Non-interactive classical verification of quantum computation

Gorjan Alagic[1,2], Andrew M. Childs[2], Alex B. Grilo[3], and Shih-Han Hung[2]

[1] National Institute of Standards and Technology
[2] Department of Computer Science, UMIACS, and QuICS, University of Maryland
[3] CWI and QuSoft

Abstract. In a recent breakthrough, Mahadev constructed an interactive protocol that enables a purely classical party to delegate any quantum computation to an untrusted quantum prover. We show that this same task can in fact be performed *non-interactively* (with setup) and in *zero-knowledge*.

Our protocols result from a sequence of significant improvements to the original four-message protocol of Mahadev. We begin by making the first message instance-independent and moving it to an offline setup phase. We then establish a parallel repetition theorem for the resulting three-message protocol, with an asymptotically optimal rate. This, in turn, enables an application of the Fiat-Shamir heuristic, eliminating the second message and giving a non-interactive protocol. Finally, we employ classical non-interactive zero-knowledge (NIZK) arguments and classical fully homomorphic encryption (FHE) to give a zero-knowledge variant of this construction. This yields the first purely classical NIZK argument system for QMA, a quantum analogue of NP.

We establish the security of our protocols under standard assumptions in quantum-secure cryptography. Specifically, our protocols are secure in the Quantum Random Oracle Model, under the assumption that Learning with Errors is quantumly hard. The NIZK construction also requires circuit-private FHE.

1 Introduction

Quantum computing devices are expected to solve problems that are infeasible for classical computers. However, as significant progress is made toward constructing quantum computers, it is challenging to verify that they work correctly, particularly when devices reach scales where classical simulation is infeasible. This problem has been considered in various models, such as with multiple entangled quantum provers [42,35,25,30,24,37,18,27] or with verifiers who have limited quantum resources [14,13,36,2]. Such solutions are not ideal since they require assumptions about the ability of the provers to communicate or require the verifier to have some quantum abilities.

In a major breakthrough, Mahadev recently described the first secure protocol enabling a purely classical verifier to certify the quantum computations of a single untrusted quantum prover [34]. The Mahadev protocol uses a quantum-secure cryptographic assumption to give the classical verifier leverage over the quantum prover. The protocol is sound under the assumption that Learning with Errors (LWE) does not admit a polynomial-time quantum algorithm. This assumption is widely accepted, and underlies some of the most promising candidates for quantum-secure cryptography [3].

The Mahadev protocol. Mahadev's result settled a major open question concerning the power of *quantum-prover interactive arguments* (QPIAs). In a QPIA, two computationally-bounded parties (a quantum prover $\mathcal{P}$ and a classical verifier $\mathcal{V}$) interact with the goal of solving a decision problem. Mahadev's result showed that there is a four-round[4] QPIA for BQP with negligible completeness error and constant soundness error $\delta \approx 3/4$. The goal of the protocol is for the verifier to decide whether an input Hamiltonian H from a certain class (which is BQP-complete) has a ground state energy that is low (YES) or high (NO).

The protocol has a high-level structure analogous to classical Σ-protocols [21]:

1. $\mathcal{V}$ generates a private-public key pair (pk, sk) and sends pk to $\mathcal{P}$;
2. $\mathcal{P}$ prepares the ground state of H and then coherently evaluates a certain classical function f_{pk}. This yields a state of the form $\sum_x \alpha_x |x\rangle_X |f_{pk}(x)\rangle_Y$, where the ground state is in a subregister of X. $\mathcal{P}$ measures Y and sends the result y to $\mathcal{V}$. $\mathcal{P}$ holds a superposition over the preimages of y.
3. $\mathcal{V}$ replies with a uniformly random *challenge* bit $c \in \{0, 1\}$.
4. If $c = 0$ ("test round"), $\mathcal{P}$ measures X in the computational basis and sends the outcome. If $c = 1$ ("Hadamard round"), $\mathcal{P}$ measures X in the Hadamard basis and sends the outcome.

After the four message rounds above are completed, the verifier uses their knowledge of H and the secret key sk to either accept or reject the instance H.

Our results. In this work, we show that the Mahadev protocol can be transformed into protocols with significantly more favorable parameters, and with additional properties of interest. Specifically, we show how to build non-interactive protocols (with setup) for the same task, with negligible completeness and soundness errors. One of our protocols enables a verifier to publish a single public "setup" string and then receive arbitrarily many proofs from different provers, each for a different instance. We also construct a non-interactive protocol that satisfies the zero-knowledge property [10].

In principle, one could ask for slightly less interaction: the prover and the verifier receive the instance from a third party, and then the prover simply sends a proof to the verifier, with no setup. While we cannot rule such a protocol out, constructing it seems like a major challenge (and may even be impossible). Such a proof must be independent of the secret randomness of the verifier, making

[4] We take one round to mean a single one-way message from the prover to the verifier, or vice-versa. The Mahadev protocol involves four such messages.

it difficult to apply Mahadev's "cryptographic leash." Without cryptographic assumptions, such a protocol would imply $\mathsf{BQP} \subseteq \mathsf{MA}$ [1], which is unlikely.

All of our results are conditioned on the hardness of the LWE problem for quantum computers; we call this *the* LWE *assumption*. This assumption is inherited from the Mahadev protocol. For the zero-knowledge protocol, we also require fully-homomorphic encryption (FHE) with circuit privacy [38]. Our security proofs hold in the Quantum Random Oracle Model (QROM) [11]. For simplicity, we assume that the relevant security parameters are polynomial in the input BQP instance size n, so that efficient algorithms run in time $\mathrm{poly}(n)$ and errors are (ideally) negligible in n.

Transforming the Mahadev protocol. We apply several transformations to the Mahadev protocol:

1. making the first message instance-independent (i.e., moving it to an offline setup phase);
2. applying parallel repetition, via a new parallel repetition theorem;
3. adding zero-knowledge, by means of classical NIZKs and classical FHE; and
4. applying Fiat-Shamir (in the QROM [11]).

Establishing that these transformations satisfy desirable properties is challenging. For instance, since cheating provers can now be quantum, classical parallel repetition theorems do not apply.

Instance-independent setup. Our first transformation is relatively simple, at a high level. Instead of setting the basis choice depending on the 2-local term of that we want to measure, we can just pick the basis uniformly at random and the choice is correct with probability $\frac{1}{4}$. When we consider multiple copies of the ground state, and each copy is assigned both a random choice of basis and a 2-local terms, then about $\frac{1}{4}$ of the copies get a consistent assignment. Thus, we can make the initial message instance-independent (and move it to an offline setup phase) by increasing the number of parallel measurements by a constant factor. We explain this transformation in more detail in Section 3. We refer to the resulting protocol as "the three-round Mahadev protocol," denoted by $\mathfrak{M}$.

Parallel repetition. Parallel repetition of a protocol is a very desirable property since it decreases the soundness error exponentially, without increasing the number of rounds of interaction (as in serial repetition). Given the importance of the Mahadev protocol, parallel repetition could be a useful tool for applying it in practice. However, several complications arise when attempting to show this. First, the Mahadev protocol is clearly private-coin, which is precisely the category of protocol that is challenging even in the classical setting [6,29]. Second, classical proofs of parallel repetition typically involve constructing a single-copy prover that uses many rounds of nested rejection sampling. The quantum analogue of such a procedure, quantum rewinding, can only be applied in special circumstances [45,5] and seems difficult to apply to parallel repetition.

We establish our new parallel repetition theorem with alternative techniques, suited specifically for the Mahadev protocol. We show that, for NO instances, the accepting paths of the verifier for the two different challenges ($c = 0$ and $c = 1$) correspond to two nearly (computationally) orthogonal projectors. We also show that this persists in k-fold parallel repetition, meaning that each pair of distinct challenge strings $\mathbf{c}, \mathbf{c}' \in \{0,1\}^k$ corresponds to nearly orthogonal projectors. From there, a straightforward argument shows that the prover cannot succeed on a non-negligible fraction of challenge strings. We show that k-fold parallel repetition yields the same optimal soundness error δ^k as sequential repetition.

Taken together with the first transformation, the result is a three-round QPIA (with offline setup) for verifying BQP. We denote the k-fold parallel repetition of $\mathfrak{M}$ by $\mathfrak{M}^k$.

Theorem 1.1. *Under the* LWE *assumption,* $\mathfrak{M}^k$ *is a three-round protocol (with offline setup) for verifying* BQP *with completeness* $1 - \text{negl}(n)$ *and soundness error* $2^{-k} + \text{negl}(n)$*.*

Zero-knowledge. Zero-knowledge is a very useful cryptographic property of proof systems. Roughly, a protocol is zero-knowledge if the verifier "learns nothing" from the interaction with the honest prover, except that they have a "yes" instance. This notion is formalized by requiring an efficient simulator whose output distribution is indistinguishable from the distribution of the protocol outcomes.

In our next result, we show how to modify the protocol $\mathfrak{M}^k$ of Theorem 1.1 to achieve zero-knowledge against arbitrary classical verifiers. Our approach is similar to that of [19], but uses a purely classical verifier. Instead of the prover providing the outcomes of the measurements to be checked by the verifier (as in $\mathfrak{M}^k$), a classical non-interactive zero-knowledge proof (NIZK) is provided. However, the NP statement "the measurements will pass verification" depends on the inversion trapdoor of the verifier, which must remain secret from the prover. To overcome this obstacle, we use classical fully homomorphic encryption (FHE). In the setup phase, an encryption of the verifier's secret keys is provided to the prover, enabling the prover to later compute the NIZK homomorphically. To establish the zero-knowledge property, we require the FHE scheme to have circuit privacy, which means that the verifier cannot learn the evaluated circuit *from the ciphertext* provided by the prover. To prove the zero-knowledge property, we also need the extra assumption that the setup phase is performed by a trusted third party, since we cannot rely on the verifier to perform it honestly anymore.

In classical zero-knowledge arguments, it is common to consider efficient provers who are provided an NP-witness of the statement to prove. In the quantum setting, if we assume that the quantum polynomial-time prover has access to a quantum proof of a QMA statement,[5] we achieve the following.

[5] QMA is a quantum analogue of NP. In QMA, an untrusted quantum proof is given to a quantum poly-time verifier.

Theorem 1.2 (informal). *Under the* LWE *assumption, if circuit-private FHE exists, then there exists a three-round zero-knowledge argument for* QMA *(with trusted setup) with negligible completeness and soundness error.*

Fiat-Shamir transformation. In the above protocols (both $\mathfrak{M}^k$ and its ZK-variant), the second message of the verifier is a uniformly random $\mathbf{c} \in \{0,1\}^k$. In the final transformation, we eliminate this "challenge" round via the well-known Fiat-Shamir transform [23]: the prover generates the challenge bits $\mathbf{c} \in \{0,1\}^k$ themselves by evaluating a public hash function $\mathcal{H}$ on the transcript of the protocol thus far. In our case, this means that the prover selects[6] $\mathbf{c} := \mathcal{H}(H, pk, y)$. Of course, the verifier also needs to adapt their actions at the verdict stage, using $\mathbf{c} = \mathcal{H}(H, pk, y)$ when deciding acceptance/rejection. The resulting protocols now only have a setup phase and a single message from the prover to the verifier.

Fiat-Shamir (FS) is typically used to establish security in the Random Oracle Model, in the sense that FS preserves soundness up to negligible loss provided $\mathcal{H}$ has superpolynomially large range [7,40]. It is straightforward to see that this last condition is required; it is also the reason we applied parallel repetition prior to FS. A well-known complication in the quantum setting is that quantum computers can evaluate any public classical function $\mathcal{H}$ in superposition via the unitary operator $U_{\mathcal{H}} \colon |x\rangle|y\rangle \mapsto |x\rangle|y \oplus \mathcal{H}(x)\rangle$. This means we must use the Quantum Random Oracle Model (QROM) [11], which grants all parties oracle access to $U_{\mathcal{H}}$. Proving the security of transformations like FS in the QROM is the subject of recent research, and newly developed techniques have largely shown that FS in the QROM preserves soundness for so-called Σ-protocols [22,33]. Extending those results to our protocols is relatively straightforward. Applying FS to $\mathfrak{M}^k$ then yields the following.

