Interactive proof systems with multiple provers sharing entanglement

Travis Hance

6.845 Paper - Fall 2012

1 Introduction

Interactive proof systems are one of the most important areas of study in modern complexity theory. One of the first exciting results of this field is the celebrated equality $\mathsf{IP} = \mathsf{PSPACE}$ due to Lund-Fortnow-Karloff-Nisan [7] and Shamir [10], which first showed the potential of interactive proof systems. Later, Babai, Fortnow, and Lund [1] show that $\mathsf{MIP} = \mathsf{NEXP}$, where MIP, introduced by Ben-Or, Goldwasser, Kilian, and Wigderson [2], is the class of languages which admit an interactive proof system with multiple provers which cannot communicate.

In this paper, we will study the class MIP^*, first defined by Cleve, Høyer, Toner, and Watrous [3]. This is the class of languages which admit an interactive proof system where the provers cannot directly communicate but have access to entangled quantum states; for example, two of the provers can share an EPR pair. Initially, it is not obvious whether or not MIP^* is more or less powerful than MIP. The only obvious bound is $\mathsf{IP} \subseteq \mathsf{MIP}$, since we can simulate an IP protocol by only talking to one prover, as the entanglement between the provers cannot possibly help in that case.

Besides MIP^*, we can also consider the class QMIP^*, introduced by Kobayashi and Matsumoto [6], where in addition to giving the provers entanglement, we let the verifier by quantum as well. Reichardt, Unger, and Vazirani [8] show that in fact $\mathsf{QMIP}^* = \mathsf{MIP}^*$.

Finally, Ito and Vidick [5] showed that $\mathsf{NEXP} \subseteq \mathsf{MIP}^*$. Although the details of the proof are quite technical, we will summarize the main ideas of the proof. The main idea is to give a multi-prover interactive protocol for an NEXP-complete problem. The protocol will be very similar to that used in [1] to show $\mathsf{NEXP} \subseteq \mathsf{MIP}$. The main contribution of Ito and Vidick [5] is that the protocol is sound even against entangled provers.

In Section 2 we describe MIP^* more formally. Then in Section 3 we give an interactive protocol for a NEXP-complete language and show that it is complete. Finally, in Sections 4 and 5, we outline the proof that the protocol is sound.

2 Description of MIP^*

We define a *multi-prover interactive proof system* with k provers. There is a classical *verifier* V and k *provers* $P_1, \ldots, P_k$ with unbounded computation. There is also a shared "quantum tape" with k registers, where each prover has access to one of the registers. The quantum tape starts in some

state $|\Psi\rangle \in \mathcal{H}^{\otimes n}$, where $\mathcal{H}$ is a Hilbert space of arbitrarily large dimension ℓ.[1] In each round, the verifier does a polynomial amount of classical computation. Then, the verifier sends a message m_i of polynomial length to each prover P_i. Depending on m_i, P_i may choose to make a measurement on his register of the quantum tape. Then, the prover may send back a message depending on m_i and the outcome of the measurement.[2]

We say that a language L is in $\mathsf{MIP}^*(k, m, p, q)$ if there is an m-round, k-prover protocol such that for any x:

- **Completeness.** If $x \in L$, there exists an ℓ, an initial quantum state $|\Psi\rangle$ and a strategy for the provers such that V accepts with probability at least p.
- **Soundness.** If $x \notin L$, for all ℓ, initial quantum states $|\Psi\rangle$, and strategies for the provers, V accepts with probability at most q.

We let $\mathsf{MIP}^* = \mathsf{MIP}^*(k, m, 2/3, 1/3)$. The constants $2/3$ and $1/3$ are arbitrary; there is no difficulty in amplifying error probabilities in MIP^*. The prover simply runs the protocol multiple times in succession, which works because the soundness condition guarantees soundness for any initial state.

Ito and Vidick [5] show that $\mathsf{NEXP} \subseteq \mathsf{MIP}^*(3, \mathsf{poly}, 1, c)$ and this is the result whose proof we will summarize in later sections.

It will be convenient to introduce the idea of a symmetric strategy. We say that a protocol is symmetric if the verifier treats all provers symmetrically. Likewise, we say that a prover strategy is symmetric if the initial state $|\Psi\rangle$ is invariant with respect to permutation of the registers and each prover employs the same strategy (with his own register).

Proposition 1. *If the verifier uses a symmetric protocol, then the provers can do no better than to use a symmetric strategy,*

Proof. Consider any prover strategy, and let $|\Psi\rangle$ be the initial state in this strategy. Let $\mathcal{A}_1, \ldots, \mathcal{A}_k$ be the registers of the k provers. Augment the registers with $\mathcal{B}_1, \ldots, \mathcal{B}_k$, large enough to store numbers in $\{1, \ldots, k\}$. For a permutation $\sigma \in S_k$, let $|\Psi_\sigma\rangle$ be $|\Psi\rangle$ with the registers permuted according to σ. Create a new strategy whose initial state is

$$\frac{1}{\sqrt{k!}} \sum_{\sigma \in S_k} |\sigma(1)\rangle_{\mathcal{B}_1} \otimes \cdots \otimes |\sigma(k)\rangle_{\mathcal{B}_k} \otimes |\Psi_\sigma\rangle_{\mathcal{A}_1 \cdots \mathcal{A}_k}$$

In this new strategy, prover P_i measures register $\mathcal{B}_i$. If he measures $|j\rangle$, then he performs the strategy of P_j in the original strategy.

