Chernoff-type bound.

Assume hereafter that the input x is fixed, and define

$$\begin{aligned} Q_y^{(0)} &= (I \otimes \langle 0^k|) A_{x,y}^\dagger \Pi_0 A_{x,y} (I \otimes |0^k\rangle), \\ Q_y^{(1)} &= (I \otimes \langle 0^k|) A_{x,y}^\dagger \Pi_1 A_{x,y} (I \otimes |0^k\rangle) \end{aligned}$$

for each $y \in \Sigma^r$. We have $Q_y^{(1)} = I - Q_y^{(0)}$, and consequently $Q_y^{(0)}$ and $Q_y^{(1)}$ share a complete set of orthonormal eigenvectors. Let $\{|\psi_{y,1}\rangle, \ldots, |\psi_{y,2^m}\rangle\}$ be such a set, and let

$$p_{y,1}^{(z)}, \ldots, p_{y,2^m}^{(z)}$$

be the corresponding eigenvalues for $Q_y^{(z)}$, $z \in \{0, 1\}$. As $Q_y^{(0)}$ and $Q_y^{(1)}$ are positive semidefinite and sum to the identity, $p_{y,i}^{(0)}$ and $p_{y,i}^{(1)}$ are nonnegative real numbers with $p_{y,i}^{(0)} + p_{y,i}^{(1)} = 1$ for each y and j. Assume without loss of generality that the eigenvectors and eigenvalues are ordered in such a way that

$$p_{y,1}^{(1)} \geq \cdots \geq p_{y,2^m}^{(1)}.$$

This implies that the maximum acceptance probability of $A_{x,y}$ is $p_{y,1}^{(1)}$.

Under the assumption that Arthur's coin-flips are given by $y_1, \ldots, y_N$, if Merlin plays the repetitions independently, and optimally for each repetition, his probability of convincing Arthur to accept is

$$\sum_{\substack{z_1, \ldots, z_N \in \Sigma \\ z_1 + \cdots + z_N \geq N \cdot \frac{a+b}{2}}} p_{y_1,1}^{(z_1)} \cdots p_{y_N,1}^{(z_N)}. \tag{3}$$

Without any assumption on Merlin's strategy, the maximum probability with which Merlin can win $N \cdot \frac{a+b}{2}$ repetitions of the original game when Arthur's coin-flips are given by

$y_1, \ldots, y_N$ is equal to the largest eigenvalue of

$$\sum_{\substack{z_1,\ldots,z_N \in \Sigma \\ z_1+\cdots+z_N \geq N \cdot \frac{a+b}{2}}} Q_{y_1}^{(z_1)} \otimes \cdots \otimes Q_{y_N}^{(z_N)}. \tag{4}$$

Therefore, to prove the proposition it suffices to show that these quantities are equal.

All of the summands in Eq. 4 share the complete set of orthonormal eigenvalues given by

$$\left\{ |\psi_{y_1,i_1}\rangle \cdots |\psi_{y_N,i_N}\rangle \,:\, i_1, \ldots, i_N \in \{1, \ldots, 2^m\} \right\},$$

and so this set also describes a complete set of orthonormal eigenvectors of the sum. The eigenvalue associated with $|\psi_{y_1,i_1}\rangle \cdots |\psi_{y_N,i_N}\rangle$ is

$$\sum_{\substack{z_1,\ldots,z_N \in \Sigma \\ z_1+\cdots+z_N \geq N \cdot \frac{a+b}{2}}} p_{y_1,i_1}^{(z_1)} \cdots p_{y_N,i_N}^{(z_N)}. \tag{5}$$

Define $u_1(X) = X$, $u_0(X) = 1 - X$, and let

$$f(X_1, \ldots, X_N) = \sum_{\substack{z_1,\ldots,z_N \in \Sigma \\ z_1+\cdots+z_N \geq N \cdot \frac{a+b}{2}}} u_{z_1}(X_1) \cdots u_{z_N}(X_N).$$