Theorem 1.3. *Let* $k = \omega(\log n)$*, and let* $\mathsf{FS}(\mathfrak{M}^k)$ *denote the protocol resulting from applying Fiat-Shamir to the* k*-fold parallel repetition of the three-round Mahadev protocol. Under the* LWE *assumption, in the QROM,* $\mathsf{FS}(\mathfrak{M}^k)$ *is a non-interactive protocol (with offline setup) for verifying* BQP *with negligible completeness and soundness errors.*

If we instead apply the Fiat-Shamir transform to the zero-knowledge protocol from Theorem 1.2, we achieve the following.[7]

Theorem 1.4 (informal). *Under the* LWE *assumption, in the QROM, there exists a classical non-interactive zero-knowledge argument (with trusted offline setup) for* QMA*, with negligible completeness and soundness errors.*

Related results. After an initial version of our work was made public, showing how the Mahadev protocol can be reduced to four rounds using parallel repetition and the Fiat-Shamir transform, Chia, Chung, and Yamakawa posted a

[6] Here pk and y are k-tuples since we are transforming parallel-repeated protocols.

[7] Note that $\mathsf{FS}(\mathfrak{M}^k)$ in Theorem 1.3 is also a protocol for verifying QMA with negligible error if the prover is given a quantum witness.

preprint [17] describing the same result, with an alternative proof of parallel repetition. They also showed how to make the verifier run in polylog time using indistinguishability obfuscation. Our work was performed independently, and we subsequently improved our result to make the protocol non-interactive with setup and zero-knowledge.

Radian and Sattath [41] recently established what they call "a parallel repetition theorem for NTCFs," which are the aforementioned classical functions f_{pk}. However, the context of [41] is very different from ours and their parallel repetition theorem follows from a purely classical result.

Broadbent, Ji, Song, and Watrous [16] presented the first quantum zero-knowledge proofs for QMA with efficient provers. Vidick and Zhang [44] combined this protocol with the Mahadev protocol [34] to make the communication classical. Broadbent and Grilo [15] showed a "quantum Σ" zero-knowledge proof for QMA (with a quantum verifier). In the non-interactive setting, Coladangelo, Vidick, and Zhang [19] constructed a non-interactive zero-knowledge argument with quantum setup and Broadbent and Grilo [15] showed a quantum statistical zero-knowledge proof in the secret parameter model.

Open problems. This work raises several natural open questions. First, is it possible to prove the soundness of our protocol when the oracle $\mathcal{H}$ is instantiated with a concrete (e.g., correlation-intractable [39]) hash function? Our current analysis only applies in an idealized model.

Another natural line of work is studying parallel repetition for other QPIAs such as [26,44,12], perhaps including small modifications such as "random termination" as needed in purely classical private-coin protocols [29,31,8].

Finally, a similar classical NIZK protocol can also be achieved using the techniques of locally simulatable proofs [28,15]. We leave as an open problem understanding whether such a protocol could give us extra useful properties.

2 Preliminaries and notation

Most algorithms we consider are efficient, meaning that they run in time polynomial in both the input size (typically n) and the security parameter (typically λ). We assume that n and λ are polynomially-related. The two main classes of algorithms of interest are PPT (probabilistic poly-time) and QPT (quantum poly-time). We say that $f = \text{negl}(n)$ if $f = o(n^{-c})$ for every constant c. We denote by U_f the efficient map that coherently implements a classical function $f\colon \{0,1\}^n \to \{0,1\}^m$, i.e., $U_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle$, when there exists an efficient deterministic circuit that computes f.

2.1 The local Hamiltonian problem and verification for BQP

Any promise problem $L = (L_{\text{yes}}, L_{\text{no}}) \in \mathsf{QMA}$ can be reduced to the local Hamiltonian problem such that for $x \in L_{\text{yes}}$, the Hamiltonian H_x has a low-energy ground state $|\psi_x\rangle$, and for $x \in L_{\text{no}}$, all quantum states have large energy [32]. While the quantum witness $|\psi_x\rangle$ may be hard to prepare for general $L \in \mathsf{QMA}$, it can be prepared efficiently if $L \in \mathsf{BQP}$. Furthermore, the problem remains QMA-complete even with a Hamiltonian that can be measured by performing standard (Z) and Hadamard (X) basis measurements [9,20,36].

Problem 2.1. The 2-local ZX-Hamiltonian promise problem $\text{ZX}_{a,b} = (\text{ZX}_{\text{yes}}, \text{ZX}_{\text{no}})$, with parameters $a, b \in \mathbb{R}$, $b > a$ and gap $b - a > \text{poly}(n)^{-1}$, is defined as follows. An instance is a local Hamiltonian $H = \sum_{i<j} J_{ij}(X_i X_j + Z_i Z_j)$, where $J_{ij} \in \mathbb{R}$ with $2\sum_{i<j} |J_{ij}| = 1$ and each X_i (resp. Z_i) is a Pauli X (resp. Pauli Z) gate acting on the ith qubit. For $H \in \text{ZX}_{\text{yes}}$, the smallest eigenvalue of H is at most a, while if $H \in \text{ZX}_{\text{no}}$, the smallest eigenvalue of H is at least b.

Note that given the normalization factors, we can see that each term ($X_i X_j$ or $Z_i Z_j$) is associated with the probability $p_{ij} = |J_{ij}|$. When working with Hamiltonian terms S, we overload the notation for convenience. First, we write S_j to denote the Pauli operator assigned by S to qubit j, so that $S = \bigotimes_j S_j$. Second, we write $i \in S$ to indicate that i is a qubit index for which S does not act as the identity, i.e., $S_i \neq \mathbb{1}$. We let $p_S := p_{ij}$ for $i, j \in S$ and $m_S \in \{\pm 1\}$ be the sign of J_{ij}.

Morimae and Fitzsimons present a protocol (the "MF protocol") with a quantum prover $\mathcal{P}$ and a limited verifier $\mathcal{V}$ who only needs to perform single-qubit X and Z basis measurements [36]. $\mathcal{P}$ prepares the ground state of the Hamiltonian and sends it to $\mathcal{V}$, who then samples a term S with probability p_S and performs the corresponding measurement $\{M_{\pm 1} = \frac{\mathbb{1} \pm S}{2}\}$. Notice that Z or X basis measurements suffice to estimate the energy of S. The success probability with input state ρ is $\sum_S p_S \operatorname{tr}(M_{-m_S}\rho) = \frac{1}{2} - \frac{1}{2}\operatorname{tr}(H\rho)$, and negligible error can be achieved with parallel repetition.[8]

In the following discussion, we encode S by an n-bit string $h(S)$: for each $i \in S$, set $h_i = 0$ (resp. 1) for a Z (resp. X) basis measurement. For other qubits, the choice is irrelevant but we set $h_i = 0$ for concreteness. We let $\alpha_{h,\rho} := \operatorname{tr}(M_{-m_S}\rho)$ denote the success probability of the MF protocol described above with the state ρ, conditioned on the event that $h = h(S)$ is sampled. Thus the success probability with ρ is $\mathbb{E}_h[\alpha_{h,\rho}]$.

2.2 The Mahadev protocol for BQP verification

The Mahadev protocol relies crucially on two special classes of functions: Noisy Trapdoor Claw-free Functions (NTCFs) $\mathcal{F}$ and Noisy Trapdoor Injective Functions (NTIFs) $\mathcal{G}$. Both can be constructed based on the LWE assumption [12,34] and come with four polynomial-time algorithms $(\mathsf{Gen}_{\mathcal{F}}, \mathsf{Chk}_{\mathcal{F}}, \mathsf{Inv}_{\mathcal{F}}, \mathsf{Samp}_{\mathcal{F}})$ and

[8] $\mathcal{V}$ receives T copies of the ground state of H and performs an independent test on each copy. By accepting if at least $(2-a-b)T/4$ copies accept, both the completeness and soundness errors are suppressed to negligible with polynomial $T(|x|)$ (cf. [34, Theorem 8.4]). See [43, Section 3] for details.

$(\mathsf{Gen}_{\mathcal{G}}, \mathsf{Chk}_{\mathcal{G}}, \mathsf{Inv}_{\mathcal{G}}, \mathsf{Samp}_{\mathcal{G}})$. For complete details, and for the LWE construction, see [12,34].

The Mahadev protocol [34] for BQP verification allows $\mathcal{V}$ to request an X or Z basis measurement outcome without revealing the basis to $\mathcal{P}$. The aim of the protocol is to verify that the prover's response, when appropriately decoded, is close to the measurement outcomes of some n-qubit quantum state ρ. Crucially, this guarantee holds simultaneously for all basis choices $h \in \{0,1\}^n$, where 0 (resp. 1) denotes a Z (resp. X) basis measurement. With this guarantee, the verifier can then apply the verification procedure of the MF protocol to the decoded responses of the prover in order to decide acceptance or rejection.

In the following protocol, for each qubit, if Z (resp. X) basis measurement is desired, then an NTIF (resp. NTCF) key is sent. Since $\mathsf{Chk}_{\mathcal{F}}$ and $\mathsf{Chk}_{\mathcal{G}}$ (resp. $\mathsf{Samp}_{\mathcal{F}}$ and $\mathsf{Samp}_{\mathcal{G}}$) are identical [34], we denote them by Chk (resp. Samp). We let $\mathsf{Gen}(1^\lambda, h)$ for $h \in \{0,1\}^*$ denote the following key generation algorithm: for every bit i of h, run $(pk_i, sk_i) \leftarrow \mathsf{Gen}_{\mathcal{G}}(1^\lambda)$ if $h_i = 0$ and $(pk_i, sk_i) \leftarrow \mathsf{Gen}_{\mathcal{F}}(1^\lambda)$ if $h_i = 1$. Set $pk = (pk_i)_i$ and $sk = (sk_i)_i$ and output the key pairs (pk, sk).

Protocol 1 (Mahadev protocol).

Setup. *Choose a security parameter $\lambda \geq n$. Both $\mathcal{P}$ and $\mathcal{V}$ receive an instance of Problem 2.1, namely $H = \sum_S p_S \frac{\mathbb{1} + m_S S}{2}$.*

Round $\mathcal{V}_1$. *$\mathcal{V}$ samples r terms $S = (S_1, \ldots, S_r)$ and computes $h = h(S)$, the concatenation of $h(S_1), \ldots, h(S_r)$. $\mathcal{V}$ generates the key pair $(pk, sk) \leftarrow \mathsf{Gen}(1^\lambda, h)$ and sends pk to $\mathcal{P}$.*

Round $\mathcal{P}_1$. *$\mathcal{P}$ prepares $|\phi\rangle^{\otimes r} = \sum_{b \in \{0,1\}^{nr}} \phi_b |b\rangle_W$, r copies of the n-qubit ground state of H. For $j \in [r], \ell \in [n]$ and each qubit $W_{j\ell}$ in W, $\mathcal{P}$ performs Samp on input the key $pk_{j\ell}$ coherently and yields a state negligibly close to $\frac{1}{|\mathcal{X}|^{n/2}} \sum_{x \in \mathcal{X}^n} \sum_{b \in \{0,1\}^{nr}} \phi_b |b\rangle_W |x\rangle_X |\psi_{f_{pk}(b,x)}\rangle_Y$, where $|\psi_{f_{pk}(b,x)}\rangle := \bigotimes_{j \in [r], \ell \in [n]} |\psi_{f_{pk_{j\ell}}(b_{j\ell}, x_{j\ell})}\rangle$. Next, $\mathcal{P}$ measures Y and sends the outcome y to $\mathcal{V}$.*

Round $\mathcal{V}_2$. *$\mathcal{V}$ responds with a uniformly random "challenge" bit $c \in \{0,1\}$. We call $c = 0 =: \mathfrak{t}$ a "test round" and $c = 1 =: \mathfrak{h}$ a "Hadamard round."*

