Efficient protocols for generating bipartite classical distributions and quantum states

Rahul Jain* Yaoyun Shi† Zhaohui Wei‡ Shengyu Zhang§

Abstract

We investigate the fundamental problem of generating bipartite classical distributions or quantum states. By designing efficient communication protocols and proving their optimality, we establish a number of intriguing connections to fundamental measures in optimization, convex geometry, and information theory.

1. To generate a classical distribution $P(x, y)$, we tightly characterize the minimum amount of quantum communication needed by the *psd-rank* of P (as a matrix), a measure recently proposed by Fiorini, Massar, Pokutta, Tiwary and de Wolf (*Proceedings of the 44th ACM Symposium on Theory of Computing*, pages 95-106, 2012) in studies of the minimum size of extended formulations of optimization problems such as TSP. This echos the previous characterization for the optimal classical communication cost by the *nonnegative rank* of P. The result is obtained via investigating the more general case of bipartite *quantum* state generation and designing an optimal protocol for it.
2. When an approximation of ϵ is allowed to generate a distribution $(X, Y) \sim P$, we present a classical protocol of the communication cost $O((C(X, Y) + 1)/\epsilon)$, where $C(X, Y)$ is *common information*, a well-studied measure in information theory introduced by Wyner (*IEEE Transactions on Information Theory*, 21(2):163-179, 1975). This also links nonnegative rank and common information, two seemingly unrelated quantities in different fields.
3. For approximately generating a quantum pure state $|\psi\rangle$, we completely characterize the minimum cost by a corresponding approximate rank, closing a possibly exponential gap left in Ambainis, Schulman, Ta-Shma, Vazirani and Wigderson (*SIAM Journal on Computing*, 32(6):1570-1585, 2003).

1 Introduction

Shared randomness and quantum entanglement among spatially separated parties are important resources for various distributed information processing tasks. This motivates us to study the minimum cost for establishing a specified shared randomness or quantum entanglement. For the case of shared randomness, we are also interested in contrasting the efficiency of quantum protocols with that of classical protocols. Such comparisons will deepen our understanding on the power and the limitation of quantum information processing, and complement numerous studies for the same purpose but on different perspectives such as computational complexity, cryptography, and nonlocal games. We will focus on the bipartite case in this study. For simplicity, we also use the binary alphabet for classical distributions and qubit space for quantum states.

Let us be more specific on the quantities involved. We use the convention of defining the *size* of a bipartite distribution as the half of the total number of bits. Similarly, the *size* of a bipartite quantum state is the half of the total number of qubits. Consider the situation in which two parties, Alice and Bob, aim to output random variables X and Y, respectively, so that (X, Y) is distributed according to a target joint distribution P. In general, when P is not a product distribution, Alice and Bob can share a seed correlation (X', Y') and each apply a local operation on their own part. The minimum size of this seed distribution is the *randomized correlation complexity*

*Department of Computer Science and Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117542. Email: rahul@comp.nus.edu.sg

†Department of Electrical Engineering and Computer Science, 2260 Hayward Street, University of Michigan, Ann Arbor, Michigan, 48109-2121, USA. Email: shiyy@eecs.umich.edu

‡Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117542. Email: cqtwz@nus.edu.sg

§Department of Computer Science and Engineering and The Institute of Theoretical Computer Science and Communications, The Chinese University of Hong Kong. Email: syzhang@cse.cuhk.edu.hk

of P, denoted $\mathsf{R}(P)$. The two parties can also share a *quantum* state σ as a seed, on which local quantum operations are applied to generate (X, Y). The minimum size of the quantum seed σ is the *quantum correlation complexity*, denoted $\mathsf{Q}(P)$. More generally, the target can be a bipartite quantum state ρ [1], and the quantum correlation complexity of generating ρ, denoted $\mathsf{Q}(\rho)$, is again defined as the minimum size of the quantum seed σ. Since Alice and Bob can always just share ρ itself, $\mathsf{Q}(\rho)$ is at most the size of ρ, so the correlation complexity is a sublinear complexity measure.