The quantity in Eq. 5 is equal to $f(p_{y_1,i_1}^{(1)}, \ldots, p_{y_N,i_N}^{(1)})$. Because f is a multilinear function that is nondecreasing in each variable, the maximum of the quantity in Eq. 5 is $f(p_{y_1,1}^{(1)}, \ldots, p_{y_N,1}^{(1)})$, which is equal to the quantity in Eq. 3. This completes the proof. ∎

4.2. An upper bound on QAM

In this section we observe the simple upper bound $\mathrm{QAM} \subseteq \mathrm{BP} \cdot \mathrm{PP}$. Recall that $L \in \mathrm{BP} \cdot \mathrm{PP}$ if and only if there exists a set $K \in \mathrm{PP}$ and a function $r \in \mathit{poly}$ such that

$$x \in L \Rightarrow \Pr\left[(x, y) \in K\right] \geq 2/3,$$
$$x \notin L \Rightarrow \Pr\left[(x, y) \in K\right] \leq 1/3,$$

where the probability is over $y \in \Sigma^{r(|x|)}$ chosen uniformly.

The following fact concerning the maximum probabilities of acceptance of $A_{x,y}$ for random y will be used. Here we let $\mu(A_{x,y})$ denote the maximum probability that $A_{x,y}$ can be made to accept (maximized over all choices of Merlin's message $|\psi_y\rangle$).

Proposition 4.3. *Suppose that*

$$\left\{ A_{x,y} \,:\, x \in \Sigma^*,\, y \in \Sigma^{r(|x|)} \right\}$$

is a QAM *verification procedure for a language L that has completeness and soundness errors bounded by 1/9. For any $x \in \Sigma^*$ and for $y \in \Sigma^r$ chosen uniformly at random,*

$$x \in L \;\Rightarrow\; \Pr[\mu(A_{x,y}) \geq 2/3] \;\geq\; 2/3$$
$$x \notin L \;\Rightarrow\; \Pr[\mu(A_{x,y}) \leq 1/3] \;\geq\; 2/3.$$

Proof. Suppose that $x \in L$. Let $z(y) = 1 - \mu(A_{x,y})$, and let Z be a random variable whose value is $z(y)$ for a uniformly chosen $y \in \Sigma^r$. The assumption of the proposition implies that $E[Z] \leq 1/9$. By Markov's inequality we have

$$\Pr[Z > 1/3] \leq \frac{E[Z]}{1/3} \leq 1/3,$$

and therefore

$$\Pr[\mu(A_{x,y}) \geq 2/3] = \Pr[Z \leq 1/3] \geq 2/3.$$

The proof for $x \notin L$ is similar. ∎

Theorem 4.4. $\mathrm{QAM} \subseteq \mathrm{BP} \cdot \mathrm{PP}$.

Proof. Let $L \in \mathrm{QAM}$, and let

$$A = \left\{ A_{x,y} \,:\, x \in \Sigma^*,\, y \in \Sigma^{r(|x|)} \right\}$$

be a QAM verification procedure for L with completeness and soundness errors bounded by 1/9. Such a procedure exists by Proposition 4.2. It follows from the proof of Theorem 3.4 that there exists a language $K \in \mathrm{PP}$ such that

$$\mu(A_{x,y}) \geq 2/3 \Rightarrow (x, y) \in K,$$
$$\mu(A_{x,y}) \leq 1/3 \Rightarrow (x, y) \notin K.$$

It is possible that $\mu(A_{x,y}) \in (1/3, 2/3)$ for some values of y, but in this case no requirement is made on whether or not $(x, y) \in K$. The theorem now follows from Proposition 4.3. ∎

Remark. At first glance one might expect the stronger relation $\mathrm{QAM} \subseteq \mathrm{BP} \cdot \mathrm{QMA}$ to hold, but we do not know whether or not this is the case.

5. QMAM

A QMAM verification procedure A consists of functions $m_1, m_2, r \in \mathit{poly}$ and a polynomial-time uniformly generated family

$$\left\{ A_{x,y} \,:\, x \in \Sigma^*,\, y \in \Sigma^{r(|x|)} \right\}$$

of quantum circuits. The functions m_1 and m_2 specify the number of qubits in Merlin's first and second messages to Arthur. Each circuit $A_{x,y}$ acts on $m_1(|x|)+m_2(|x|)+k(|x|)$ qubits, where as before $k(|x|)$ denotes the number of qubits corresponding to Arthur's workspace.