Round $\mathcal{P}_2$. *If $c = \mathfrak{t}$, $\mathcal{P}$ measures WX in the computational basis. If $c = \mathfrak{h}$, $\mathcal{P}$ measures WX in the Hadamard basis. In either case, $\mathcal{P}$ sends the measurement outcome (w, t) to $\mathcal{V}$.*

Verdict. *If $c = \mathfrak{t}$, $\mathcal{V}$ accepts if $\bigwedge_{j \in [r], \ell \in [n]} \mathsf{Chk}(pk_{j\ell}, w_{j\ell}, t_{j\ell}, y_{j\ell}) = 1$.*
If $c = \mathfrak{h}$, $\mathcal{V}$ performs the following: for each copy j and qubit $\ell \in [n]$,

1. *if $h_{j\ell} = 0$, run $(e_{j\ell}, x_{b_{j\ell}, y_{j\ell}}) \leftarrow \mathsf{Inv}_{\mathcal{G}}(sk_{j\ell}, y_{j\ell})$. If $h_{j\ell} = 1$, run $x_{0, y_{j\ell}} \leftarrow \mathsf{Inv}_{\mathcal{F}}(sk_{j\ell}, 0, y_{j\ell})$ and $x_{1, y_{j\ell}} \leftarrow \mathsf{Inv}_{\mathcal{F}}(sk_{j\ell}, 1, y_{j\ell})$ to get both preimages. Then compute $e_{j\ell} = t_{j\ell} \cdot (x_{0, y_{j\ell}} \oplus x_{1, y_{j\ell}}) \oplus w_{j\ell}$. If any of the above the algorithms returns* reject *for any j, ℓ, or $t_{j\ell}$ is trivial (e.g., 0^n, see [12]), reject. Otherwise store $e = (e_{j\ell})_{j \in [r], \ell \in [n]}$ as the witness to the next check.*
2. *$\mathcal{V}$ sets $v_j = 1$ if the witness $(e_{j\ell})_{\ell \in S_j}$ satisfies $M_{-m_{S_j}}$ (defined in Section 2.1).*[9]

[9] A string e satisfies a projector M if $\langle e|M|e\rangle = 1$, i.e., $|e\rangle$ lies in the support of M.

Finally, $\mathcal{V}$ accepts if $\sum_{j\in[r]} v_j \geq (2-a-b)r/4$.

Theorem 2.1 (Theorems 1.1 and 8.6 in [34]). *Under the* LWE *assumption, Protocol 1 is a four-message quantum-prover interactive argument for the class* BQP *with completeness error* $\mathrm{negl}(n)$ *and soundness error* $3/4 + \mathrm{negl}(n)$*.*

3 Instance-independent key generation

We now show how to generate the keys in the Mahadev protocol before the parties receive the input Hamiltonian, in an offline setup phase. To that end, we modify the MF protocol so the sampling of the Hamiltonian term is independent of the performed measurements. In our variant, for some $r = \mathrm{poly}(n)$, $\mathcal{V}$ samples n-bit strings $h_1, \dots, h_r$ uniformly and independent 2-local terms $S_1, \dots, S_r$ according to distribution π (in which S is sampled with the probability p_S from Section 2.1). We say the bases h_i and the terms S_i are *consistent* if, when the observable for the jth qubit in S_i is Z (resp., X) then the jth bit of h_i is 0 (resp., 1). Since h_i is uniformly sampled and S_i is 2-local, they are consistent with probability at least $\frac{1}{4}$.

In an r-copy protocol, we let $A := \{i \in [r] : h_i \text{ and } S_i \text{ are consistent}\}$ and denote $t = |A|$. For each $i \in A$, $\mathcal{V}_i$ decides as in the MF protocol: if $i \notin A$, then $\mathcal{V}_i$ accepts. Thus we consider the following protocol.

Protocol 2 (A modified parallel-repeated MF protocol for $\mathbf{zx}_{a,b}$).

Setup. *$\mathcal{V}$ samples the bases $h_1, \dots, h_r \leftarrow \{0,1\}^n$ uniformly.*

Round 1. *$\mathcal{P}$ sends the witness state ρ (r copies of the ground state).*

Round 2. *$\mathcal{V}$ measures the quantum state ρ in the bases $h_1, \dots, h_r$. For each copy $i \in [r]$, $\mathcal{V}$ samples terms $S_1, \dots, S_r \leftarrow \pi$. $\mathcal{V}$ records the subset $A \subseteq [r]$ of consistent copies. For each copy $i \in A$, $\mathcal{V}$ sets $v_i = 1$ if the outcome satisfies M_{-m_S} and 0 otherwise. $\mathcal{V}$ accepts if $\sum_{i\in A} v_i \geq (2-a-b)|A|/4$.*

For sufficiently large r, with high probability, there are around $r/4$ consistent copies. Thus to achieve the same completeness and soundness, it suffices to increase the number of copies by a constant factor. We thus have the following fact.

Lemma 3.1. *The completeness error and soundness error of Protocol 2 are negligible, provided $r = \omega\left(\frac{\log n}{(b-a)^2}\right)$ copies are used.*

Proof. First we observe that for each copy, with probability $1/4$, $\mathcal{V}$ measures the quantum state with a term sampled from the distribution π; otherwise $\mathcal{V}$ accepts. Thus for an instance H, the effective Hamiltonian to verify is $\widetilde{H}^{\otimes r}$ where $\widetilde{H} = \frac{3\mathbb{1}+H}{4}$. Following the standard parallel repetition theorem for QMA, we know that $\mathcal{P}$'s optimal strategy is to present the the ground state of $\widetilde{H}$, which is also the ground state of H.

With probability $\binom{r}{t}(\frac{1}{4})^t(\frac{3}{4})^{r-t}$, there are t consistent copies. Now for $i \in A$, we let X_i be a binary random variable corresponding to the decision of $\mathcal{V}_i$. For soundness, by Hoeffding's inequality[10] the success probability for A such that $|A| = t$ is

$$\begin{aligned}\Pr[\text{accept}|A] &= \Pr\left[\frac{1}{t}\sum_{i\in A} X_i \geq \frac{c+s}{2}\right] \\ &\leq \Pr\left[\frac{1}{t}\sum_{i\in A} X_i - s \geq \frac{c-s}{2}\right] \leq 2e^{-\frac{tg^2}{2}},\end{aligned}$$

where $g = c - s$ is the promise gap. Then the overall success probability is

$$\begin{aligned}\Pr[\text{accept}] &= 2\cdot 4^{-r}\sum_{t=0}^{r}\binom{r}{t}3^{r-t}e^{-tg^2/2} \\ &= 2\left(\frac{e^{-g^2/2}+3}{4}\right)^r \leq 2(1-g^2/16)^r \leq 2e^{-rg^2/16} \qquad (1)\end{aligned}$$

since $1 - x/2 \geq e^{-x}$ for $x \in [0,1]$ and $1 - x \leq e^{-x}$ for $x \geq 0$. Thus $r = \omega(g^{-2}\log n)$ suffices to suppress the soundness error to $n^{-\omega(1)}$. Since $g^{-1} = \text{poly}(n)$, polynomially many copies suffice to achieve negligible soundness error.

For completeness, again by Hoeffding's inequality,

$$\begin{aligned}\Pr[\text{reject}|A] &= \Pr\left[\frac{1}{t}\sum_{i\in A} X_i < \frac{c+s}{2}\right] \\ &\leq \Pr\left[c - \frac{1}{t}\sum_{i\in A} X_i > \frac{c-s}{2}\right] \leq 2e^{-\frac{tg^2}{2}}.\end{aligned}$$

By the same calculation as in (1), the completeness error is negligible if we set $r = \omega(g^{-2}\log n)$. □

Remark 3.1. The terms S_i are sampled *independently* of the interaction in the protocol. We let $\mathsf{term}(H, s)$ denote the deterministic algorithm that outputs a term from H according to distribution π when provided the randomness $s \in \{0,1\}^p$ for sufficiently large polynomial p. For bases $h \in \{0,1\}^{nr}$ and $s \in \{0,1\}^p$, $\alpha_{h,s,\rho}$ denotes the success probability when $\mathcal{P}$ sends the quantum state ρ.

The modifications to the MF protocol which resulted in Protocol 2 above can also be made (with minor adjustments) to the Mahadev protocol (Protocol 1). These changes are as follows:

1. In **Round** $\mathcal{V}_1$, the measurement bases h are sampled uniformly at random and S is not sampled.

[10] $\Pr[\frac{1}{n}\sum_i X_i - \mu \geq \delta] \leq e^{-2t\delta^2}$ for i.i.d. $X_1, \ldots, X_n \in [0,1]$.

2. In the **Verdict** stage for a Hadamard round ($c = 1$), $\mathcal{V}$ computes the measurement outcomes, as in check 1. Then $\mathcal{V}$ samples terms $S_1, \ldots, S_r \leftarrow \pi$ and for the consistent copies, $\mathcal{V}$ performs the check in 2.

We refer to this variant of Protocol 1 as "the three-round Mahadev protocol", and denote it by $\mathfrak{M}$.

4 A parallel repetition theorem for the Mahadev protocol

In a k-fold parallel repetition of $\mathfrak{M}$, an honest prover runs the honest single-fold prover independently for each copy of the protocol. Meanwhile, the honest verifier runs the single-fold verifier independently for each copy, accepting if and only if all k verifiers accept. The completeness error clearly remains negligible. To control soundness error, we establish a parallel repetition theorem.

In preparation, we fix the following notation related to the Verdict stage of $\mathfrak{M}$. We refer frequently to the notation from our description of Protocol 1 above, which applies to $\mathfrak{M}$ as well. First, the check $\bigwedge_{j \in [r], \ell \in [n]} \mathsf{Chk}(pk_{j\ell}, w_{j\ell}, t_{j\ell}, y_{j\ell}) = 1$ in a test round is represented by a projection $\mathit{\Pi}_{sk,\mathfrak{t}}$ acting on registers WXY. Specifically, this is the projector whose image is spanned by all inputs (w, t, y) that are accepted by the verifier in the Verdict stage. Note that running Chk does not require the trapdoor sk, but the relation implicitly depends on it. For notational convenience, we also denote $\mathit{\Pi}_{sk,\mathfrak{t}}$ as $\mathit{\Pi}_{s,sk,\mathfrak{t}}$, though the projector does not depend on s (defined in Remark 3.1). Second, the two Hadamard round checks 1 and 2 of the Verdict stage are represented by a projector $\mathit{\Pi}_{s,sk,\mathfrak{h}}$.

4.1 A lemma for the single-copy protocol

We begin by showing an important fact about the single-copy protocol: the verifier's accepting paths associated to the two challenges correspond to nearly orthogonal[11] projectors. Moreover, in a certain sense this property holds even for input states that are adaptively manipulated by a dishonest prover after they have learned which challenge will take place. This fact is essential in our analysis of the parallel repetition of many copies in the following sections.

The setup. As discussed in [34], any prover $\mathcal{P}$ can be characterized as follows. First, pick a state family $|\mathit{\Psi}_{pk}\rangle$; this state is prepared on registers $WXYE$ after receiving pk. Here Y is the register that will be measured in Round $\mathcal{P}_1$, W and X are the registers that will be measured in Round $\mathcal{P}_2$, and E is the private workspace of $\mathcal{P}$. Then, choose two unitaries $U_\mathfrak{t}$ and $U_\mathfrak{h}$ to describe the Round $\mathcal{P}_2$ actions of $\mathcal{P}$ before any measurements, in the test round and Hadamard round, respectively. Both $U_\mathfrak{t}$ and $U_\mathfrak{h}$ act on $WXYE$, but can only be classically controlled on Y, as they must be implemented after $\mathcal{P}$ has measured Y and sent the

[11] Strictly speaking, the projectors are only nearly orthogonal when applied to states prepared by efficient provers.

result to the verifier. (Of course, a cheating prover is not constrained to follow the honest protocol, but we can nevertheless designate a fixed subsystem Y that carries their message.) We will write $\mathcal{P} = (|\Psi_{pk}\rangle, U_{\mathfrak{t}}, U_{\mathfrak{h}})$, where it is implicit that $|\Psi_{pk}\rangle$ is a family of states parameterized by pk.