Instead of sharing seed states, Alice and Bob can also generate a correlation by communication. The *quantum communication complexity* of ρ, denoted $\mathsf{QComm}(\rho)$, is defined as the minimum number of qubits exchanged between Alice and Bob, initially sharing nothing, to produce ρ at the end of the protocol. Again, when the target state ρ is a classical distribution P, one can also define the *randomized communication complexity* of P, denoted $\mathsf{RComm}(P)$, as the minimum number of bits exchanged to produce P. It turns out that for any state ρ, the correlation complexity and the communication complexity are always the same, namely $\mathsf{QComm}(\rho) = \mathsf{Q}(\rho)$ (and $\mathsf{RComm}(P) = \mathsf{R}(P)$ when ρ is a classical distribution P) [18]. Therefore in the following, we will ignore the difference between correlation and communication complexity, and just use notation Q and R to denote the quantity in quantum and classical settings.[2]

Throughout the paper, the target distribution P is over $\mathcal{X} \times \mathcal{Y}$, where set $\mathcal{X}$ is at Alice's side, and set $\mathcal{Y}$ is at Bob's. We will use x, x' to range over $\mathcal{X}$, and y, y' over $\mathcal{Y}$. For quantum states, Alice's output space is $\mathcal{H}_A = \mathbb{C}^{\mathcal{X}}$, and $|x\rangle$, $|x'\rangle$ range over $\mathcal{H}_A$; similarly for $\mathcal{H}_B = \mathbb{C}^{\mathcal{Y}}$ and $|y\rangle$, $|y'\rangle$ over $\mathcal{H}_B$. We usually identify a distribution P with the matrix $P = [P(x, y)]_{x,y}$ (whose (x, y)-th entry is $P(x, y)$).

We summarize below our main results in three directions, in the context of previous related works.

Distribution generation. Correlation complexity and communication complexity in the above model were proposed and studied in [18], while the communication complexity of generating certain special distributions (arising from Boolean functions on distributed inputs) was considered much earlier in [1]. In [18] a distribution P of size n is exhibited with $\mathsf{R}(P) \geq \log_2(n)$ and $\mathsf{Q}(P) = 1$. Later developments [11, 8] [3] gave a distribution P of size n with $\mathsf{R}(P) \geq \Omega(n)$ and $\mathsf{Q}(P) = O(\log(n))$. These results can be viewed as complexity-theoretic versions of results on nonlocal games, a central area of study in quantum mechanics and quantum information that contrasts quantum and classical input-output correlations (see, e.g. [5], for a survey). Note that a crucial difference between the non-local game model and in ours is that in non-local games, the two parties are given *private* (and random) inputs, which are necessary to differentiate the power of classical and quantum shared resource in those settings. In comparison, there is no private inputs in the model that we adopt, which is thus simpler and more basic. So it is somewhat surprising that it can still separate the power of shared classical randomness and quantum entanglement.

These separations [18, 11, 8] are obtained as follows. First, $\mathsf{R}(P)$ is fully characterized as $\lceil \log_2 \mathtt{rank}_+(P) \rceil$, where $\mathtt{rank}_+(P)$ is the nonnegative rank, a measure in linear algebra with numerous applications in combinatorial optimization [17], nondeterministic communication complexity [12], algebraic complexity theory [14], and many other fields [6]. Concretely, $\mathtt{rank}_+(P)$ is defined as the minimum number r such that P can be decomposed as the summation of r nonnegative matrices of rank 1. Showing large $\mathsf{R}(P)$ thus boils down to lower bounding nonnegative rank, which are proven either via merely looking at the nonzero locations [4, 18], or via nondeterministic communication complexity [11, 8]. Second, one needs to upper bound $\mathsf{Q}(P)$. It is showed in [18] that

$$\frac{1}{4} \log_2 \mathtt{rank}(P) \leq \mathsf{Q}(P) \leq \min_{M:\ M \circ \bar{M} = P} \log_2 \mathtt{rank}(M), \tag{1.1}$$

where $\circ$ is the Hadamard product (*i.e.* entry-wise product) and $\bar{M}$ is the complex conjugate of M. In [8], en route to discovering an exponential lower bound in extended formulation of TSP, the authors also derived the same upper bound as above though for a different communication complexity measure Q'. This was not a coincidence as we will later show that the measure Q' is the same as the $\mathsf{Q}(P)$ above. Also see more work related to the upper bound in convex geometry [9, 3].