Since V implements a symmetric protocol, this new strategy has the same success probability as the original. But this new strategy is a symmetric strategy, proving the proposition. □

[1] We could have instead allowed $|\Psi\rangle$ to start in a mixed state. This definition would be equivalent for the same reason that shared randomness does not help in the classical MIP setting. It is always best to just use the pure state which gives the highest probability of success for a given input.

[2] We do not need to explicitly account for the possibility that P_i may use a randomized strategy, since he can obtain random bits from the measurement.

3 An interactive protocol for NEXP

3.1 NEXP-complete problems and arithmetization

To prove that NEXP $\subseteq$ MIP, Babai, Fortnow, and Lund [1] use the following NEXP-complete problem:

Problem 2. Consise 3-SAT

Input: An exponentially large 3-SAT formula F in exponentially many variables, indexed by $\{0,1\}^n$. The formula is given by a polynomial-sized circuit C which, given 3 variables, gives constraints on those 3 variables (i.e., a set of clauses, each of which has those three variables).

Problem: Is F satisfiable?

One can show that this problem is NEXP-complete just as the standard Cook-Levin proof shows that 3-SAT in NP-complete. Indeed, if one considers the tableau of an exponentially long computation and converts it into a 3-SAT formula, it is easy to see that, given three variables, it is easy to compute the constraints between them.

To give an interactive protocol for Consise 3-SAT, Babai, Fortnow, and Lund [1] use the technique of *arithmetization*, a trick which was also used to give an interactive protocol for PSPACE, showing IP = PSPACE [7, 10]. The idea behind arithmetization is that we can convert a boolean circuit into an arithmetic circuit over a field $\mathbb{F}$. We will work with finite fields, and as a technical condition, we will need $|\mathbb{F}|$ to be sufficiently large.

Problem 3. Consise 3-SAT (arithmetized)

Input: Integers n, r, and d (in unary), a field $\mathbb{F}$, and an arithmetic circuit for a polynomial $f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, a_1, a_2, a_3)$, where $\boldsymbol{z}$ reprents r variables, and $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$ represent n variables each. You are promised that f has degree at most d in each variable.

Problem: Does there exist a map $\mathcal{W} : \{0,1\}^n \to \{0,1\}$ such that

$$f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, \mathcal{W}(\boldsymbol{b}_1), \mathcal{W}(\boldsymbol{b}_2), \mathcal{W}(\boldsymbol{b}_3)) = 0 \tag{1}$$

for all $\boldsymbol{z} \in \{0,1\}^r$ and $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \{0,1\}^n$?

In the arithmetized version, the map $\mathcal{W}$ corresponds to the map from variables to truth values, and f is the arithmetized version of a circuit which checks if three variables are consistent. The auxiliary variables in $\boldsymbol{z}$ appear in the reduction in order to "flatten out" the circuit, since we need to keep the degree d low. (These variables appear for the same reason that extra variables appear when reducing an arbitrary SAT instance to 3-SAT.)

Ito and Vidick [5] use a slightly different problem, but they use the same arithmetization idea:

Problem 4. Consise 3-coloring

Input: An exponentially large graph G with vertices indexed by $\{0,1\}^n$. The adjacency matrix is given by a polynomial-sized circuit $C : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$.

Problem: Is G 3-colorable?

Problem 5. Consise 3-coloring (arithmetized)

Input: Integers n, r, and d (in unary), a field $\mathbb{F}$, $\alpha \in \mathbb{F} \backslash \{0,1\}$, and an arithmetic circuit for a polynomial $f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, a_1, a_2)$, where $\boldsymbol{z}$ reprents r variables, and $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ represent n variables each. You are promised that f has degree at most d in each variable.

Problem: Does there exist a map $\mathcal{W} : \{0,1\}^n \to \{0,1,\alpha\}$ such that

$$f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \mathcal{W}(\boldsymbol{b}_1), \mathcal{W}(\boldsymbol{b}_2)) = 0 \tag{2}$$

for all $\boldsymbol{z} \in \{0,1\}^r$ and $\boldsymbol{b}_1, \boldsymbol{b}_2 \in \{0,1\}^n$?

This problem is the same in spirit as Consise 3-SAT, but it is more convenient because the constraints in (2) depend only on two value of $\mathcal{W}$, rather than three values as in (1). This lets us use one less prover than we would need otherwise, since ultimately, we will need an extra prover for each additional value of $\mathcal{W}$ we query. Babai, Fortnow, and Lund [1] do not worry about the optimizing for the number of proves as we do, since it was already known that any language with an MIP protocol also admits a protocol with only two provers (Ben-Or, Goldwasser, Kilian, and Wigderson [2]).

3.2 The AND Test

Consider the following problem:

Problem 6. AND problem
Input: Given oracle access to a function $h : \mathbb{F}^m \to \mathbb{F}$, promised that it is a polynomial of degree at most d in each variable.
Problem: Does $h(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \{0,1\}^m$?