In the QMAM case, it becomes necessary to discuss possible actions that Merlin may perform rather than just discussing states that he may send. This is because Merlin's strategy could involve preparing some quantum state, sending part of that state to Arthur on the first message, and transforming the part of that state he did not send to Arthur (after receiving Arthur's coin-flips) in order to produce his second message.

Definition 5.1. A language $L \subseteq \Sigma^*$ is in $\mathrm{QMAM}(a,b)$ if there exists a QMAM verification procedure $\{A_{x,y}\}$ as above such that the following conditions are satisfied.

1. If $x \in L$ then for some l there exists a quantum state $|\psi\rangle$ on $m_1 + m_2 + l$ qubits and a collection of unitary operators $\{U_y : y \in \Sigma^r\}$ acting on $m_2 + l$ qubits such that
$$\frac{1}{2^r} \sum_{y \in \Sigma^r} \Pr[A_{x,y} \text{ accepts } (I_{m_1} \otimes U_y)|\psi\rangle] \geq a.$$
2. If $x \notin L$ then for every l, every quantum state $|\psi\rangle$ on $m_1 + m_2 + l$ qubits, and every collection of unitary operators $\{U_y : y \in \Sigma^r\}$ acting on $m_2 + l$ qubits,
$$\frac{1}{2^r} \sum_{y \in \Sigma^r} \Pr[A_{x,y} \text{ accepts } (I_{m_1} \otimes U_y)|\psi\rangle] \leq b.$$

The same assumptions regarding a and b apply in this case as in the QMA and QAM cases.

In the above definition, the circuit $A_{x,y}$ is acting on m_1+m_2 qubits sent by Merlin in addition to Arthur's k workspace qubits, while $(I_{m_1} \otimes U_y)|\psi\rangle$ is a state on m_1+m_2+l qubits. It is to be understood that the last l qubits of $(I_{m_1} \otimes U_y)|\psi\rangle$ remain in Merlin's possession, so $A_{x,y}$ is effectively tensored with the identity acting on these qubits.

5.1. Background on quantum interactive proofs

This section contains background information on quantum interactive proof systems that will be used to prove that quantum Arthur-Merlin games have the same power as arbitrary quantum interactive proof systems. A more complete discussion of quantum interactive proof systems can be found in Ref. [13].

As in the classical case, a quantum interactive proof system consists of two parties, a prover with unlimited computation power and a computationally bounded verifier. The prover and verifier may processes and exchange quantum information; the prover can perform arbitrary quantum computations while the verifier's computations must be described by polynomial-time uniform families of quantum circuits. It will only be necessary for us to discuss the particular case of three-message quantum interactive proof systems, as any (polynomial-message) quantum interactive proof system can be simulated by a three-message quantum interactive proof. Moreover, such a proof system may be taken to have perfect completeness and exponentially small soundness error.

For a fixed input x, a three-message quantum interactive proof system operates as follows. The verifier begins with a k-qubit register V and the prover begins with two registers: an m-qubit register M and an l-qubit register P. The register V corresponds to the verifier's work-space, the register M corresponds to the message qubits that are sent back and forth between the prover and verifier, and the register P corresponds to the prover's workspace. The register M begins in the prover's possession because the prover sends the first message. The verifier's work-space register V begins initialized to the state $|0^k\rangle$, while the prover initializes the pair (M,P) to some arbitrary quantum state $|\psi\rangle$.

In the first message, the prover sends M to the verifier. The verifier applies some unitary transformation V_1 to the pair (V,M) and returns M to the prover in the second message. The prover now applies some arbitrary unitary transformation U to the pair (M,P) and returns M to the verifier in the third and final message. Finally, the verifier applies a second unitary transformation V_2 to the pair (V,M) and measures the first qubit of the resulting collection of qubits in the standard basis. The outcome 1 is interpreted as accept and 0 is interpreted as reject.