[1] As always, one can think of classical distributions as a special class of quantum states.

[2] One can also consider an intermediate model where Alice and Bob use both shared state and communication, and the measure is the size of the shared state plus the amount of the communication. It is not hard to see that this complexity is also the same as the correlation and communication complexities.

[3] The two papers independently obtained the same distribution and used the same upper bound, though in [8] the purpose of constructing the distribution was not exactly for separating classical and quantum correlation complexities.

Despite the application of the upper bound in Eq.(1.1), it is not clear how tight the bound is. Indeed, a full characterization of $\mathsf{Q}(P)$ was called for in [18] as an open question. In this paper, we answer this by showing a tight characterization in terms of *psd-rank* of P, a concept recently proposed in [8] by Fiorini, Massar, Pokutta, Tiwary and de Wolf. For an entry-wise nonnegative matrix P, its psd-rank, denoted $\mathtt{rank}_{\mathtt{psd}}(P)$, is the minimum r such that there are $r \times r$ positive semi-definite matrices C_x, D_y, satisfying that $P(x,y) = \mathtt{tr}(C_x D_y)$, for all x and y. We show the following result.

THEOREM 1.1. *For any bipartite distribution $P(x,y)$,*

$$\mathsf{Q}(P) = \lceil \log_2 \mathtt{rank}_{\mathtt{psd}}(P) \rceil. \tag{1.2}$$

Along with the foregoing characterization $\mathsf{R}(P) = \lceil \log_2 \mathtt{rank}_+(P) \rceil$, it is interesting to see that the classical complexity for generating a bipartite distribution P is all about its nonnegative rank, and in comparison, its quantum complexity is all about the psd-rank.

The above theorem is shown via studying a more general task of generating a bipartite *quantum* state ρ, for which we also show the following characterization of $\mathsf{Q}(\rho)$.

THEOREM 1.2. *Let ρ be a quantum state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $|x\rangle$, $|x'\rangle$ range over the computational basis states for $\mathcal{H}_A$, and $|y\rangle$, $|y'\rangle$ for $\mathcal{H}_B$. Then $\mathsf{Q}(\rho) = \lceil \log_2 r \rceil$, where r is the minimum number such that there exist matrices $\{A_x\}$ and $\{B_y\}$, each with r columns, and*

$$\rho = \sum_{x,x';y,y'} |x\rangle\langle x'| \otimes |y\rangle\langle y'| \cdot \mathtt{tr}\Big((A_{x'}^\dagger A_x)^T (B_{y'}^\dagger B_y)\Big).$$

Distribution approximation. While in the aforementioned connections to optimization, the quantity that matters corresponds to *exact* generation of a distribution, if we go back to the original motivation of distributively sampling, it is natural to relax the requirement by allowing a small deviation. After all, if two distributions are close, then they are hard to distinguish information theoretically. Next, we give a protocol approximating a classical distribution P, with the cost bounded by the *common information* introduced by Wyner [16].

DEFINITION 1.1. (WYNER, [16]) *The* common information *$C(X:Y)$ between two random variables X and Y is the minimum value of $I(XY:W)$, the mutual information between XY and W, where the minimum is over all random variables W conditioned on which X and Y are independent.*

For $(X,Y) \sim P$, it is not hard to see that $\mathsf{R}(P) \geq C(X:Y)$. Though the bound can be loose, we show that if we allow a small deviation of δ in sampling P, then the complexity is roughly *upper bounded* by $C(X:Y)/\delta$. The approximate version of correlation complexity is defined as follows.