Babai, Fortnow, and Lund [1] give a single prover protocol for the AND problem which works over finite fields to help solve Succinct 3-SAT. Based on this, Ito and Vidick [5] prove the slightly more general statement:

Proposition 7. *There is a single-prover protocol for the AND problem which has perfect completeness and soundness error at most $5/8 + t(m,d)/|\mathbb{F}|$, where t is some polynomial. Furthermore, the protocol has the following form: the verifier reads only one value from the oracle. Specifically, he chooses a random vector $\boldsymbol{x} \in \mathbb{F}^m$, reads the value $h(\boldsymbol{x})$ from the oracle, and then interacts with the prover.*

3.3 A protocol for Concise 3-coloring

How can we use the AND test to give a protocol for Consise 3-coloring? First, given an instance of Consise 3-coloring, we reduce it to the arithmetized version, using a sufficiently large field $\mathbb{F}$. (Various things, such as Proposition 7, will require large $\mathbb{F}$, but none will require it to be more than polynomially large.) can be applied. Now, the provers want to convince the verifier that there exists $\mathcal{W}$ such that (2) holds for all $\boldsymbol{z}$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$. To apply the AND problem protocol, we want to view $f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \mathcal{W}(\boldsymbol{b}_1), \mathcal{W}(\boldsymbol{b}_2))$ as a polynomial in $\boldsymbol{z}$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$. To do this, we extend $\mathcal{W} : \{0,1\}^n \to \{0,1,\alpha\}$ to a multilinear function $\widetilde{\mathcal{W}} : \mathbb{F}^n \to \mathbb{F}$. Then if we let

$$h(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2) := f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, \widetilde{\mathcal{W}}(\boldsymbol{b}_1), \widetilde{\mathcal{W}}(\boldsymbol{b}_2)) \tag{3}$$

we get that h is a polynomial in $\boldsymbol{z}$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$ of bounded degree.

Now, we can construct a proof system for arithmetized succinct 3-colorability as follows (informally): the provers want to prove that they know some $\mathcal{W}$ such that (2) holds for all $\boldsymbol{z}$, $\boldsymbol{b}_1$, and

$\boldsymbol{b}_2$. One prover will run the protocol for the AND problem on h, by Proposition 7. This involves uniformly and randomly choosing $\boldsymbol{z}$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$. The verifier also has to compute $h(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2)$, which involves finding $\widetilde{\mathcal{W}}(\boldsymbol{b}_1)$ and $\widetilde{\mathcal{W}}(\boldsymbol{b}_2)$. The provers will also provide these values of A. Of course, the verifier also needs to verify that the provers are not cheating in giving the values of $\widetilde{\mathcal{W}}$.[3] Thus, we only do the AND test with some probability; otherwise, we do either the *consistency test* or the *linearity test*. Intuitively, the consistency test is used to check that different provers are applying the same function $\widetilde{\mathcal{W}}$, while the linearity test is used to check that the function provided is linear.

Finally, we now formally state the protocol:

Protocol 8. This protocol involves a verifier V and three provers P_1, P_2, and P_3.

First, V randomly decides whether to do a consistency test, linearity test, or an AND test. In any kind of test, each prover will act as either a *lookup prover* or an *AND test prover*. V tells each prover which kind of prover they are, but not which kind of test he is doing.

- **Consistency Test.** V randomly chooses $\boldsymbol{x} \in \mathbb{F}^n$, and sends $\boldsymbol{x}$ to all three provers, treating them as lookup provers. Then, he receives the responses, say $y_1, y_2, y_3 \in \mathbb{F}$, and accepts if and only if $y_1 = y_2 = y_3$.

- **Linearity Test.** V randomly chooses $i \in \{1, \ldots, n\}$ and $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in \mathbb{F}^n$ such that for any $j \neq i$, $x_{1,j} = x_{2,j} = x_{3,j}$, but $x_{1,i}$, $x_{2,j}$, and $x_{3,j}$ are all distinct. (That is, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ are all distinct points on a line in the direction of the i^{th} direction.) V sends $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ to P_1, P_2, and P_3 and receives their responses, say $y_1, y_2, y_3 \in \mathbb{F}$. Then, V accepts if and only if
$$\frac{y_2 - y_1}{x_{1,i} - x_{2,i}} = \frac{y_3 - y_2}{x_{2,i} - x_{3,i}} = \frac{y_1 - y_3}{x_{3,i} - x_{1,i}}. \tag{4}$$
That is, V accepts if and only if it is possible to have $y_1 = \widetilde{\mathcal{W}}(\boldsymbol{x}_1)$, $y_2 = \widetilde{\mathcal{W}}(\boldsymbol{x}_2)$, and $y_3 = \widetilde{\mathcal{W}}(\boldsymbol{x}_3)$ for some multilinear function $\widetilde{\mathcal{W}}$.