Let Π_0, Π_1, Δ_0, and Δ_1 be projections defined as
$$\begin{aligned}
\Pi_1 &= |1\rangle\langle 1| \otimes I_{k+m-1}, \\
\Pi_0 &= |0\rangle\langle 0| \otimes I_{k+m-1}, \\
\Delta_1 &= |0^k\rangle\langle 0^k| \otimes I_m, \\
\Delta_0 &= I_{k+m} - \Delta_1.
\end{aligned}$$

In other words, these are $k+m$ qubit projections that act on the pair of registers (M,V); Π_1 and Π_0 are projections onto those states for which the first qubit of the register V is 1 or 0, respectively, and Δ_1 and Δ_0 are projections onto those states for which the register V contains the state $|0^k\rangle$ or contains a state orthogonal to $|0^k\rangle$, respectively. These are similar definitions to Eq. 1, but for notational convenience the first k qubits refer to the work-space qubits V and the last m qubits refer to the message qubits M.

The maximum probability with which a verifier specified by V_1 and V_2 can be made to accept is
$$\left\| (\Pi_1 V_2 \otimes I_l)(I_k \otimes U)(V_1 \otimes I_l)(|0^k\rangle|\psi\rangle) \right\|^2,$$
maximized over all choices of the state $|\psi\rangle$ and the unitary transformation U. The number l is determined by the prover's strategy, so one may maximize over this number as well. However, there is no loss of generality in assuming $l = m+k$, as it is always possible for a quantum prover to play optimally with this many work-space qubits.

There is another way to characterize the maximum acceptance probability for a given verifier based on the *fidelity* function: for two mixed-states ρ and ξ, fidelity between ρ and ξ is defined as
$$F(\rho,\xi) = \mathrm{tr}\sqrt{\sqrt{\rho}\,\xi\sqrt{\rho}}.$$

To describe this characterization we will need to define various sets of states of the pair of registers (V,M). For any

projection Λ on $k+m$ qubits let $\mathcal{S}(\Lambda)$ denote the set of all mixed states ρ of (V,M) that satisfy $\rho = \Lambda\rho\Lambda$, i.e., the collection of states whose support is contained in the space onto which Λ projects. Also let $\mathcal{S}_{\mathsf{V}}(\Lambda)$ denote the set of all reduced states of V that result from some state $\rho \in \mathcal{S}(\Lambda)$, i.e.,

$$\mathcal{S}_{\mathsf{V}}(\Lambda) = \{\mathrm{tr}_{\mathsf{M}}\,\rho \,:\, \rho \in \mathcal{S}(\Lambda)\},$$

where tr_{M} denotes the partial trace over the register M.

Lemma 5.2. *The maximum probability with which a verifier specified by V_1 and V_2 can be made to accept is*

$$\max\left\{F(\rho,\xi)^2 \,:\, \rho \in \mathcal{S}_{\mathsf{V}}(V_1\Delta_1 V_1^\dagger),\ \xi \in \mathcal{S}_{\mathsf{V}}(V_2^\dagger \Pi_1 V_2)\right\}.$$

This lemma is implicit in Ref. [13].

5.2. Equivalence of QMAM and QIP

In this section we prove that QMAM = QIP. Because quantum Arthur-Merlin games are a restricted form of quantum interactive proof systems, QMAM $\subseteq$ QIP is obvious. To prove the containment QIP $\subseteq$ QMAM, we will need the following lemmas in addition to the facts summarized in the previous section. The first lemma is a corollary of Uhlmann's Theorem (v. [17]).

Lemma 5.3. *Suppose the pair of registers (V,M) is in some mixed quantum state for which the reduced state of V is σ. If the pair (V,M) is measured with respect to a binary valued measurement described by orthogonal projections $\{\Lambda_0,\Lambda_1\}$, then the probability of obtaining the outcome 1 is at most $F(\sigma,\rho)^2$ for some $\rho \in \mathcal{S}_{\mathsf{V}}(\Lambda_1)$.*

The second lemma is a simple property of the fidelity function.

Lemma 5.4 (Refs. [16, 20]). *For density matrices ρ, ξ, and σ, we have $F(\rho,\sigma)^2 + F(\sigma,\xi)^2 \leq 1 + F(\rho,\xi)$.*

Now we have the required tools to prove the main theorem of this section, which follows.