DEFINITION 1.2. *Let $\epsilon > 0$. Let p be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. The* ϵ-correlation complexity *is defined as $\mathsf{R}_\epsilon(p) \stackrel{\text{def}}{=} \min\{\mathsf{R}(q) : q$ is a probability distribution on $\mathcal{X} \times \mathcal{Y}$ with $\|p-q\|_1 \leq \epsilon\}$, where $\|p-q\|_1$ is the total variation distance between the two distributions and for convenience we ignore the factor* $1/2$.

We show the following result which can be viewed as a single-shot version of Wyner's result in the asymptotic setting that $\lim_{n\to\infty} \mathsf{R}(X^n, Y^n)/n = C(X:Y)$ [16].

THEOREM 1.3. *Let $\delta > 0$. Let $(X,Y) \sim P$ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Then,*

$$\mathsf{R}_{8\delta}(P) \leq \frac{1}{\delta}(C(X:Y)+1) + O(\log(1/\delta)).$$

Combining the above theorem with the characterization of $\mathsf{R}(P)$, we have the following corollary, which relates an approximation version of nonnegative rank and common information, defined naturally as $\mathtt{rank}_+^{(\delta)} = \min\{\mathtt{rank}_+(P') : \|P'-P\|_1 \leq \delta\}$, and $C_\delta(X:Y) = \min\{C(X':Y') : \|(X',Y')-(X,Y)\|_1 \leq \delta\}$.

COROLLARY 1.1. *Suppose that $\delta > 0$ and $(X,Y) \sim P$, then*

$$C_\delta(X:Y) \leq \log_2 \mathtt{rank}_+^{(\delta)}(P) \leq O((C(X:Y)+1)/\delta).$$

Pure state approximation. Apart from classical distributions, we also consider the other end of the spectrum of general quantum states, namely quantum *pure* states. While the correlation complexity of exactly generating a pure state $|\psi\rangle$ is easily seen to be the logarithm of the Schmidt-rank of $|\psi\rangle$, its approximation version is more complicated. In [1], Ambainis, Schulman, Ta-Shma, Vazirani and Wigderson considered to approximate a pure state $|\psi\rangle$ by another *pure* state. Define the approximate communication complexity by $\mathsf{Q}_\epsilon^{pure}(|\psi\rangle) = \min\{\mathsf{Q}(|\psi'\rangle) : |\langle\psi|\psi'\rangle| \geq 1-\epsilon\}$. They showed that for any pure state $|\psi\rangle = \sum_{x,y} a_{x,y}|x\rangle \otimes |y\rangle$,

$$\lceil \log_2 \mathtt{rank}_{2\epsilon}(A) \rceil \leq \mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|) \leq \lceil \log_2 \mathtt{rank}_\epsilon(A) \rceil.$$

Here $A = [a_{x,y}]$, and $\mathtt{rank}_\epsilon(A) \stackrel{\text{def}}{=} \min\{\mathtt{rank}(B) : \|A-B\|_2^2 \leq \epsilon\}$. Using Lemma 2.1 (as mentioned in the next section), one can easily construct a state $|\psi\rangle \in \mathbb{C}^N \otimes \mathbb{C}^N$ such that $\mathtt{rank}_{2\epsilon}(A) = 1$ but $\mathtt{rank}_\epsilon(A) = N/2$, making the above two bounds arbitrarily far from each other.

In this paper, we consider approximate versions of $\mathsf{Q}(\rho)$ for general quantum states as follows. We first extend the above definition by allowing to use a *mixed* state to approximate a target state. (In what follows, $\mathsf{F}(\rho, \rho')$ represents the *fidelity* between ρ and ρ'. See the next section for formal definitions.)