- **AND Test.** V randomly chooses $\boldsymbol{z} \in \mathbb{F}^r$, and $\boldsymbol{b}_1, \boldsymbol{b}_2 \in \mathbb{F}^n$. Also, V randomly chooses one of the provers (here, we will say P_3) to be the AND test prover. Then, V sends $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ to V_1 and V_2, receiving back a_1 and a_2. Then V computes $f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, a_1, a_2)$. Then he runs the AND test with prover P_3, using $\boldsymbol{x} = (\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2)$ and $h(\boldsymbol{x}) = f(\boldsymbol{z}, \boldsymbol{b}_1, \boldsymbol{b}_2, a_1, a_2)$ in Proposition 7.

Theorem 9. *Protocol 8 has perfect completeness.*

Proof. If $\mathcal{W} : \{0,1\}^n \to \{0, 1, \alpha\}$ exists such that (2) holds for all $\boldsymbol{z}$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$, then extend $\mathcal{W}$ to a multilinear function $\widetilde{\mathcal{W}} : \mathbb{F}^n \to \mathbb{F}$. Each prover P_i acts as follows: if P_i is assigned to be a lookup prover, then he responds with $\widetilde{\mathcal{W}}(\boldsymbol{x})$ when given $\boldsymbol{x}$. If P_i is assigned to be an AND test prover, then P_i runs the AND test with the polynomial h as defined in (3). □

The technical result of Ito and Vidick [5] is to show

Theorem 10. *Protocol 8 has soundness* $1 - 1/\mathsf{poly}$*, even when the provers have shared entanglement.*

We will summarize the proof in their remaining sections.

[3] Babai, Fortnow, and Lund [1] discuss *oracle protocols*, where the verifier has access to an untrusted but fixed oracle, and they show easily, citing Fortnow, Rompel, and Sipser [4], that this means we can find a pure interactive protocol. It is more convenient for us to not to deal with this additional but thin layer of abstraction.

4 Measurement Strategies

To prove soundness, we note that Protocol 8 is a symmetric protocol. Thus by Proposition 1, we can assume that the provers' strategy is symmetric. Suppose the initial state of the quantum tape is $|\Psi\rangle$ with density matrix $\rho = |\Psi\rangle\langle\Psi|$. To describe a measurement, we will use a class of positive semidefinite matrices $\{A^a\}_a$, where a ranges over the set of outcomes of the measurement. If P_1 measures using measurement A, then the probability of measuring an outcome $\mathcal{W}$ is

$$\langle\Psi|(A^a \otimes I \otimes I)|\Psi\rangle = \mathrm{Tr}((A^a \otimes I \otimes I)\rho)$$

We will often denote this simply by $\mathrm{Tr}_\rho(A^a \otimes I \otimes I)$. Since the inital state is invariant under permutation of the registers, we have

$$\begin{aligned}\mathrm{Tr}_\rho(A^a \otimes I \otimes I) &= \mathrm{Tr}_\rho(I \otimes A^a \otimes I)\\ &= \mathrm{Tr}_\rho(I \otimes I \otimes A^a).\end{aligned}$$

Hence, we will often denote this value simply by $\mathrm{Tr}_\rho(A^a)$.

We will also assume that the strategies used by the provers are *projective*, that is, $(A^a)^2 = A^a$. However, non-projective strategies will be introduced in the course of the proof.

Now, if $\sum_a A^a = I$, then we say that A is a *full measurement*. Otherwise, if $\sum_a A^a \prec I$, then we say that A is a *sub-measurement*. Such measurements have some probability of failing. That is, with probability

$$1 - \sum_a \mathrm{Tr}_\rho(A^a)$$

the measurement returns some "fail" state. We imagine that a prover aborts the protocol whenever he measures a "fail" outcome, and that the verifier rejects.

Now, let us consider various measurement strategies that a lookup prover could use. The most natural choice is for the prover to have, for every possible $\boldsymbol{x} \in \mathbb{F}^n$, a measurement $\{A^a_x\}_a$ with outcomes $a \in \mathbb{F}$. Upon receiving the vector $\boldsymbol{x}$, the prover perform the measurement $\{A^a_x\}_a$ and returns the outcome. It is not hard to see that such a strategy is fully general. Any additional non-quantum computation that the prover may wish to do can be encoded in the measurement.

However, there are other, more restricting, measurement strategies that a prover could implement. We define such strategies now:

Definition 11. *A **measurement strategy of arity** k is a strategy for a lookup prover of the following form. For a vector $\boldsymbol{x} \in \mathbb{F}^n$, we let $\boldsymbol{x}_{\leq k} = (x_1, \ldots, x_k) \in \mathbb{F}^k$ and $\boldsymbol{x}_{>k} = (x_{k+1}, \ldots, x_n) \in \mathbb{F}^{n-k}$. For any possible value of $\boldsymbol{x}_{>k}$, the prover has a measurement $\{A^g_{\boldsymbol{x}_{>k}}\}_g$, with outcomes g in the set of multilinear functions $\mathbb{F}^k \to \mathbb{F}$. Upon receiving $\boldsymbol{x}$, the prover measures using $\{A^g_{\boldsymbol{x}_{>k}}\}_g$ to obtain a multilinear function g. The prover then sends $g(\boldsymbol{x}_{\leq k}) = g(x_1, \ldots, x_k)$ to the verifier.*