At the end of the protocol, the registers WXY are measured and given to the verifier. Recall that we can view the final actions of the verifier as applying one of two measurements: a test-round measurement or a Hadamard-round measurement. Let $\Pi_{s,sk,\mathfrak{t}}$ and $\Pi_{s,sk,\mathfrak{h}}$ denote the "accept" projectors for those measurements, respectively. For a given prover $\mathcal{P}$, we additionally define

$$\Pi^{U_{\mathfrak{t}}}_{s,sk,\mathfrak{t}} := U_{\mathfrak{t}}^{\dagger}(\Pi_{s,sk,\mathfrak{t}} \otimes \mathbb{1}_E)U_{\mathfrak{t}}\,, \quad \Pi^{U_{\mathfrak{h}}}_{s,sk,\mathfrak{h}} := U_{\mathfrak{h}}^{\dagger}(H_{WX}\Pi_{s,sk,\mathfrak{h}}H_{WX} \otimes \mathbb{1}_E)U_{\mathfrak{h}}\,,$$

where H_{WX} denotes the Hadamard transform on registers WX, i.e., the Hadamard gate applied to every qubit in those registers. These projectors have a natural interpretation: they describe the action of the two accepting projectors of the verifier on the initial state $|\Psi_{pk}\rangle$ of the prover, taking into account the (adaptive) attacks the prover makes in Round $\mathcal{P}_2$.

A key lemma. We now prove a fact about the single-copy protocol. The proof is largely a matter of making some observations about the results from [34], and then combining them in the right way.

Recall that, after the setup phase, for any instance H of the ZX-Hamiltonian problem (Problem 2.1), $\mathfrak{M}$ begins with the verifier $\mathcal{V}$ making a measurement basis choice $h \in \{0,1\}^{nr}$ for all the qubits. After interacting with a prover $\mathcal{P}$, the verifier either rejects or produces a candidate measurement outcome, which is then tested as in Protocol 2. We let $D_{\mathcal{P},h}$ denote the distribution of this candidate measurement outcome for a prover $\mathcal{P}$ and basis choice h, averaged over all measurements and randomness of $\mathcal{P}$ and $\mathcal{V}$. It is useful to compare $D_{\mathcal{P},h}$ with an "ideal" distribution $D_{\rho,h}$ obtained by simply measuring some (nr)-qubit quantum state ρ (i.e., a candidate ground state) according to the basis choices specified by h, with no protocol involved. Formally, we state the following lemma.

Lemma 4.1. *Let $\mathcal{P} = (|\Psi_{pk}\rangle, U_{\mathfrak{t}}, U_{\mathfrak{h}})$ be a prover in $\mathfrak{M}$ such that, for every $h \in \{0,1\}^{nr}$ and $s \in \{0,1\}^{p}$,*

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow \mathsf{Gen}(1^\lambda,h)}[\langle\Psi_{pk}|\Pi^{U_{\mathfrak{t}}}_{s,sk,\mathfrak{t}}|\Psi_{pk}\rangle] \geq 1 - \mathrm{negl}(n)\,. \tag{2}$$

Then there exists an (nr)-qubit quantum state ρ such that, for every h, s,

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow \mathsf{Gen}(1^\lambda,h)}[\langle\Psi_{pk}|\Pi^{U_{\mathfrak{h}}}_{s,sk,\mathfrak{h}}|\Psi_{pk}\rangle] \leq \alpha_{h,s,\rho} + \mathrm{negl}(n)\,,$$

where $\alpha_{h,s,\rho}$ (see Remark 3.1) is the success probability in the MF protocol with basis choice h and quantum state ρ.

Proof. Up to negligible terms, (2) means that $\mathcal{P}$ is what Mahadev calls a *perfect prover*. She establishes two results ([34, Claim 7.3] and [34, Claim 5.7]) which, when taken together, directly imply the following fact about perfect provers. For every perfect prover $\mathcal{P}$, there exists an efficiently preparable quantum state ρ such that $D_{\mathcal{P},h}$ is computationally indistinguishable from $D_{\rho,h}$ for all basis choices $h \in \{0,1\}^{nr}$. In particular, the proof is obtained in two steps. First, for every perfect prover, there exists a nearby "trivial prover" whose attack in a Hadamard round commutes with standard basis measurement on the committed state [34, Claim 5.7]. Second, for every trivial prover, the distribution is computationally indistinguishable from measuring a consistent quantum state ρ in any basis h [34, Claim 7.3]. Mahadev shows this for exactly perfect provers, but the proofs can be easily adapted to our "negligibly-far-from-perfect" case.

Now consider two ways of producing a final accept/reject output of the verifier. In the first case, an output is sampled from the distribution $D_{\mathcal{P},h}$ and the verifier applies the final checks in $\mathfrak{M}$. In this case, the final outcome is obtained by performing the measurement $\{\Pi^{U_{\mathfrak{h}}}_{s,sk,\mathfrak{h}}, \mathbb{1} - \Pi^{U_{\mathfrak{h}}}_{s,sk,\mathfrak{h}}\}$ on the state $|\Psi_{pk}\rangle$, and accepting if the first outcome is observed. In the second case, an output is sampled from the distribution $D_{\rho,h}$ and the verifier applies the final checks in the MF protocol. In this case, the acceptance probability is $\alpha_{h,s,\rho}$ simply by definition. The result then follows directly. □

Notice that for the soundness case, there is no state that succeeds non-negligibly in the MF protocol. In this case, Lemma 4.1 implies that for perfect provers the averaged projection

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow \mathsf{Gen}(1^\lambda,h),h,s}[\langle\Psi_{pk}|\Pi^{U_{\mathfrak{h}}}_{s,sk,\mathfrak{h}}|\Psi_{pk}\rangle]$$

is negligible. In other words, provers who succeed almost perfectly in the test round must almost certainly fail in the Hadamard round. We emphasize that this is the case even though the prover can adaptively change their state (by applying $U_{\mathfrak{t}}$ or $U_{\mathfrak{h}}$) after learning which round will take place. This formalizes the intuitive claim we made at the beginning of the section about "adaptive orthogonality" of the two acceptance projectors corresponding to the two round types.

4.2 The parallel repetition theorem

Characterization of a prover in the k-fold protocol. We now discuss the behavior of a general prover in a k-fold protocol. We redefine some notation, and let $\mathcal{V}$ be the verifier and $\mathcal{P}$ an arbitrary prover in the k-fold protocol.

In the Setup phase, the key pairs $(pk_1, sk_1), \ldots, (pk_k, sk_k)$ are sampled according to the correct NTCF/NTIF distribution.[12] The secret keys $sk = (sk_1, \ldots, sk_k)$[13] are given to $\mathcal{V}$, whereas $pk = (pk_1, \ldots, pk_k)$ is given to $\mathcal{P}$.

In Round $\mathcal{P}_1$, without loss of generality, the action of $\mathcal{P}$ prior to measurement is to apply a unitary $U_{0,pk}$ on the zero state $|0\rangle_{WXYE}$, producing the

[12] Recall that the keys are sampled by choosing uniform bases h and running $\mathsf{Gen}(1^\lambda, h)$.
[13] The verifier can learn the corresponding bases h from sk; see [34] for details.

state $|\Psi_{pk}\rangle_{WXYE} := U_{0,pk}|0\rangle_{WXYE}$. Each of W, X, Y is now a k-tuple of registers, and E is the prover's workspace. To generate the "commitment" message to $\mathcal{V}$, $\mathcal{P}$ performs standard basis measurement on Y. We write $|\Psi_{pk}\rangle_{WXYE} = \sum_y \beta_y|\Psi_{pk,y}\rangle_{WXE}|y\rangle_Y$. When the measurement outcome is y, the side state $\mathcal{P}$ holds is then $|\Psi_{pk,y}\rangle_{WXE}$. In the following analysis of the success probability of $\mathcal{P}$, we consider the superposition $|\Psi_{pk}\rangle_{WXYE}$ instead of a classical mixture of the states $|\Psi_{pk,y}\rangle_{WXE}$ using the principle of deferred measurement.

In Round $\mathcal{P}_2$, without loss of generality, the action of $\mathcal{P}$ consists of a general operation (that can depend on c), followed by the honest action. The general operation is some efficient unitary U_c on $WXYE$. The honest action is measurement in the right basis, i.e., for each i, W_iX_i is measured in the standard basis (if $c_i = 0$) or the Hadamard basis (if $c_i = 1$). Equivalently, the honest action is *(i.)* apply $\mathfrak{H}^c_{WX} := \bigotimes_{i=1}^k (H^{c_i})_{W_iX_i}$, i.e., for each $\{i : c_i = 1\}$ apply a Hadamard to every qubit of W_iX_i, and then *(ii.)* apply standard basis measurement.

In the Verdict stage, $\mathcal{V}$ first applies for each i the two-outcome measurement corresponding to the Π_{s_i,sk_i,c_i} from the single-copy protocol. The overall decision is then to accept if the measurements accept for all i. We let

$$(\Pi_{s,sk,c})_{WXY} := \bigotimes_{i=1}^{k} (\Pi_{s_i,sk_i,c_i})_{W_iX_iY_i} \tag{3}$$

denote the corresponding acceptance projector for the entire k-copy protocol. The effective measurement on $|\Psi_{pk}\rangle_{WXYE}$ is then described by the projection

$$\left(\Pi^{U_c}_{s,sk,c}\right)_{WXYE} := (U_c^\dagger)_{WXYE}(\mathfrak{H}^c \Pi_{s,sk,c,y} \mathfrak{H}^c \otimes \mathbb{1}_E)(U_c)_{WXYE} \,.$$

The success probability of $\mathcal{P}$, which is characterized by the state $|\Psi_{pk}\rangle$ and family of unitaries $\{U_c\}_{c\in\{0,1\}^n}$, is thus $\mathbb{E}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h),h,s,c}\left[\langle\Psi_{pk}|\Pi^{U_c}_{s,sk,c}|\Psi_{pk}\rangle\right]$.

The proof of parallel repetition. Recall that Lemma 4.1 states that the projectors corresponding to the two challenges in $\mathfrak{M}$ are nearly orthogonal, even when one takes into account the prover's adaptively applied unitaries. We show that this property persists in the k-copy protocol. Specifically, we show that all 2^k challenges are nearly orthogonal (in the same sense as in Lemma 4.1) with respect to any state $|\Psi_{pk}\rangle$ and any post-commitment unitaries U_c of the prover.

This can be explained informally as follows. For any two distinct challenges $c \neq c'$, there exists a coordinate i such that $c_i \neq c'_i$, meaning that one enters a test round in that coordinate while the other enters a Hadamard round. In coordinate i, by the single-copy result (Lemma 4.1), the prover who succeeds with one challenge should fail with the other. A complication is that, since we are dealing with an interactive argument, we must show that a violation of this claim leads to an *efficient* single-copy prover that violates the single-copy result. Once we have shown this, we can then apply it to any distinct challenge pairs $c \neq c'$. It then follows that we may (approximately) decompose $|\Psi_{pk}\rangle$ into components accepted in each challenge, each of which occurs with probability 2^{-k}. We can then use this decomposition to express the overall success probability of $\mathcal{P}$ in terms of this decomposition. As $|\Psi_{pk}\rangle$ is of course a normalized state, it will follow that the overall soundness error is negligibly close to 2^{-k}.

The "adaptive orthogonality" discussed above is formalized in Lemma 4.2. Recall that any prover in the k-fold parallel repetition of $\mathfrak{M}$ can be characterized by a state family $\{|\Psi_{pk}\rangle\}_{pk}$ that is prepared in Round $\mathcal{P}_1$ and a family of unitaries $\{U_c\}_{c\in\{0,1\}^k}$ that are applied in Round $\mathcal{P}_2$.