Definition 1.3. *For $\epsilon > 0$ and ρ a bipartite quantum state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Define*

$$\mathsf{Q}_\epsilon(\rho) \stackrel{\text{def}}{=} \min\{\mathsf{Q}(\rho') : \mathsf{F}(\rho,\rho') \geq 1-\epsilon;\ \rho' \in \mathcal{H}_A \otimes \mathcal{H}_B\}.$$

$$\mathsf{Q}_\epsilon^{pure}(\rho) \stackrel{\text{def}}{=} \min\{\mathsf{Q}(|\phi\rangle\langle\phi|) : \mathsf{F}(\rho,|\phi\rangle\langle\phi|) \geq 1-\epsilon;\ |\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B\}.$$

That is, $\mathsf{Q}_\epsilon^{pure}(\rho)$ and $\mathsf{Q}_\epsilon(\rho)$ are the complexities to generate an approximation of the target state ρ by a pure and mixed state, respectively. We show the following tight characterization.

Theorem 1.4. *Let $\epsilon > 0$. Let $|x\rangle$ and $|y\rangle$ range over the computational basis states for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Let $|\psi\rangle = \sum_{x,y} a_{x,y}|x\rangle \otimes |y\rangle$, and matrix $A = [a_{x,y}]$. Then*

$$\mathsf{Q}_\epsilon(|\psi\rangle\langle\psi|) = \mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|) = \lceil \log_2 \mathtt{rank}_{2\epsilon-\epsilon^2}(A) \rceil.$$

Note that our result not only improves the bounds in [1] to the optimal, but also shows that allowing a *mixed* state to approximate a pure state $|\psi\rangle$ does not help, for any $|\psi\rangle$ and any ϵ.

We conclude this section by relating our work to several additional works. The paper [10] studies the communication complexity for generating a correlation (X, Y), where the complexity is measured in expectation as follows. Suppose Alice samples $x \leftarrow X$ and tries to let Bob sample from $Y|(X = x)$, [10] asks the minimum *expected* communication needed (where the expectation is over the randomness of protocol as well as the initial sample $x \leftarrow X$). For comparison, the model here measures the worst-case cost, over all randomness in the protocol. And also note that protocols in [10] uses a large amount of public coins, which is exactly the resource we hope to save.

The complexity for generating correlation in the asymptotic setting has a long history in information theory, dating back to Wyner's classic paper [16]. Asymptotic complexity in the corresponding quantum case was studied by Winter [15].

When introduced in [8], psd-rank is also used to characterize one-way quantum communication complexity of computing $P(x, y)$ in expectation, where Bob uses a POVM $\{E_\theta^y\}$ on a message ρ_x sent by Alice *s.t.* the expected outcome θ is equal to $P(x, y)$. Unfortunately, we do not know any direct argument for why the task and our distributive sampling have the same complexity.

After finishing this work, we learned that Lemma 5.1 was known before (for example Fact 2.1 of [2]), but previous results seem to be concerned with approximating a pure state by another *pure* state. Note that Theorem 1.4 in the present paper considers approximations also by mixed states.

2 Preliminaries

Matrix theory. For a natural number n we let $[n]$ represent the set $\{1, 2, \ldots, n\}$. We sometimes write $A = [a_{x,y}]$ to mean that A is a matrix with the (x, y)-th entry being $a_{x,y}$. For a matrix A, we let A^T represent the transpose of A, A^* or $\bar{A}$ represent the conjugate of A and $A^\dagger$ represent the conjugate transpose of A. An operator A is said to be *Hermitian* if $A^\dagger = A$. A Hermitian operator A is said to be *positive semi-definite* (PSD) if all its eigenvalues are non-negative. For any vectors $|v_1\rangle, \ldots, |v_r\rangle$ in $\mathbb{C}^n$, the $r \times r$ matrix M defined by $M(i,j) \stackrel{\text{def}}{=} \langle v_i|v_j\rangle$ is positive semi-definite.