Theorem 5.5. *Let $L \in$ QIP and let $p \in poly$. Then L has a three message quantum Arthur-Merlin game with completeness error 0 and soundness error at most $1/2 + 2^{-p(n)}$ on inputs of length n. Moreover, in this quantum Arthur-Merlin game, Arthur's message consists of a single coin-flip.*

Proof. Let $L \in$ QIP, which implies that L has a three-message quantum interactive proof system with completeness error 0 and soundness error $\varepsilon(n) = 2^{-2p(n)}$ on inputs of length n.

Using the notation described in Section 5.1, consider a QMAM verification procedure A that corresponds to the following actions for Arthur. (It will be assumed that the input x is fixed, as the uniformity of this QMAM verification procedure is clear given the uniformity of the verifier being simulated.)

1. Receive register V from Merlin.
2. Flip a fair coin and send the result to Merlin.
3. Receive register M from Merlin. If the coin flipped in step 2 was HEADS, apply V_2 to (V,M) and *accept* if the first qubit of V (i.e., the output qubit of the quantum interactive proof system) is 1, otherwise *reject*. If the coin in step 2 was TAILS, apply $V_1^\dagger$ to (V,M) and *accept* if all qubits of V are set to 0, otherwise *reject*.

Suppose first that $x \in L$, so that some prover, whose actions are described by a state $|\psi\rangle$ and a unitary operator U as discussed in the previous section, can convince V to accept with certainty. Then Merlin can convince Arthur to accept with certainty by acting as follows:

1. Prepare state $|0^k\rangle$ in register V and state $|\psi\rangle$ in registers (M,P). Apply V_1 to registers (V,M), and send V to Arthur.
2. If Arthur flips HEADS, apply U to (M,P) and send M to Arthur. If Arthur flips TAILS, send M to Arthur without applying U.

Now assume $x \notin L$, so that no prover can convince V to accept with probability exceeding ε. Suppose that the reduced density matrix of register V sent by Merlin is σ. By Lemmas 5.3 and 5.4, the probability that Arthur can be made to accept is at most

$$\frac{1}{2}F(\rho,\sigma)^2 + \frac{1}{2}F(\xi,\sigma)^2 \leq \frac{1}{2} + \frac{1}{2}F(\rho,\xi)$$

maximized over $\rho \in \mathcal{S}_{\mathsf{V}}(V_1\Delta_1 V_1^\dagger)$ and $\xi \in \mathcal{S}_{\mathsf{V}}(V_2^\dagger\Pi_1 V_2)$. By Lemma 5.2 this probability is at most

$$\frac{1}{2} + \frac{\sqrt{\varepsilon}}{2} \leq \frac{1}{2} + 2^{-p(|x|)},$$

which completes the proof. ∎

Corollary 5.6. *For any function $p \in poly$ we have*

$$\mathrm{QIP} \subseteq \mathrm{QMAM}(1, 1/2 + 2^{-p}).$$

Now, suppose that we have a QMAM protocol for a language L with perfect completeness and soundness error δ, and we repeat the protocol N times in parallel, accepting if and only if all N of the repetitions accept. It is clear that this resulting protocol has perfect completeness, because Merlin can play optimally for each parallel repetition independently and achieve an acceptance probability of 1 for any $x \in L$. In the case that $x \notin L$, Merlin can gain no advantage whatsoever over playing the repetitions independently, and so the soundness error decreases to δ^N as we would hope. This follows from the fact that the same holds for arbitrary three-message quantum interactive proof systems [13], of which three-message quantum Arthur-Merlin games are a restricted type. This implies the following corollary.

Corollary 5.7. *For any function $p \in poly$ we have*

$$\mathrm{QIP} = \mathrm{QMAM}(1, 2^{-p}).$$

5.3. More than three messages

Finally, we note that one may define quantum Arthur-Merlin games having any polynomial number of messages in a similar way to three-message quantum Arthur-Merlin games. Such games are easily seen to be equivalent in power to three-message quantum Arthur-Merlin games. Specifically, polynomial-message quantum Arthur-Merlin games will be special cases of quantum interactive proof systems, and can therefore be parallelized to three-message interactive proofs and simulated by three-message quantum Arthur-Merlin games as described in the previous section.