A measurement strategy of arity 0 is just the general measurement strategy described above. However, measurement strategies of higher arity have significantly more structure. In particular, a measurement strategy of arity n has the prover make a measurement completely independently of the value of $\boldsymbol{x}$ and returns $g(\boldsymbol{x})$, where g is some multilinear function. This is very similar to what we would hope an "honest" prover would do, since we want an honest prover to apply some multilinear $\widetilde{\mathcal{W}}$. Of course, the multilinear function g that the prover applies may still depend on the outcome of the measurement. However, we can still show:

Theorem 12. *Protocol 8 is sound if we assume that all lookup provers use arity-n measurement strategies.*

The proof will rely on a famous result, known as the Schwartz-Zippel Lemma [9, 12]:

Lemma 13 (Schwartz-Zippel Lemma)**.** *If $\mathbb{F}$ is a finite field, then a nonzero polynomial function $g : \mathbb{F}^n \to \mathbb{F}$ has at most $d|\mathbb{F}|^{n-1}$ zeroes, where d is the total degree of g.*

In particular, this means that if g is multilinear, it has at most $n|\mathbb{F}|^{n-1}$ zeroes.

Proof of Theorem 12. Suppose that the provers pass Protocol 8 with probability at least $1 - \varepsilon$, where ε is a small constant. Since the lookup provers' measurements do not depend at all on the values they receive, we can imagine that they make the measurements before $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are chosen. Suppose we run the AND test with (say) P_3 as the AND test prover. Then P_1 and P_2 each measure some multilinear functions g_1 and g_2, and they respond with $g_1(\boldsymbol{b}_1)$ and $g_2(\boldsymbol{b}_2)$.

We claim that we must have $g_1 \neq g_2$ with probability at most 6ε. Suppose otherwise. Then since there is a $1/3$ probability that V runs the consistency test, there would be at least a 2ε chance that they play the consistency test with $g_1 \neq g_2$. Furtheremore, by Schwartz-Zippel, if $g_1 \neq g_2$ there is at least a $1/2$ chance that $g_1(\boldsymbol{x}) \neq g_2(\boldsymbol{x})$ for randomly chosen $\boldsymbol{x}$ (recall our assumption that $|\mathbb{F}|$ is large). Then the provers would fail with probability at least ε, a contradiction.

Hence, with probability at least $1 - 6\varepsilon$, $g_1 = g_2$. Now, when running the AND test with P_3, there is at most a 6ε chance that the "oracle" to $\widetilde{\mathcal{W}}$ is corrupt. By the soundness of the AND test (Proposition 7), the entire protocol is sound if the "oracle" is not corrupt. Hence the protocol is sound is ε is small enough. (Since P_1 and P_2's measurements are independent of $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, P_3 gains no information about them, not even from the entanglement.) □

4.1 Replacing measurements

Theorem 12 is nice, but general strategies can have much less well-behaved measurement strategies than arity-n measurement strategies. The key idea is to show that we can replace a prover's arity-0 strategy with an arity-n strategy which does not change the outcome by much. First, we need a metric to determine if two measurements are "close".

Definition 14. *Given two multilinear functions $f : \mathbb{F}^k \to \mathbb{F}$ and $g : \mathbb{F}^\ell \to \mathbb{F}$, with $k \leq \ell$, we say that f and g are **consistent** for $\boldsymbol{x}$ if*

$$f = g|_{x_{k+1},\ldots,x_\ell}$$

where $g|_{x_{k+1},\ldots,x_\ell}$ is the restriction of g and is a multilinear function $\mathbb{F}^k \to \mathbb{F}$. Then given two measurement strategies P and Q of arities k and ℓ, we define

$$\begin{aligned}\mathit{CON}(P, Q) &:= \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{f,g \,:\, f=g|_{x_{k+1},\ldots,x_\ell}} \mathrm{Tr}_\rho\left(P^f_{\boldsymbol{x}_{>k}} \otimes Q^g_{\boldsymbol{x}_{>\ell}}\right) \\ &= \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \Pr\left[\begin{array}{c}\textit{Two provers given } \boldsymbol{x} \textit{ measuring with } P \\ \textit{and } Q \textit{ measure consistent } f \textit{ and } g\end{array}\right]\end{aligned}$$

and likewise

$$\begin{aligned}\mathsf{INC}(P,Q) &:= \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{f,g\,:\, f\neq g|_{x_{k+1},\ldots,x_\ell}} \mathrm{Tr}_\rho\left(P^f_{\boldsymbol{x}_{>k}} \otimes Q^g_{\boldsymbol{x}_{>\ell}}\right) \\ &= \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \Pr\left[\begin{array}{c}\textit{Two provers given } \boldsymbol{x} \textit{ measuring with } P \\ \textit{and } Q \textit{ measure inconsistent } f \textit{ and } g\end{array}\right]\end{aligned}$$

In the summations, f and g are multilinear functions $\mathbb{F}^k \to \mathbb{F}$ and $\mathbb{F}^\ell \to \mathbb{F}$, respectively.

Note that if P and Q are full measurement strategies, then $\mathsf{CON}(P,Q)+\mathsf{INC}(P,Q) = 1$, because in that case, the provers will always successfully measure *some* f and g, and then f and g are either consistent or inconsistent.