If A is positive semi-definite then so is $A^T = A^*$. We let $\sigma_1(A) \geq \cdots \geq \sigma_n(A) \geq 0$ denote singular values of A. The rank of A, denoted $\mathtt{rank}(A)$, is defined to be the number of non-zero singular values of A. The *Frobenius norm* of A is $\|A\|_2 = \sqrt{\sum_i \sigma_i(A)^2}$ and the *trace norm* is $\|A\|_1 = \sum_i \sigma_i(A)$. For $\epsilon > 0$, define $\mathtt{rank}_\epsilon(A) = \min\{\mathtt{rank}(B) : \|A-B\|_2^2 \leq \epsilon\}$. The following well-known result says that the best way to approximate A (under the Frobenius norm) with the least rank is by taking the large singular values part.

Lemma 2.1. (Eckart-Young, [7]) *Let $\|A\|_2 = 1$ and $\epsilon > 0$. Then, $\mathtt{rank}_\epsilon(A) =$ the minimum k such that $\sum_{i=1}^k \sigma_i(A)^2 \geq 1 - \epsilon$.*

The following definition of psd-rank of a matrix was proposed in [8].

Definition 2.1. *For a matrix $P \in \mathbb{R}_+^{n\times m}$, its psd-rank, denoted $\mathtt{rank}_{\mathtt{psd}}(P)$, is the minimum number r such that there are PSD matrices $C_x, D_y \in \mathbb{C}^{r\times r}$ with $\mathtt{tr}(C_x D_y) = P(x,y)$, $\forall x \in [n], y \in [m]$.*

Quantum information. A quantum state ρ in Hilbert space $\mathcal{H}$, denoted $\rho \in \mathcal{H}$, is a trace one positive semi-definite operator acting on $\mathcal{H}$. The *size* of a state ρ is defined to be half the number of qubits of ρ. [4] A quantum state ρ is called *pure* if it is rank one, namely $\rho = |\psi\rangle\langle\psi|$ for some vector $|\psi\rangle$ of unit ℓ_2 norm; in this case, we often identify ρ with $|\psi\rangle$. For quantum states ρ and σ, their fidelity is defined as $\mathsf{F}(\rho,\sigma) \stackrel{\text{def}}{=} \mathtt{tr}(\sqrt{\sigma^{1/2}\rho\sigma^{1/2}})$. For $\rho, |\psi\rangle \in \mathcal{H}$, we have $\mathsf{F}(\rho, |\psi\rangle\langle\psi|) = \sqrt{\langle\psi|\rho|\psi\rangle}$. We define norm of $|\psi\rangle$ as $\||\psi\rangle\| \stackrel{\text{def}}{=} \sqrt{\langle\psi|\psi\rangle}$. For a quantum state $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$, we let $\mathtt{tr}_{\mathcal{H}_B}\rho$ represent the partial trace of ρ in $\mathcal{H}_A$ after tracing out $\mathcal{H}_B$. Let $\rho \in \mathcal{H}_A$ and $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be such that $\mathtt{tr}_{\mathcal{H}_B}|\phi\rangle\langle\phi| = \rho$, then we call $|\phi\rangle$ a *purification* of ρ. For a pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, its *Schmidt decomposition* is defined as $|\psi\rangle = \sum_{i=1}^r \sqrt{p_i} \cdot |v_i\rangle \otimes |w_i\rangle$, where the states $|v_i\rangle \in \mathcal{H}_A$ are orthonormal, and so are the states $|w_i\rangle \in \mathcal{H}_B$, and p is a probability distribution. It is easily seen that r is also equal to $\mathtt{rank}(\mathtt{tr}_{\mathcal{H}_A}|\psi\rangle\langle\psi|) = \mathtt{rank}(\mathtt{tr}_{\mathcal{H}_B}|\psi\rangle\langle\psi|)$ and is therefore the same in all Schmidt decompositions of $|\psi\rangle$. This number is also referred to as the *Schmidt rank* of $|\psi\rangle$ and denoted $\mathtt{S\text{-}rank}(|\psi\rangle)$. Sometimes we absorb the coefficients $\sqrt{p_i}$ in $|v_i\rangle \otimes |w_i\rangle$, in which case $|v_i\rangle, |w_i\rangle$ may not be unit vectors. It is not hard to verify that local unitary operations do not change Schmidt rank of a bipartite state. The next fact follows by considering Schmidt decomposition of the pure states involved; see, for example, Ex(2.81) of [13].