6. Open questions

Several interesting questions about quantum Arthur-Merlin games remain unanswered. Some examples include the following questions.

- Are there interesting examples of problems in QMA or QAM that are not known to be in AM? A similar question may be asked for QMAM vs. PSPACE.
- The question of whether there exists an oracle relative to which BQP is outside of PH appears to be a difficult problem. In fact it is currently not even known if there is an oracle relative to which $\mathrm{BQP} \not\subseteq \mathrm{AM}$. Is there an oracle relative to which QMA or QAM is not contained in AM? If so, what about QMA or QAM versus PH? Such results might shed some light on the problem of BQP vs. PH.
- Nisan and Wigderson [18] proved $\mathrm{almost\text{-}NP} = \mathrm{AM}$. Is it the case that $\mathrm{almost\text{-}QMA} = \mathrm{QAM}$?

Acknowledgments

Thanks to Dorit Aharonov, Oded Regev, and Umesh Vazirani for their comments on error reduction for QMA, Ashwin Nayak for helpful references, and Alexei Kitaev for discussions about quantum proof systems. This research was supported by CIAR and the Canada Research Chairs program.

References

[1] D. Aharonov and T. Naveh. Quantum NP – a survey. arXiv.org e-Print quant-ph/0210077, 2002.

[2] D. Aharonov and O. Regev. A lattice problem in quantum NP. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 210–219, 2003.

[3] L. Babai. Trading group theory for randomness. In *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, pages 421–429, 1985.

[4] L. Babai and S. Moran. Arthur-Merlin games: a randomized proof system, and a hierarchy of complexity classes. *Journal of Computer and System Sciences*, 36(2):254–276, 1988.

[5] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16): article 167902, 2001.

[6] L. Fortnow and J. Rogers. Complexity limitations on quantum computation. *Journal of Computer and System Sciences*, 59(2):240–252, 1999.

[7] S. Goldwasser and M. Sipser. Private coins versus public coins in interactive proof systems. In S. Micali, editor, *Randomness and Computation*, volume 5 of *Advances in Computing Research*, pages 73–90. JAI Press, 1989.

[8] D. Janzing, P. Wocjan, and T. Beth. "Identity check" is QMA-complete. arXiv.org e-Print quant-ph/0305050, 2003.

[9] J. Kempe and O. Regev. 3-Local Hamiltonian is QMA-complete. *Quantum Information and Computation*, 3(3):258–264, 2003.

[10] A. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[11] A. Kitaev. "Quantum NP". Talk at AQIP'99: Second Workshop on Algorithms in Quantum Information Processing, DePaul University, January 1999.

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

[13] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof system. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 608–617, 2000.

[14] E. Knill. Quantum randomness and nondeterminism. Technical Report LAUR-96-2186, Los Alamos National Laboratory, 1996. arXiv.org e-Print quant-ph/9610012.

[15] H. Kobayashi, K. Matsumoto, and T. Yamakami. Quantum Merlin-Arthur proof systems: Are multiple Merlins more helpful to Arthur? In *Proceedings of the 14th Annual International Symposium on Algorithms and Computation*, 2003.

[16] A. Nayak and P. Shor. On bit-commitment based quantum coin flipping. arXiv.org e-Print quant-ph/0206123, 2002.

[17] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[18] N. Nisan and A. Wigderson. Hardness vs. randomness. *Journal of Computer and System Sciences*, 49(2):149–167, 1994.

[19] R. Raz and A. Shpilka. On the power of quantum proofs. In *Proceedings of the Nineteenth Annual IEEE Conference on Computational Complexity*, 2004.

[20] R. Spekkens and T. Rudolph. Degrees of concealment and bindingness in quantum bit-commitment protocols. *Physical Review A*, 65: article 123410, 2002.

[21] M. Vyalyi. QMA=PP implies that PP contains PH. Electronic Colloquium on Computational Complexity Report TR03-021, 2003.

[22] J. Watrous. Succinct quantum proofs for properties of finite groups. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 537–546, 2000.

[23] J. Watrous. PSPACE has constant-round quantum interactive proof systems. *Theoretical Computer Science*, 292(3):575–588, 2003.