Also note that if A is an arity-0 measurement, then $\mathsf{CON}(A,A)$ is the probability of two provers passing the consistency test, since for arity-0 measurements, for f and g to be consistent it just means that $f(\boldsymbol{x}) = g(\boldsymbol{x})$.

One important fact is that if a measurement P is consistent with itself (i.e., $\mathsf{INC}(P,P)$ is small) then applying P to one register should be approximately the same as applying it to a separate register. This is intuitively obvious: if P is consistent with itself and we measure one register using P, then measuring other registers should give the same result most of the time. Thus each other register should be close to the state it would be in if P was applied directly to it. Here is the formal statement:

Lemma 15. *Let $\{Z^a_{\boldsymbol{x}_{>k}}\}_a$ be such that $\mathrm{E}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} Z^a_{\boldsymbol{x}_{>k}} (Z^a_{\boldsymbol{x}_{>k}})^\dagger \preceq I$, and let P be a (sub-) measurement strategy of arity k. Let $P_{\boldsymbol{x}_{>k}} = \sum_{a\in\mathbb{F}^n} P^a_{\boldsymbol{x}_{>k}}$ (so $P_{\boldsymbol{x}_{>k}} = I$ if P is a full measurement strategy). Then*

$$\left| \mathop{\mathrm{E}}_{x\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho(Z^a_{\boldsymbol{x}_{>k}} P^a_{\boldsymbol{x}_{>k}} \otimes P_{\boldsymbol{x}_{>k}}) - \mathop{\mathrm{E}}_{x\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho(Z^a_{\boldsymbol{x}_{>k}} P_{\boldsymbol{x}_{>k}} \otimes P^a_{\boldsymbol{x}_{>k}}) \right| \leq \sqrt{\mathsf{INC}(P,P)}$$

Proof. This follows from an application of Cauchy-Schwartz and the definition of $\mathsf{INC}(P,P)$. For full details, see Appendix B, Lemma 40 of [5]. □

Now, we come to the brilliant fact from Ito and Vidick [5].

Theorem 16. *For sufficiently small ε,[4] suppose the provers have an arity-0 measurement strategy $\{A^a_x\}_a$ which is full and projective, and $\{A^a_{\boldsymbol{x}}\}_a$ passes the linearity and consistency tests with probability $1-\varepsilon$. Then there exists an arity-n (sub-)measurement strategy $\{V^g\}_g$ such that $\mathsf{CON}(V,A) \geq 1-\varepsilon^{O(1)}$.*

This lets us show:

Proof of Theorem 10. Suppose the provers are using the 0-arity strategy A. Let V be as in the statement of Theorem 16. We will prove that we can replace the lookup prover's strategy by V without affecting the outcome by much. After replacing all the strategies, we can apply Theorem 12 to complete the proof.

[4] By "sufficiently small" we mean $1/\mathsf{poly}(n)$. The reader interested in the details of error terms should, of course, consult Ito and Vidick [5].

Define

$$V_{\boldsymbol{x}}^a := \sum_{g\,:\,g(\boldsymbol{x})=a} V^g. \tag{5}$$

This is useful because if a prover is using strategy V, the probability that he responds with a when given $\boldsymbol{x}$ is just $\mathrm{Tr}_\rho(V_{\boldsymbol{x}}^a)$. Now, we will show that

$$\mathop{\mathrm{E}}_{x\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(\left(A_{\boldsymbol{x}}^a - \sqrt{V_{\boldsymbol{x}}^a}\right)^2\right) = \varepsilon^{O(1)}. \tag{6}$$

To see this, first expand

$$\mathop{\mathrm{E}}_{x\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(\left(A_{\boldsymbol{x}}^a - \sqrt{V_{\boldsymbol{x}}^a}\right)^2\right) = \mathop{\mathrm{E}}_{x\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left((A_{\boldsymbol{x}}^a)^2 + V_{\boldsymbol{x}}^a - 2A_{\boldsymbol{x}}^a\sqrt{V_{\boldsymbol{x}}^a}\right)$$

We have $\sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho((A_{\boldsymbol{x}}^a)^2) = \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho(A_{\boldsymbol{x}}^a) \le 1$ and $\sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho(V_{\boldsymbol{x}}^a) \le 1$, so

$$\mathop{\mathrm{E}}_{x\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(\left(A_{\boldsymbol{x}}^a - \sqrt{V_{\boldsymbol{x}}^a}\right)^2\right) \le 2 - 2\mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(A_{\boldsymbol{x}}^a\sqrt{V_{\boldsymbol{x}}^a}\right). \tag{7}$$

By Lemma 15, and using that A passes the consistency test with probability ε, we get

$$\begin{aligned}\left|\mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(A_{\boldsymbol{x}}^a\sqrt{V_{\boldsymbol{x}}^a}\right) - \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(A_{\boldsymbol{x}}^a \otimes \sqrt{V_{\boldsymbol{x}}^a}\right)\right| &\le \sqrt{\mathsf{INC}(A,A)} \\ &\le \varepsilon^{1/2}.\end{aligned} \tag{8}$$

Now, $\sqrt{V_{\boldsymbol{x}}^a} \succeq V_{\boldsymbol{x}}^a$ since $V_{\boldsymbol{x}}^a \preceq I$. Then by the definition of CON, we get

$$\begin{aligned}\mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(A_{\boldsymbol{x}}^a \otimes \sqrt{V_{\boldsymbol{x}}^a}\right) &\ge \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}_\rho\left(A_{\boldsymbol{x}}^a \otimes V_{\boldsymbol{x}}^a\right) \\ &= \mathsf{CON}(V,A) \\ &\ge 1 - \varepsilon^{O(1)}.\end{aligned} \tag{9}$$

Combining (7), (8), and (9) proves (6). Now, we show that if a prover P_i uses strategy A, we can replace P_i with a prover $\widehat{P}_i$ which uses V without changing the protocol's acceptance probability by much.