Fact 1. *Let $|\psi\rangle, |\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be such that $\mathtt{tr}_{\mathcal{H}_B}|\phi\rangle\langle\phi| = \mathtt{tr}_{\mathcal{H}_B}|\psi\rangle\langle\psi|$. There exists a unitary operation U on $\mathcal{H}_B$ such that $(I_{\mathcal{H}_A} \otimes U)|\psi\rangle = |\phi\rangle$, where $I_{\mathcal{H}_A}$ is the identity operator on $\mathcal{H}_A$.*

We will also need another fundamental fact, shown by Uhlmann [13].

Fact 2. (Uhlmann, [13]) *Let $\rho, \sigma \in \mathcal{H}_A$. Let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a purification of ρ and $\dim(\mathcal{H}_A) \leq \dim(\mathcal{H}_B)$. There exists a purification $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ of σ such that $\mathsf{F}(\rho,\sigma) = |\langle\phi|\psi\rangle|$.*

We define the *approximate Schmidt rank* as follows.

Definition 2.2. *Let $\epsilon > 0$. Let $|\psi\rangle$ be a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Define*

$$\mathtt{S\text{-}rank}_\epsilon(|\psi\rangle) \stackrel{\text{def}}{=} \min\{\mathtt{S\text{-}rank}(|\phi\rangle) : |\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \text{ and } \mathsf{F}(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) \geq 1 - \epsilon\}. \tag{2.3}$$

Define linear map $\mathsf{vecinv} : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{L}(\mathcal{H}_B, \mathcal{H}_A)$, where $\mathcal{L}(\mathcal{H}_B, \mathcal{H}_A)$ is the set of operators from $\mathcal{H}_B$ to $\mathcal{H}_A$, by $\mathsf{vecinv}(|x\rangle \otimes |y\rangle) \stackrel{\text{def}}{=} |x\rangle\langle y|$, and extend to all vectors in $\mathcal{H}_A \otimes \mathcal{H}_B$ by linearity. For $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, it is easily seen that $\||\psi\rangle\| = \|\mathsf{vecinv}(|\psi\rangle)\|_2$.

In the following sections we assume Hilbert spaces $\mathcal{H}_A, \mathcal{H}_{A_1}, \mathcal{H}_{A_2}$ etc. are possessed by Alice and Hilbert spaces $\mathcal{H}_B, \mathcal{H}_{B_1}, \mathcal{H}_{B_2}$ etc. are possessed by Bob. We start by showing the following key lemma which we will use many times in the following sections.

[4] We take the factor of half because we shall talk about a correlation as a *shared* resource. It is consistent with the convention that when the two parties shares a classical correlation (X,Y), where $Y = X = R$ for a r-bit random string R, we say that they share a random variable R of size r.

LEMMA 2.2. *Let ρ be a quantum state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Then,*

$$\mathsf{Q}(\rho) = \min_{\mathcal{H}_{A_1}, \mathcal{H}_{B_1}} \{ \lceil \log_2 \texttt{S-rank}(|\psi\rangle) \rceil : |\psi\rangle \textit{ is a pure state in } \mathcal{H}_{A_1} \otimes \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}, \rho = \texttt{tr}_{\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}} |\psi\rangle\langle\psi| \}.$$

Proof. Let $r \stackrel{\text{def}}{=} \min_{\mathcal{H}_{A_1}, \mathcal{H}_{B_1}} \{ \lceil \log_2 (\texttt{S-rank}(|\psi\rangle)) \rceil : \rho = \texttt{tr}_{\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}} |\psi\rangle\langle\psi| \}$. We first show $\mathsf{Q}(\rho) \leq r$. Let $|\psi\rangle$ be such that

$$r = \lceil \log_2 (\texttt{S-rank}(|\psi\rangle)) \rceil \text{ and } \rho = \texttt{tr}_{\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}} |\psi\rangle\langle\psi|.$$