Lemma 4.2. *Let $\mathcal{P}$ be a prover in the k-fold parallel repetition of $\mathfrak{M}$ that prepares $|\Psi_{pk}\rangle$ in Round $\mathcal{P}_1$ and performs U_c in Round $\mathcal{P}_2$. Let $a, b \in \{0,1\}^k$ such that $a \neq b$ and choose i such that $a_i \neq b_i$. Then there is an (nr)-qubit quantum state ρ such that for every basis choice h and randomness s,*

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}\left[\langle\Psi_{pk}|\Pi^{U_b}_{s,sk,b}\Pi^{U_a}_{s,sk,a} + \Pi^{U_a}_{s,sk,a}\Pi^{U_b}_{s,sk,b}|\Psi_{pk}\rangle\right] \leq 2\alpha^{1/2}_{h_i,s_i,\rho} + \mathrm{negl}(n)\,,$$

where $\alpha_{h_i,s_i,\rho}$ (see Remark 3.1) is the success probability with ρ conditioned on the event that h_i is sampled.

Proof. Since we are proving an upper bound for a quantity that is symmetric under the interchange of b and a, we can assume that $b_i = 0$ and $a_i = 1$ without loss of generality.

We first claim that there exists a quantum state ρ such that

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}\left[\langle\Psi_{pk}|\Pi^{U_b}_{s,sk,b}\Pi^{U_a}_{s,sk,a}\Pi^{U_b}_{s,sk,b}|\Psi_{pk}\rangle\right] \leq \alpha_{h_i,s_i,\rho} + \mathrm{negl}(n) \tag{4}$$

for all basis choices h and randomness s. For a contradiction, suppose that is not the case. Then there exists a basis choice h^* and s^* and a polynomial η such that for every state ρ,

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h^*)}\left[\langle\Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}\Pi^{U_a}_{s^*,sk,a}\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\right] > \alpha_{h_i^*,s_i^*,\rho} + 1/\eta(n)\,.$$

We show that this implies the existence of an efficient prover $\mathcal{P}^*$ for the single-copy three-round Mahadev protocol $\mathfrak{M}$ who violates Lemma 4.1. Define the following projector on $WXYE$:

$$\Sigma_a := U_a^\dagger(H^a \otimes \mathbb{1}_E)((\mathbb{1}\otimes\cdots\otimes\mathbb{1}\otimes\Pi\otimes\mathbb{1}\otimes\cdots\otimes\mathbb{1})\otimes\mathbb{1}_E)(H^a\otimes\mathbb{1}_E)U_a\,.$$

Here Π denotes the single-copy protocol acceptance projector for the Hadamard round, with key sk_i and basis choice h_i^*, s_i^*. In the above, Π acts on the ith set of registers, i.e., $W_iX_iY_i$. The projector Σ_a corresponds to performing the appropriate Hadamard test in the ith protocol copy, and simply accepting all other copies unconditionally. It follows that $\Pi^{U_a}_{s,sk,a} \preceq \Sigma_a$, and we thus have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h^*)}\left[\langle\Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}\Sigma_a\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\right]\\
&\geq \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h^*)}\left[\langle\Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}\Pi^{U_a}_{s^*,sk,a}\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\right]\\
&> \alpha_{h_i^*,s_i^*,\rho} + 1/\eta.
\end{aligned} \tag{5}$$

The single-copy prover $\mathcal{P}^*$ interacts with the single-copy verifier $\mathcal{V}^*$ as follows.

- In the Setup phase, after receiving the public key pk^*, initialize $k-1$ internally simulated verifiers, and set pk to be the list of their keys, with pk^* inserted in the ith position. Let $h = (h_1, \ldots, h_k)$ be the basis choices, and note that all but h_i are known to $\mathcal{P}^*$.
- Using the algorithms of $\mathcal{P}$, perform the following repeat-until-success (RUS) procedure for at most $q = \eta^4$ steps.
  1. Prepare the state $|\Psi_{pk}\rangle$ on registers $WXYE$, and then apply the unitary U_b.
  2. Apply the measurement determined by $\Pi_{s,sk,b}$ (defined in (3)); for index i we can use pk^* because $b_i = 0$; for the rest we know the secret keys.
  3. If the measurement rejects, go to step (1.), and otherwise apply $U_b^\dagger$ and output the state.

  If the RUS procedure does not terminate within q steps, then $\mathcal{P}^*$ prepares a state[14] $|\Phi^*_{pk}\rangle$ by performing Samp coherently on $|0^n\rangle_W$ (see Round 2 of Protocol 1).

  Note that if $\mathcal{P}^*$ terminates within q steps, the resulting state is

  $$|\Phi_{pk}\rangle := \frac{\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle}{\|\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\|}\,;$$

  otherwise $|\Phi^*_{pk}\rangle$ is prepared.
- For the Round $\mathcal{P}_1$ message, measure the Y_i register of $|\Phi_{pk}\rangle$ and send the result to $\mathcal{V}^*$.
- When $\mathcal{V}^*$ returns the challenge bit w in Round 3, if $w = b_i = 0$, apply U_b (resp. $\mathbb{1}$) to $|\Phi_{pk}\rangle$ (resp. $|\Phi^*_{pk}\rangle$), and otherwise apply U_a. Then behave honestly, i.e., measure W_iX_i in computational or Hadamard bases as determined by w, and send the outcomes.

By the RUS construction and the fact that $b_i = 0$, the state $|\Phi_{pk}\rangle$ or $|\Phi^*_{pk}\rangle$ is in the image of the test-round acceptance projector in the ith coordinate. This means that, when $\mathcal{V}^*$ enters a test round, i.e., $w = 0 = b_i$, $\mathcal{P}^*$ is accepted perfectly. In other words, $\mathcal{P}^*$ is a perfect prover[15] and thus satisfies the hypotheses of Lemma 4.1.

Now consider the case when $\mathcal{V}^*$ enters a Hadamard round, i.e., $w = 1$. Let

$$\Omega := \{(pk, sk) : \langle\Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle > q^{-1/2}\}$$

denote the set of "good" keys. For $(pk, sk) \in \Omega$, the probability of not terminating within $q = \mathrm{poly}(n)$ steps is at most $(1 - q^{-1/2})^q \leq e^{-\sqrt{q}}$. Therefore, the

[14] To pass the test round, any efficiently preparable state suffices.

[15] While we used $\Pi_{h^*,sk,b}$ in the RUS procedure, and h_i^* is (almost always) not equal to the h_i selected by $\mathcal{V}^*$, the result is still a perfect prover state. This is because, as described in Protocol 1, the acceptance test in the test round is independent of the basis choice.

success probability of RUS for the good keys is $1 - \mathrm{negl}(n)$. Thus we have

$$\mathop{\mathbb{E}}_{sk|\Omega}[\langle \Phi_{pk}|\Sigma_a|\Phi_{pk}\rangle] \Pr[\Omega] \leq \alpha_{h_i^*,s_i^*,\rho} + \mathrm{negl}(n)$$

where we let $\mathbb{E}_{X|E}[f(X)] := \frac{1}{\Pr[E]}\sum_{x\in E} p(x)f(x)$ denote the expectation value of $f(X)$ conditioned on event E for random variable X over finite set $\mathcal{X}$ with distribution p and function $f\colon \mathcal{X} \to [0,1]$. Now we divide (5) into two terms and find

$$\begin{aligned}
\alpha_{h_i^*,s_i^*,\rho} + \eta^{-1} < {} & \mathop{\mathbb{E}}_{(pk,sk)}\left[\langle \Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}\Sigma_a\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\right] \\
= {} & \Pr[\Omega] \mathop{\mathbb{E}}_{(pk,sk)|\Omega}\left[\langle \Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}\Sigma_a\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\right] \\
& + \Pr[\overline{\Omega}] \mathop{\mathbb{E}}_{(pk,sk)|\overline{\Omega}}\left[\langle \Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}\Sigma_a\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\right] \\
\leq {} & \Pr[\Omega] \mathop{\mathbb{E}}_{(pk,sk)|\Omega}\left[\langle \Psi_{pk}|\Pi^{U_b}_{s^*,sk,b}\Sigma_a\Pi^{U_b}_{s^*,sk,b}|\Psi_{pk}\rangle\right] + q^{-1/2} \\
\leq {} & \alpha_{h_i^*,\rho} + \mathrm{negl}(n) + q^{-1/2}.
\end{aligned}$$

Since $q = \eta^4$, this is a contradiction. Therefore (4) holds for every h, s, i.e.,

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}[\langle \Psi_{pk}|\Pi^{U_b}_{s,sk,b}\Pi^{U_a}_{s,sk,a}\Pi^{U_b}_{s,sk,b}|\Psi_{pk}\rangle] \leq \alpha_{h_i,s_i,\rho} + \mathrm{negl}(n).$$

It then follows that

$$\begin{aligned}
& \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}\left[\langle \Psi_{pk}|\Pi^{U_b}_{h,sk,b}\Pi^{U_a}_{h,sk,a} + \Pi^{U_a}_{h,sk,a}\Pi^{U_b}_{h,sk,b}|\Psi_{pk}\rangle\right] \\
& = 2 \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}\left[\mathrm{Re}(\langle \Psi_{pk}|\Pi^{U_b}_{h,sk,b}\Pi^{U_a}_{h,sk,a}|\Psi_{pk}\rangle)\right] \\
& \leq 2 \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}\left[|\langle \Psi_{pk}|\Pi^{U_b}_{h,sk,b}\Pi^{U_a}_{h,sk,a}|\Psi_{pk}\rangle|\right] \\
& \leq 2 \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}\left[\langle \Psi_{pk}|\Pi^{U_b}_{h,sk,b}\Pi^{U_a}_{h,sk,a}\Pi^{U_b}_{h,sk,b}|\Psi_{pk}\rangle^{1/2}\right] \\
& \leq 2 \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}\left[\langle \Psi_{pk}|\Pi^{U_b}_{h,sk,b}\Pi^{U_a}_{h,sk,a}\Pi^{U_b}_{h,sk,b}|\Psi_{pk}\rangle\right]^{1/2} \leq 2\alpha^{1/2}_{h_i,s_i,\rho} + \mathrm{negl}(n)
\end{aligned}$$

as claimed. □

We remark that this adaptive orthogonality is guaranteed under a computational assumption. Assuming that no efficient quantum adversary can break the underlying security properties based on plain LWE, the projections are pairwise orthogonal in the sense of averaging over the key pairs (pk, sk) and with respect to any quantum state $|\Psi_{pk}\rangle$ prepared by an efficient quantum circuit.

We also emphasize that, in Lemma 4.2, for each pair $a \neq b$ the left-hand side is upper-bounded by the acceptance probability of measuring some state ρ in the basis h_i, and the quantum state ρ may be different among distinct choices of (a, b)

and i. This implies that if $\mathcal{P}$ succeeds with one particular challenge perfectly[16] when we average over h and s, Lemma 4.2 and standard amplification techniques (see Section 3) imply that it succeeds on challenge $b \neq a$ with probability at most $\mathbb{E}_{(pk,sk)\leftarrow \mathsf{Gen}(1^\lambda)} \langle \Psi_{pk} | \mathit{\Pi}_{s,sk,b} | \Psi_{pk} \rangle \leq \mathrm{negl}(n)$. We note that this strategy leads to acceptance probability at most $2^{-k} + \mathrm{negl}(n)$.

Since pairwise orthogonality holds with respect to *any* efficiently preparable quantum state by Lemma 4.2, our parallel repetition theorem follows.

First, we state a key technical lemma.