Suppose that P_i's register in the quantum tape is $\mathcal{P}_i$. Imagine two more registers $\mathcal{A}$ and $\mathcal{B}$. We think of them as "input" and "output" registers. $\mathcal{A}$ will hold the value $\boldsymbol{x}$, and $\mathcal{B}$ will hold the value $a \in \mathbb{F}$ that P_i returns. Let $\mathcal{D}$ denote everything else, that is, it represents the entire global state except $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{P}_i$. Then we can imagine the global state, before measurement and even before choice of $\boldsymbol{x}$, for prover P_i:

$$\sigma := \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|_{\mathcal{A}} \otimes \sum_{a\in\mathbb{F}} \left(|a\rangle\langle a|_{\mathcal{B}} \otimes (A_{\boldsymbol{x}}^a \otimes I)\sigma_{\mathcal{P}_i,\mathcal{D}}^{\boldsymbol{x}}(A_{\boldsymbol{x}}^a \otimes I)^\dagger\right)$$

where $\sigma^{\boldsymbol{x}}_{\mathcal{P}_i,\mathcal{D}}$ is the density matrix describing the state of $\mathcal{P}_i$ and $\mathcal{D}$ conditioned on $\boldsymbol{x}$. Typically, we imagine that V chooses $\boldsymbol{x}$ in some fashion, along with other parameters (e.g., type of test), sends it to P_i; P_i makes a measurement, and then sends some value a back. However, this entire interaction can be thought of as measuring $\mathcal{A}$ and $\mathcal{B}$. The "other parameters" are encoded in $\mathcal{D}$.

If we use the prover $\widehat{P}_i$ instead, we get the global state to be

$$\sigma' := \mathop{\mathrm{E}}_{\boldsymbol{x}\in\mathbb{F}^n} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|_{\mathcal{A}} \otimes \sum_{a\in\mathbb{F}} \left(|a\rangle\langle a|_{\mathcal{B}} \otimes (\sqrt{V^a_{\boldsymbol{x}}} \otimes I)\sigma^{\boldsymbol{x}}_{\mathcal{P}_i,\mathcal{D}}(\sqrt{V^a_{\boldsymbol{x}}} \otimes I)^\dagger \right)$$

Now, the difference in acceptance probabilities must be at most the 1-norm between the two states

$$\begin{aligned}&\|\sigma - \sigma'\| \\ &= \left\| \mathop{\mathrm{E}}_{x\in\mathbb{F}^n} |\boldsymbol{x}\rangle\langle\boldsymbol{x}|_{\mathcal{A}} \otimes \sum_{a\in\mathbb{F}} \left(|a\rangle\langle a|_{\mathcal{B}} \otimes \left((A^a_{\boldsymbol{x}} \otimes I)\sigma^{\boldsymbol{x}}_{\mathcal{P}_i,\mathcal{D}}(A^a_{\boldsymbol{x}} \otimes I)^\dagger - (\sqrt{V^a_{\boldsymbol{x}}} \otimes I)\sigma^{\boldsymbol{x}}_{\mathcal{P}_i,\mathcal{D}}(\sqrt{V^a_{\boldsymbol{x}}} \otimes I)^\dagger \right) \right) \right\|\end{aligned}$$

By some matrix inequalities (see Claim 36 in Appendix A and the proof of Claim 12 in [5]), this gives

$$\|\sigma - \sigma'\| \le 2\sqrt{\mathop{\mathrm{E}}_{x\in\mathbb{F}^n} \sum_{a\in\mathbb{F}} \mathrm{Tr}\left(\left(A^a_{\boldsymbol{x}} - \sqrt{V^a_{\boldsymbol{x}}} \right)^2 \rho \right)}$$

where ρ is the density matrix of just the $\mathcal{P}_i$ register. Finally, this value is $\varepsilon^{O(1)}$ by (6).

Thus replacing A by V as P_i's strategy changes the acceptance probability of the entire by only a small amount. Thus we can assume that all provers are using arity-n measurement strategies, and then Theorem 12, completes the proof. □

5 From arity-0 to arity-n

As we saw, Theorem 16 is an excellent hammer which lets us fairly easily (modulo some matrix computation) prove the soundness of Protocol 8 by reducing to Theorem 12. But how do we prove Theorem 16? This turns out still to be quite a challenge.

Recall that we are given some arity-0 measurement strategy A, and we want to show that we can find an arity-n measurement. Recall from Definition 11 that, in general, a measurement strategy of arity k is a strategy whose measurement depends only on the last $n-k$ coordinates of $\boldsymbol{x}$, and which applies a multilinear function of the first k coordinates of $\boldsymbol{x}$.

The main idea for constructing V is to inductively construct a sequence of measurement strategies $A = V_0, V_1, V_2, \ldots, V_n = V$, where V_i has arity i.

Again, suppose that A passes the consistency and linearity tests with probability at least $1-\varepsilon$. When constructing V_{k+1} from V_k, we will obviously want $\mathsf{INC}(V_{i+1}, A)$ to be low. But this is not enough: the inductive procedure creates sub-measurements, not full measurements, so we have to keep track of how much of the measurement we lose. (It is trivial to get a low value of INC if we make the measurement empty, but we want to have a high CON value as well.)

We have to prove quantum analogues of two lemmas from [1]: the *Pasting Lemma* and the *Self-improvement Lemma*. The Pasting Lemma does the most visible work; it takes a sub-measurement V_k and constructs the sub-measurement V_{k+1}. However, we cannot just apply the Pasting Lemma

iteratively, because the constructed V_{k+1} has $\mathsf{INC}(V_{k+1}, A)$ too large for us to apply the Pasting Lemma again.[5] This is where the Self-improvement Lemma comes in. Given V_{k+1}, the Self-improvement Lemma constructs a measurement strategy T_{k+1}, also of arity $k+1$, which has even lower inconsistency, low enough that the Pasting Lemma can be applied.[6] Then we can construct the sequence

$$V_0, T_0, V_1, T_1, \ldots, V_{n-1}, T_{n-1}, V_n$$

to prove Theorem 16. The cost of the Self-improvement Lemma is that the measurements T_k are less full than their respective V_k measruements. This is the reason that sub-measurements are introduced in the first place.

The proofs of both lemmas are quite technical, and we do not have the space to go into details. However, the interested reader is encouraged to see the classical versions in [1] (Lemmas 5.10 and 5.11).

Also, the reader may have noticed that, in proving soundness, we have not yet applied the fact that the provers' strategy A passes the linearity test. This fact is used critically in the proofs of the Self-improvement lemma and Pasting Lemma. Also relevant is Claim 15 from [5]. This is essentially a simplified version of the Pasting Lemma. It only applies to full measurements but it serves as useful motivation for the proof of the Pasting Lemma. Furthermore, Chapter 2 of [11] is a good introduction to the subject although it deals only with linear (not multilinear) functions over $\mathbb{F}_2^n$. It also makes nice use of the same intuition behind Lemma 15 (namely that a measurement strategy with high self-consistency has nearly the same effect on the global state no matter which register it is applied to).

6 Conclusion

In this survey, we gave a high level overview of the proof that $\mathsf{NEXP} \subseteq \mathsf{MIP}^*$ by Ito and Vidick [5]. The result is fascinating and is a large step towards our understanding of MIP^*. However, there is still much work to be done. No reasonable upper bounds on MIP^* are known, and in fact, it is not even known if it is computable (since there is no upper bound on the amount of entanglement that the provers can use).

Another question is: how many provers are needed? Since [2] showed that two provers always suffices for MIP, we might ask the same for MIP^*. Do three provers suffice for MIP^*? Can we improve the NEXP protocol to require only two provers?

References

[1] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3-40, 1991.

[2] M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson. Multi-prover interactive proofs: How to remove the intractability assumptions. *Proc. 20th Ann. ACM Symp. Theory of Computing*, 113-131, 1988.

[5] Again, see [5] for more explicit details about these terms.

[6] The intertwining of the Self-improvement and Pasting lemmas is in fact slightly more complex; the Pasting Lemma relies on the specific form of the measurements constructed by the Self-improvement Lemma.

[3] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. *Proc. of IEEE Complexity*, 236-249, 2004.

[4] L. Fortnow, J. Rompel, and M. Sipser. On the power of multi-prover interactive protocols. *Proc. 3rd Structure in Complexity Theory Conf.*, 156-161, 1988.

[5] T. Ito and T. Vidick. A multi-prover interactive proof for NEXP sound against entangled provers. *FOCS*, to appear, 2012.

[6] H. Kobayashi and K. Matsumoto. Quantum multi-prover interactive proof systems with limited prior entanglement. *Journal of Computer and System Sciences*, 66(3):429-450, 2003.

[7] C. Lund, L. Fortnow, H. Karloff, and N. Nisan. Algebraic methods for interactive proof systems. *Proc. 31st Ann. IEEE Symp. Foundations of Comp. Sci.*, 1-10, 1990.

[8] B. Reichardt, F. Unger, and U. Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games. Technical report, arXiv:1209.0448v1, 2012.

[9] J. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):707-717, 1980.

[10] A. Shamir. IP = PSPACE. *Proc. 31st Ann. IEEE Symp. Foundations of Comp. Sci.*, 11-15, 1990.

[11] T. Vidick. The Complexity of Entangled Games. Thesis, 2011.

[12] R. Zippel. Probabilistic algorithms for sparse polynomials. *Symbolic and Algebraic Computation: An Internatial Symposium on Symbolic and Algebraic Manipulation (EUROSM)*, volume 72 of *Lecture Notes in Computer Science*, 216-226, 1979.