Lemma 4.3. *Let $A_1, \dots, A_m$ be projectors and $|\psi\rangle$ be a quantum state. Suppose there are real numbers $\delta_{ij} \in [0,2]$ such that $\langle \psi | A_i A_j + A_j A_i | \psi \rangle \leq \delta_{ij}$ for all $i \neq j$. Then $\langle \psi | A_1 + \cdots + A_m | \psi \rangle \leq 1 + \left(\sum_{i<j} \delta_{ij}\right)^{1/2}$.*

Proof. Let $\alpha := \langle \psi | A_1 + \ldots + A_m | \psi \rangle$. We have

$$\begin{aligned} \alpha^2 &\leq \langle \psi | (A_1 + \cdots + A_m)^2 | \psi \rangle \\ &= \alpha + \sum_{i<j} \langle \psi | A_i A_j + A_j A_i | \psi \rangle \qquad (6) \\ &\leq \alpha + \sum_{i<j} \delta_{ij} \end{aligned}$$

The first inequality holds since $|\psi\rangle\langle\psi| \preceq \mathbb{1}$, and thus

$$\langle \psi | (A_1 + \cdots + A_m) | \psi \rangle \langle \psi | (A_1 + \cdots + A_m) | \psi \rangle \leq \langle \psi | (A_1 + \cdots + A_m)^2 | \psi \rangle .$$

The equality (6) holds since each A_i is idempotent, and thus

$$\begin{aligned} \langle \psi | (A_1 + \cdots + A_m)^2 | \psi \rangle &= \langle \psi | A_1^2 + \cdots + A_m^2 | \psi \rangle + \sum_{i<j} \langle \psi | A_i A_j + A_j A_i | \psi \rangle \\ &= \langle \psi | A_1 + \cdots + A_m | \psi \rangle + \sum_{i<j} \langle \psi | A_i A_j + A_j A_i | \psi \rangle . \end{aligned}$$

Now observe that for $\beta > 0$, $x^2 \leq x + \beta$ implies $x \leq \frac{1}{2}(1 + \sqrt{1 + 4\beta}) \leq \frac{1}{2}(1 + (1 + 2\sqrt{\beta})) = 1 + \sqrt{\beta}$. Thus $\alpha \leq 1 + \sqrt{\sum_{i<j} \delta_{ij}}$ as claimed. □

Observe that when the projectors are mutually orthogonal, we have $A_1 + \cdots + A_m \preceq \mathbb{1}$ and the bound clearly holds. Lemma 4.3 describes a relaxed version of this fact. In our application, the projectors and the state are parameterized by the key pair, and we use this bound to show that the average of pairwise overlaps is small. We are now ready to establish our parallel repetition theorem.

Lemma 4.4. *Let k be a positive integer and let $\{U_c\}_{c \in \{0,1\}^k}$ be any set of unitaries that may be implemented by $\mathcal{P}$ after the challenge coins are sent. Let $|\Psi_{pk}\rangle$ be any state $\mathcal{P}$ holds in the commitment round, and suppose $\mathcal{P}$ applies U_c followed by honest measurements when the coins are c. Then there exists a negligible function ϵ such that $\mathcal{V}_1, \dots, \mathcal{V}_k$ accept $\mathcal{P}$ with probability at most $2^{-k} + \epsilon(n)$.*

[16] More concretely, if for some fixed a, $\mathit{\Pi}_{s,sk,a} | \Psi_{pk} \rangle = | \Psi_{pk} \rangle$.

Proof. The success probability of any prover in the k-fold protocol is

$$\Pr[\text{success}] = 2^{-k} \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h),h,s}[\langle\Psi_{pk}|\sum_c \Pi^{U_c}_{s,sk,c}|\Psi_{pk}\rangle]$$

where h, s are drawn from uniform distributions.

Define a total ordering on $\{0,1\}^k$ such that $a < b$ if $a_i < b_i$ for the smallest index i such that $a_i \neq b_i$. Then by Lemma 4.3, we have

$$\Pr[\text{success}] \le 2^{-k} + 2^{-k} \mathop{\mathbb{E}}_{h,s}\left[\sum_{a<b} \mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}[\langle\Psi_{pk}|\Pi^{U_a}_{s,sk,a}\Pi^{U_b}_{s,sk,b} + \Pi^{U_b}_{s,sk,b}\Pi^{U_a}_{s,sk,a}|\Psi_{pk}\rangle]\right]^{1/2}.$$

By Lemma 4.2, there exists a negligible function δ such that

$$\mathop{\mathbb{E}}_{(pk,sk)\leftarrow\mathsf{Gen}(1^\lambda,h)}[\langle\Psi_{pk}|\Pi^{U_a}_{s,sk,a}\Pi^{U_b}_{s,sk,b} + \Pi^{U_b}_{s,sk,b}\Pi^{U_a}_{s,sk,a}|\Psi_{pk}\rangle] \le 2\alpha^{1/2}_{h_{i(a,b)},\rho_{ab}} + \delta$$

for every pair (a, b). Here $i(a, b)$ is the smallest index i such that $a_i \neq b_i$ and ρ_{ab} is the reduced quantum state associated with a, b, as guaranteed by Lemma 4.2. Let μ be the soundness error of the MF protocol. We have

$$\begin{aligned}
\Pr[\text{success}] &\le 2^{-k} + 2^{-k} \mathop{\mathbb{E}}_{h,s}\left[\sum_{a<b}\left(2\alpha^{1/2}_{h_{i(a,b)},s_{i(a,b)},\rho_{ab}} + \delta\right)\right]^{1/2} \\
&\le 2^{-k} + 2^{-k} \mathop{\mathbb{E}}_{h,s}\left[\sum_{a<b} 2\alpha^{1/2}_{h_{i(a,b)},s_{i(a,b)},\rho_{ab}}\right]^{1/2} + 2^{-k}\sqrt{\binom{2^k}{2}}\delta^{1/2} \\
&\le 2^{-k} + 2^{-k}\left[\sum_{a<b} 2\left(\mathop{\mathbb{E}}_{h,s}[\alpha_{h_{i(a,b)},s_{i(a,b)},\rho_{ab}}]\right)^{1/2}\right]^{1/2} + \delta^{1/2} \\
&\le 2^{-k} + 2^{-k}\left[\sum_{a<b} 2\mu^{1/2}\right]^{1/2} + \delta^{1/2} \\
&\le 2^{-k} + 2^{-k}\left[2^k(2^k-1)\mu^{1/2}\right]^{1/2} + \delta^{1/2} \\
&\le 2^{-k} + \mu^{1/4} + \delta^{1/2}
\end{aligned}$$

where the second and third inequalities hold by Jensen's inequality. Amplifying the soundness of the MF protocol, μ is negligible using polynomially many copies by Lemma 3.1. Thus the soundness error is negligibly close to 2^{-k}. □

We note that Mahadev shows the soundness error for a single-copy protocol is negligibly close to $3/4$ [34], whereas Lemma 4.4 implies the error can be upper bounded by $1/2 + \mathrm{negl}(n)$. Mahadev obtains soundness error $3/4 + \mathrm{negl}(n)$ by considering a general prover $\mathcal{P}$ who, for each basis h, succeeds in the test round (characterized by $\Pi_{h,sk,\mathfrak{t}}$) with probability $1 - p_{h,\mathfrak{t}}$, in the first stage of the Hadamard round with probability $1 - p_{h,\mathfrak{h}}$, and in the second stage of the

Hadamard round with probability at most $\sqrt{p_{h,\mathfrak{t}}} + p_{h,\mathfrak{h}} + \alpha_{h,\rho} + \text{negl}(n)$ for some state ρ [34, Claim 7.1]. These contributions are combined by applying the triangle inequality for trace distance. This analysis is loose since the two stages are both classical, and $\mathcal{P}$ must pass both stages to win the Hadamard round.

Finally, Lemma 4.4 immediately implies the following theorem.

Theorem 4.1. *Let $\mathfrak{M}^k$ be the k-fold parallel repetition of the three-round Mahadev protocol $\mathfrak{M}$. Under the* LWE *assumption, the soundness error of $\mathfrak{M}^k$ is at most $2^{-k} + \text{negl}(n)$.*

For completeness, we present the three-round protocol $\mathfrak{M}^k$.

Protocol 3 (Verification with instance-independent setup).

Setup. *$\mathcal{V}$ samples random bases $h \in \{0,1\}^{nrk}$ and runs the key generation algorithm $(pk, sk) \leftarrow \mathsf{Gen}(1^\lambda, h)$. $\mathcal{V}$ samples a string $s \in \{0,1\}^{prk}$ uniformly. $\mathcal{V}$ sends the public keys pk to $\mathcal{P}$.*

Round $\mathcal{P}_1$. *$\mathcal{P}$ queries Samp coherently on the witness state $|\psi\rangle^{\otimes rk}$, followed by a standard basis measurement on register Y. The outcome is sent to $\mathcal{V}$.*

Round $\mathcal{V}_2$. *$\mathcal{V}$ samples $c_1, \ldots, c_k \leftarrow \{0,1\}$ and sends $c = (c_1, \ldots, c_k)$ to $\mathcal{P}$.*

Round $\mathcal{P}_2$. *For each $i \in [k]$, $j \in [r]$, $\ell \in [n]$,*

1. *if $c_i = 0$, $\mathcal{P}$ performs a standard basis measurement and gets $u_{ij\ell} = (w_{ij\ell}, t_{ij\ell})$;*
2. *if $c_i = 1$, $\mathcal{P}$ performs a Hadamard basis measurment and gets $u_{ij\ell} = (w_{ij\ell}, t_{ij\ell})$.*

$\mathcal{P}$ sends u to $\mathcal{V}$.

Verdict. *For each $i \in [k]$,*

1. *If $c_i = 0$, $\mathcal{V}$ accepts iff $\bigwedge_{j,\ell} \mathsf{Chk}(pk_{j\ell}, w_{j\ell}, t_{j\ell}, y_{j\ell}) = 1$.*
2. *If $c_i = 1$, $\mathcal{V}$ records the set $A_i \subseteq [r]$ of consistent copies. For each $j \in A_i$ and $\ell \in [n]$:*
   (a) *If $h_{ij\ell} = 0$, run $(b_{ij}, x_{b_{ij},y_{ij}}) \leftarrow \mathsf{Inv}_{\mathcal{G}}(sk_{ij}, y_{ij})$. Set $e_{ij\ell} = b_{ij\ell}$; if $h_{ij} = 1$, compute $x_{0,y_{ij\ell}}, x_{1,y_{ij\ell}}$ and $e_{ij\ell} = t_{ij\ell} \cdot (x_{0,y_{ij\ell}} \oplus x_{1,y_{ij\ell}}) \oplus w_{ij}$. If any of the algorithms rejects or any of $t_{ij\ell}$ is trivial (e.g., $t_{ij\ell} = 0$, see [34]), $\mathcal{V}$ sets $v_{ij} = 0$; otherwise enters the next step.*
   (b) *$\mathcal{V}$ computes the terms $S_{ij} = \mathsf{term}(H, s_{ij})$ for each $i \in [k], j \in [r]$. Set $v_{ij} = 1$ if $(e_{ij\ell})_{\ell \in S_{ij}}$ satisfies $M_{-m_{S_{ij}}}$ and $v_{ij} = 0$ otherwise.*

   Then $\mathcal{V}$ sets $v_i = 1$ if $\sum_{j \in A_i} v_{ij} \geq (2 - a - b)|A_i|/4$ and 0 otherwise.

$\mathcal{V}$ accepts iff $v_i = 1$ for every $i \in [k]$.

The verdict function is $\mathsf{verdict}(H, s, sk, y, c, u) := \bigwedge_{i=1}^{k} v_i$.

Theorem 4.2. *For $r = \omega(\frac{\log n}{(b-a)^2})$ and $k = \omega(\log n)$, Protocol 3 is a quantum prover interactive argument for $\text{ZX}_{a,b}$ with negligible completeness error and soundness error.*

5 A classical zero-knowledge argument for QMA

To turn $\mathfrak{M}^k$ into a zero-knowledge protocol, we first consider an intermediate protocol in which $\mathcal{P}$ first encrypts the witness state $|\psi\rangle^{\otimes rk}$ with a quantum one-time pad. Then in Round $\mathcal{P}_2$, $\mathcal{P}$ sends the one-time pad key β, γ along with the response u. In the verdict stage, $\mathcal{V}$ uses the keys to decrypt the response. We denote the verdict function as

$$\mathsf{verdict}'(H, s, sk, y, c, \beta, \gamma, u) := \mathsf{verdict}(H_{\beta,\gamma}, s, sk, y, c, u) \tag{7}$$

where $H_{\beta,\gamma} := X^{\beta} Z^{\gamma} H Z^{\gamma} X^{\beta}$ is the instance conjugated by the one-time pad.

Obviously, this is not zero-knowledge yet, as the verifier can easily recover the original measurement outcomes on the witness state. To address this issue, we take the approach of [16,19] and invoke a NIZK protocol for NP languages. The language $\mathcal{L}$ that we consider is defined as follows:

$$\begin{aligned}\mathcal{L} := \{(H, s, sk, \xi, y, c, \chi) :\ \exists\, \tau = (\beta, \gamma, u, r_1, r_2),\\ \xi = \mathsf{commit}(u; r_1) \wedge \chi = \mathsf{commit}(\beta, \gamma; r_2)\\ \wedge\, \mathsf{verdict}'(H, s, sk, y, c, \beta, \gamma, u) = 1\},\end{aligned}$$

where r_1, r_2 are the randomness for a computationally secure bit commitment scheme. However, this alone is insufficient since, to agree on an instance without introducing more message exchanges, $\mathcal{V}$ must reveal sk, s before $\mathcal{P}$ sends a NIZK proof. Revealing sk, s enables a simple attack on soundness: $\mathcal{P}$ can ensure the verifier accepts all instances by using the secret key to forge a valid response u, committing to it, and computing the NIZK proof.

The solution is to invoke a quantum-secure classical FHE scheme and to let $\mathcal{P}$ homomorphically compute a NIZK proof. This requires $\mathcal{P}$ to only use an encrypted instance. In the setup phase, $\mathcal{P}$ is given the ciphertexts $csk = \mathsf{FHE.Enc}_{hpk}(sk)$ and $cs = \mathsf{FHE.Enc}_{hpk}(s)$. Next, in Round $\mathcal{P}_2$, $\mathcal{P}$ computes $cx = \mathsf{FHE.Enc}_{hpk}(x)$ where $x := (H, s, sk, \xi, y, c, \chi)$ since the partial transcript (y, c, ξ, χ) has already been fixed. $\mathcal{P}$ then computes

$$ce = \mathsf{FHE.Eval}_{hpk}(\mathsf{NIZK.P}, cc, cx, c\tau) = \mathsf{FHE.Enc}_{hpk}(\mathsf{NIZK.P}(\mathsf{crs}, x, \tau)),$$

where $c\tau = \mathsf{FHE.Enc}_{hpk}(\tau)$, and sends ce to $\mathcal{V}$. Finally, $\mathcal{V}$ decrypts the encrypted NIZK proof ce and outputs $\mathsf{NIZK.V}(\mathsf{crs}, x, e)$. The above transformation yields a three-message zero-knowledge protocol for quantum computation with trusted setup from a third party, described as follows.

Protocol 4 (Setup phase $\mathsf{setup}(\lambda, N, M)$). *The algorithm* setup *takes two integers N, M as input, and outputs two strings $\mathsf{st}_{\mathcal{V}}, \mathsf{st}_{\mathcal{P}}$ with the following steps.*

1. *Run* $\mathsf{crs} \leftarrow \mathsf{NIZK.Setup}(1^{\lambda})$.
2. *Sample uniform bases $h \leftarrow \{0,1\}^N$ and run* $(pk, sk) \leftarrow \mathsf{Gen}(1^{\lambda}, h)$.
3. *Run the FHE key generation algorithm* $(hpk, hsk) \leftarrow \mathsf{FHE.Gen}(1^{\lambda})$.

4. *Run encryption on the secret key* $csk \leftarrow \mathsf{FHE.Enc}_{hpk}(sk)$.
5. *Choose keys* (β, γ) *and randomness* r_1 *uniformly and compute* $\xi = \mathsf{commit}(\beta, \gamma; r_1)$.
6. *Sample a random string* $s_1, \ldots, s_M \in \{0,1\}^p$ *(see Remark 3.1) uniformly and compute its encryption* $cs = \mathsf{FHE.Enc}_{hpk}(s)$.

Output $\mathsf{st}_{\mathcal{V}} = (\mathsf{crs}, sk, hsk, hpk, \xi)$ *and* $\mathsf{st}_{\mathcal{P}} = (\mathsf{crs}, pk, hpk, csk, cs, \beta, \gamma, r_1)$.

Protocol 5 (An interactive protocol).

Setup. *Run* $\mathsf{st}_{\mathcal{V}}, \mathsf{st}_{\mathcal{P}} \leftarrow \mathsf{setup}(\lambda, nrk, rk)$. *Send* $\mathsf{st}_{\mathcal{V}} = (\mathsf{crs}, sk, hsk, hpk, \xi)$ *to* $\mathcal{V}$ *and* $\mathsf{st}_{\mathcal{P}} = (\mathsf{crs}, pk, hpk, csk, cs, \beta, \gamma, r_1)$ *to* $\mathcal{P}$.

Round $\mathcal{P}_1$. $\mathcal{P}$ *aborts if* pk *is invalid.* $\mathcal{P}$ *queries* Samp *coherently on the witness state* $X^\beta Z^\gamma |\psi\rangle^{\otimes rk}$.

Round $\mathcal{V}_2$. $\mathcal{V}$ *samples* $c_1, \ldots, c_k \leftarrow \{0,1\}$ *and sends* $c = (c_1, \ldots, c_k)$ *to* $\mathcal{P}$.

Round $\mathcal{P}_2$. *For each* $i \in [k]$, $j \in [r]$, $\ell \in [n]$,

1. *if* $c_i = 0$, $\mathcal{P}$ *performs a standard basis measurement and gets* $u_{ij\ell} = (w_{ij\ell}, t_{ij\ell})$.
2. *if* $c_i = 1$, $\mathcal{P}$ *performs a Hadamard basis measurement and gets* $u_{ij\ell} = (w_{ij\ell}, t_{ij\ell})$.

$\mathcal{P}$ *sends* $\chi := \mathsf{commit}(u; r_2)$ *and*

$$ce := \mathsf{FHE.Eval}_{hpk}(\mathsf{NIZK.P}, cc, cx, c\tau),$$

where cc, cx *and* $c\tau$ *are the encryptions of* crs, x *and* τ *respectively.*

Verdict. $\mathcal{V}$ *accepts if* $\mathsf{NIZK.V}(\mathsf{crs}, x, \mathsf{FHE.Dec}_{hsk}(ce)) = 1$.

We show Protocol 5 is complete, sound, and zero-knowledge. For the detailed proofs, see the full version [4].

Theorem 5.1. *Protocol 5 has negligible completeness and soundness errors.*

Theorem 5.2. *Assuming the existence of a non-interactive bit commitment scheme with perfect binding and computational hiding, Protocol 5 is zero-knowledge.*

6 Round reduction by Fiat-Shamir transformation

In this section we show that the Fiat-Shamir transformation can be used make the k-fold parallel three-round Mahadev protocol $\mathfrak{M}$ non-interactive with a setup phase, while keeping both the completeness and the soundness errors negligible. This will also be the case for the zero-knowledge variant of the same, i.e., Protocol 5.

6.1 Fiat-Shamir for Σ-protocols in the QROM

The Fiat-Shamir (FS) transformation turns any public-coin three-message interactive argument, also called a Σ-protocol, into a single-message protocol in the random oracle model (ROM). In the standard approach, one proves that the Fiat-Shamir transformation preserves soundness in the ROM. In this idealized cryptographic model, all parties receive oracle access to a uniformly random function $\mathcal{H}$. Against quantum adversaries, there is a well-known complication: a quantum computer can easily evaluate any actual instantiation of $\mathcal{H}$ (with a concrete public classical function) in superposition via

$$U_{\mathcal{H}} \colon |x, y\rangle|z\rangle \mapsto |x, y\rangle|z \oplus \mathcal{H}(x, y)\rangle \, .$$

We thus work in the Quantum Random Oracle Model (QROM), in which all parties receive quantum oracle access to $U_{\mathcal{H}}$.

We make use of the following theorem of [22]; we describe the underlying reduction in the full version [4].

Theorem 6.1 (Quantum security of Fiat-Shamir [22, Theorem 2]). *For every QPT prover $\mathcal{A}^{\mathcal{H}}$ in the transformed protocol, there exists a QPT prover $\mathcal{S}$ for the underlying Σ-protocol such that*

$$\begin{aligned}&\Pr_{\Theta}[V(x, y, \Theta, m) = 1 : (y, m) \leftarrow \langle \mathcal{S}^{\mathcal{A}}, \Theta \rangle] \\ &\geq \frac{1}{2(2q+1)(2q+3)} \Pr_{\mathcal{H}}[V(x, y, \mathcal{H}(x, y), m) = 1, \;\; (y, m) \leftarrow \mathcal{A}^{\mathcal{H}}(x)] - \frac{1}{(2q+1)|\mathcal{Y}|} .\end{aligned}$$

In the above, $(y, m) \leftarrow \langle \mathcal{S}^{\mathcal{A}}, \Theta \rangle$ indicates that y and m are the first-round and third-round (respectively) messages of $\mathcal{S}^{\mathcal{A}}$, when it is given the random challenge Θ in the second round.

6.2 Extension to generalized Σ-protocols

In this section, we show that Fiat-Shamir also preserves soundness for a more general family of protocols, which we call "generalized Σ-protocols." In such a protocol, $\mathcal{V}$ can begin the protocol by sending an initial message to $\mathcal{P}$.

Protocol 6 (Generalized Σ-protocol). *Select a public function $f \colon \mathcal{R} \times L \to \mathcal{W}$, a finite set $\mathcal{C}$, and a distribution D over $\mathcal{R}$. The protocol begins with $\mathcal{P}$ and $\mathcal{V}$ receiving an input x.*

Round 1. *$\mathcal{V}$ samples randomness $r \in \mathcal{R}$ from distribution D and computes message $w = f(r, x)$, which is sent to $\mathcal{P}$.*
Round 2. *$\mathcal{P}$ sends a message y to $\mathcal{V}$.*
Round 3. *$\mathcal{V}$ responds with a uniformly random classical challenge $c \in \mathcal{C}$.*
Round 4. *$\mathcal{P}$ sends a response m to $\mathcal{V}$.*
Verdict. *$\mathcal{V}$ outputs a bit computed by a Boolean function $V(r, x, y, c, m)$.*

Notice that the original Mahadev protocol [34] is a generalized Σ-protocol: the distribution D describes the distribution for the secret key, and f computes the public key. Similarly, the k-fold parallel repetition of our instance-independent protocol is also a generalized Σ-protocol since our trusted setup phase can be seen as a message from the verifier.

Fiat-Shamir for generalized Σ protocols. The FS transformation for generalized Σ-protocols is similar to standard ones: in the Verdict stage, $\mathcal{V}$ computes $c = \mathcal{H}(x, w, y)$ and accepts if and only if $V(r, x, y, c, m) = 1$.

Protocol 7 (FS-transformed generalized Σ-protocol). *Select a public function $f : \mathcal{R} \times L \to \mathcal{W}$, a finite set $\mathcal{C}$, and a distribution D over $\mathcal{R}$. $\mathcal{P}$ and $\mathcal{V}$ receive an input x and are given access to a random oracle $\mathcal{H}$.*

Round 1. *$\mathcal{V}$ samples randomness $r \in \mathcal{R}$ from distribution D, and computes message $w = f(r, x)$, which is sent to $\mathcal{P}$.*
Round 2. *$\mathcal{P}$ sends a message (y, m) to $\mathcal{V}$.*
Verdict. *$\mathcal{V}$ computes $c = \mathcal{H}(x, w, y)$ and then outputs a bit computed by a Boolean function $V(r, x, y, c, m)$.*

To show that generalized Σ-protocols remain secure under the FS transformation, similarly to the idea for Σ-protocols, we give a reduction. Conditioned on any randomness r, the prover is $\mathcal{A}_r^{\mathcal{H}}(x) := \mathcal{A}^{\mathcal{H}}(x, f(r, x))$.[17] The prover $\mathcal{B}$ in the Σ-protocol runs $\mathcal{S}^{\mathcal{A}_r}$ and outputs its decision. Given the success probability of $\mathcal{A}$, we establish a lower bound on that of $\mathcal{B}$, as follows. For the proof, see the full version [4].

Lemma 6.1 (Fiat-Shamir transformation for generalized Σ-protocols). *Suppose that*

$$\Pr_{r,\mathcal{H}}[V(r, x, y, \mathcal{H}(x, f(r, x), y), m) = 1 : \; (y, m) \leftarrow \mathcal{A}^{\mathcal{H}}(x, f(r, x))] = \epsilon.$$

Then

$$\Pr_{r,\Theta}[V(r, x, y, \Theta, m) = 1 : \; (y, m) \leftarrow \langle \mathcal{B}, \Theta \rangle] \geq \frac{\epsilon}{2(2q+1)(2q+3)} - \frac{1}{(2q+1)|\mathcal{Y}|}.$$

Lemma 6.1 immediately gives the following theorem.

Theorem 6.2. *If a language L admits a generalized Σ-protocol with soundness error s, then after the Fiat-Shamir transformation, the soundness error against provers who make up to q queries to a random oracle is $O(sq^2 + q|\mathcal{Y}|^{-1})$.*

Proof. Suppose there is a prover who succeeds in the transformed protocol with success probability ϵ. Then by Lemma 6.1, we may construct a prover who succeeds with probability at least $\frac{\epsilon}{O(q^2)} - O\left(\frac{1}{q|\mathcal{Y}|}\right)$. By the soundness guarantee, we have $\frac{\epsilon}{O(q^2)} - O\left(\frac{1}{q|\mathcal{Y}|}\right) \leq s$ and thus $\epsilon \leq O(q^2 s + q|\mathcal{Y}|^{-1})$. □

By Theorem 6.2, if both s and $|\mathcal{Y}|^{-1}$ are negligible in security parameter λ, the soundness error of the transformed protocols remains negligible against an efficient prover who makes $q = \text{poly}(\lambda)$ queries. Theorem 1.3 follows directly from Theorem 6.2.

[17] Though the prover does not learn the private randomness r, its action depends on r implicitly.

6.3 Non-interactive zero-knowledge for QMA

We now show that, using the Fiat-Shamir transformation, our three-round protocol proposed in Protocol 5 can be converted into a non-interactive zero-knowledge argument (with trusted setup) for QMA in the Quantum Random Oracle model. The resulting protocol is defined exactly as Protocol 5, with two modifications: (i.) instead of Round $\mathcal{V}_2$, the prover $\mathcal{P}$ computes the coins c by evaluating the random oracle $\mathcal{H}$ on the protocol transcript thus far, and (ii.) the NIZK instance x is appropriately redefined using these coins.

We remark that since the setup in this protocol is trusted, it follows from Theorem 6.2 that the compressed protocol is complete and sound, and therefore we just need to argue about the zero-knowledge property.

Theorem 6.3. *The Fiat-Shamir transformation of Protocol 5 is zero-knowledge.*

Proof. The simulator $\mathcal{S}^{\mathcal{V}_2^*}$ can sample the trapdoor keys for NTCF/NTIF functions and private keys for the FHE scheme, enabling simulation of the transcript for every challenge sent by the verifier. In particular, one can run the same proof with the variant $\mathcal{S}^{\mathcal{H}}$ that queries the random oracle $\mathcal{H}$ for the challenges instead of receiving it from a malicious verifier $\mathcal{V}^*$. □

Acknowledgments

We thank Kai-Min Chung, Andrea Coladangelo, Bill Fefferman, Serge Fehr, Christian Majenz, Christian Schaffner, Umesh Vazirani, and Thomas Vidick for helpful discussions.

AMC and SHH acknowledge support from NSF grant CCF-1813814 and from the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Testbed Pathfinder program under Award Number DE-SC0019040. GA acknowledges support from the NSF under grant CCF-1763736, from the U.S. Army Research Office under Grant Number W911NF-20-1-0015, and from the U.S. Department of Energy under Award Number DE-SC0020312. Part of this work was completed while GA was supported by the Dutch Research Council (NWO) through travel grant 040.11.708. Part of this work was completed while AG was visiting the Joint Center for Quantum Information and Computer Science, University of Maryland and the Simons Institute for the Theory of Computing.

References

1. Scott Aaronson. BQP and the polynomial hierarchy. In *STOC 2010*, pages 141–150, 2010. arXiv:0910.4698.
2. Dorit Aharonov, Michael Ben-Or, Elad Eban, and Urmila Mahadev. Interactive proofs for quantum computations. 2017. arXiv:1704.04487.

3. Gorjan Alagic, Jacob Alperin-Sheriff, Daniel Apon, David N. Cooper, Quynh Dang, Yi-Kai Liu, Carl Frederick Miller, Dustin Moody, Rene Peralta, Ray Perlner, Angela Robinson, and Daniel Smith-Tone. Status report on the first round of the NIST post-quantum cryptography standardization process. 2019.
4. Gorjan Alagic, Andrew M. Childs, Alex B. Grilo, and Shih-Han Hung. Non-interactive classical verification of quantum computation. 2019. arXiv:1911.08101.
5. Andris Ambainis, Ansis Rosmanis, and Dominique Unruh. Quantum attacks on classical proof systems: The hardness of quantum rewinding. In *FOCS 2014*, pages 474–483, 2014. Cryptology ePrint Archive, Report 2014/296.
6. Mihir Bellare, Russell Impagliazzo, and Moni Naor. Does parallel repetition lower the error in computationally sound protocols? In *FOCS 1997*, pages 374–383, 1997.
7. Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *CCS 1993*, pages 62–73, 1993.
8. Itay Berman, Iftach Haitner, and Eliad Tsfadia. A tight parallel-repetition theorem for random-terminating interactive arguments, 2019. Cryptology ePrint Archive, Report 2019/393.
9. Jacob D. Biamonte and Peter J. Love. Realizable Hamiltonians for universal adiabatic quantum computers. *Physical Review A*, 78(1):012352, 2008. arXiv:0704.1287.
10. Manuel Blum, Paul Feldman, and Silvio Micali. Non-interactive zero-knowledge and its applications. In *STOC 1988*, page 103–112, 1988.
11. Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random oracles in a quantum world. In *ASIACRYPT 2011*, pages 41–69, 2011. arXiv:1008.0931.
12. Zvika Brakerski, Paul Christiano, Urmila Mahadev, Umesh Vazirani, and Thomas Vidick. A cryptographic test of quantumness and certifiable randomness from a single quantum device. In *FOCS 2018*, pages 320–331, 2018. arXiv:1804.00640.
13. Anne Broadbent. How to verify a quantum computation. *Theory of Computing*, 14(11):1–37, 2018. arXiv:1509.09180.
14. Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In *FOCS 2009*, pages 517–526, 2009. arXiv:0807.4154.
15. Anne Broadbent and Alex B. Grilo. Zero-knowledge for QMA from locally simulatable proofs, 2019. arXiv:1911.07782.
16. Anne Broadbent, Zhengfeng Ji, Fang Song, and John Watrous. Zero-knowledge proof systems for QMA. In *FOCS 2016*, pages 31–40, 2016. arXiv:1604.02804.
17. Nai-Hui Chia, Kai-Min Chung, and Takashi Yamakawa. Classical verification of quantum computations with efficient verifier, 2019. arXiv:1912.00990.
18. Andrea Coladangelo, Alex B. Grilo, Stacey Jeffery, and Thomas Vidick. Verifier-on-a-leash: New schemes for verifiable delegated quantum computation, with quasi-linear resources. In *EUROCRYPT 2019*, pages 247–277, 2019. arXiv:1708.07359.
19. Andrea Coladangelo, Thomas Vidick, and Tina Zhang. Non-interactive zero-knowledge arguments for QMA, with preprocessing, 2019. arXiv:1911.07546.
20. Toby Cubitt and Ashley Montanaro. Complexity classification of local Hamiltonian problems. *SIAM Journal on Computing*, 45(2):268–316, 2016. arXiv:1311.3161.
21. Ivan Damgård. On Σ-protocols. *Lecture Notes, University of Aarhus, Department for Computer Science*, 2002.
22. Jelle Don, Serge Fehr, Christian Majenz, and Christian Schaffner. Security of the fiat-shamir transformation in the quantum random-oracle model. In *CRYPTO 2019*, pages 356–383, 2019. Cryptology ePrint Archive, Report 2019/190.
23. Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In *CRYPTO 1986*, pages 186–194, 1986.

24. Joseph F. Fitzsimons and Michal Hajdušek. Post hoc verification of quantum computation, 2015. arXiv:1512.04375.
25. Alexandru Gheorghiu, Elham Kashefi, and Petros Wallden. Robustness and device independence of verifiable blind quantum computing. *New Journal of Physics*, 17(8):083040, 2015. arXiv:1502.02571.
26. Alexandru Gheorghiu and Thomas Vidick. Computationally-secure and composable remote state preparation, 2019. arXiv:1904.06320.
27. Alex B. Grilo. A simple protocol for verifiable delegation of quantum computation in one round. In *ICALP 2019*, pages 28:1–28:13, 2019.
28. Alex Bredariol Grilo, William Slofstra, and Henry Yuen. Perfect zero knowledge for quantum multiprover interactive proofs. In *FOCS 2019*, pages 611–635, 2019. arXiv:1905.11280.
29. Iftach Haitner. A parallel repetition theorem for any interactive argument. In *FOCS 2009*, pages 241–250, 2009.
30. Michal Hajdušek, Carlos A. Pérez-Delgado, and Joseph F. Fitzsimons. Device-independent verifiable blind quantum computation, 2015. arXiv:1502.02563.
31. Johan Håstad, Rafael Pass, Douglas Wikström, and Krzysztof Pietrzak. An efficient parallel repetition theorem. In *TCC 2010*, pages 1–18, 2010.
32. Alexei Yu. Kitaev, Alexander H. Shen, and Mikhail N. Vyalyi. *Classical and quantum computation*. American Mathematical Society, 2002.
33. Qipeng Liu and Mark Zhandry. Revisiting post-quantum Fiat-Shamir. In *CRYPTO 2019*, 2019. Cryptology ePrint Archive, Report 2019/262.
34. Urmila Mahadev. Classical verification of quantum computations. In *FOCS 2018*, pages 259–267, 2018. arXiv:1804.01082.
35. Matthew McKague. Interactive proofs for BQP via self-tested graph states. *Theory of Computing*, 12(3):1–42, 2016. arXiv:1309.5675.
36. Tomoyuki Morimae and Joseph F. Fitzsimons. Post hoc verification with a single prover, 2016. arXiv:1603.06046.
37. Anand Natarajan and Thomas Vidick. A quantum linearity test for robustly verifying entanglement. In *STOC 2017*, pages 1003–1015, 2017. arXiv:1610.03574.
38. Rafail Ostrovsky, Anat Paskin-Cherniavsky, and Beni Paskin-Cherniavsky. Maliciously circuit-private FHE. In *CRYPTO 2014*, pages 536–553, 2014. Cryptology ePrint Archive, Report 2013/307.
39. Chris Peikert and Sina Shiehian. Noninteractive zero knowledge for NP from (plain) learning with errors, 2019. Cryptology ePrint Archive, Report 2019/158.
40. David Pointcheval and Jacques Stern. Security arguments for digital signatures and blind signatures. *Journal of Cryptology*, 13(3):361–396, 2000.
41. Roy Radian and Or Sattath. Semi-quantum money, 2019. arXiv:1908.08889.
42. Ben W. Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013. arXiv:1209.0448.
43. Thomas Vidick and John Watrous. Quantum proofs. *Foundations and Trends in Theoretical Computer Science*, 11(1-2):1–215, 2016. arXiv:1610.01664.
44. Thomas Vidick and Tina Zhang. Classical zero-knowledge arguments for quantum computations, 2019. arXiv:1902.05217.
45. John Watrous. Zero-knowledge against quantum attacks. *SIAM Journal on Computing*, 39(1):25–58, 2009. arXiv:quant-ph/0511020.