Let $t \stackrel{\text{def}}{=} \texttt{S-rank}(|\psi\rangle)$. Let $|\psi\rangle$ have a Schmidt decomposition

$$|\psi\rangle = \sum_{i=1}^{t} \sqrt{p_i} \cdot |v_i\rangle \otimes |w_i\rangle,$$

Let Alice and Bob start with the state

$$|\phi\rangle = \sum_{i=1}^{t} \sqrt{p_i} \cdot |i\rangle \otimes |i\rangle,$$

and transform $|\phi\rangle$ to $|\psi\rangle$ using local unitary transformations. This shows that $\mathsf{Q}(\rho) \leq \lceil \log_2 t \rceil = r$.

For the other direction let $s \stackrel{\text{def}}{=} \mathsf{Q}(\rho)$. Let Alice and Bob start with the seed state σ and apply local *completely positive trace preserving maps* Φ_A, Φ_B respectively to produce ρ. Let us assume without loss of generality that the number of qubits of $\sigma_A \stackrel{\text{def}}{=} \texttt{tr}_{\mathcal{H}_B} \sigma$ is at most s. For convenience, here we suppose that σ lies in $\mathcal{H}_A \otimes \mathcal{H}_B$, and if this is not the case, similar proof still holds. Let $\sigma_A = \sum_{i=1}^{2^s} a_i |v_i\rangle\langle v_i|$, where $a_i \geq 0$ is the i-th eigenvalue of σ_A with eigenvector $|v_i\rangle$. Define

$$|\phi\rangle \stackrel{\text{def}}{=} \sum_{i=1}^{2^s} \sqrt{a_i} \cdot |v_i\rangle \otimes |v_i\rangle$$

and let

$$|\phi'\rangle = \sum_{i=1}^{2^s} \sqrt{a_i} \cdot |v_i\rangle \otimes |w_i\rangle$$

be a purification of σ, where $\forall i : \; |v_i\rangle \in \mathcal{H}_A$ and $|w_i\rangle \in \mathcal{H}_B \otimes \mathcal{H}_{B_2}$.

Now consider the following operations by Alice and Bob. They start with the shared state $|\phi\rangle$. Bob using local unitary (after attaching ancilla $|0\rangle$ if needed) transforms $|\phi\rangle$ to $|\phi'\rangle$ (Bob can do this due to Fact 1). Alice and Bob then simulate their maps Φ_A, Φ_B on σ by local unitaries (each after attaching ancilla $|0\rangle$ if needed on their parts; such a simulation is a standard fact, please refer to [13]) and finally produce a purification $|\theta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$ of ρ. Since Alice and Bob, using local unitary operations and attaching ancilla $|0\rangle$, transform $|\phi\rangle$ to $|\theta\rangle$, we have $2^s \geq \texttt{S-rank}(|\phi\rangle) = \texttt{S-rank}(|\theta\rangle)$. This shows that $r \leq s$.

The following lemma is credited to Nayak (personal communication) in [18]; we include a proof in Appendix for completeness.

LEMMA 2.3. *For a quantum state ρ in $\mathcal{H}_A \times \mathcal{H}_B$, $\mathsf{Q}(\rho) = \mathsf{QComm}(\rho)$.*

Classical information theory. Let p be a probability distribution on $\mathcal{X}$. For $x \in \mathcal{X}$, we let $p(x)$ denote the probability of x under p. For $\mathcal{X}_1 \subseteq \mathcal{X}$, we define $p(\mathcal{X}_1) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}_1} p(x)$. The *entropy* of p, denoted $H(p)$ is defined as $H(p) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} -p(x) \log p(x)$. Let p, q be probability distributions on $\mathcal{X}$. Then the ℓ_1 distance between them, denoted $\|p - q\|_1$ is defined as $\sum_{x \in \mathcal{X}} |p(x) - p(y)|$. Let p be a distribution on $\mathcal{X} \times \mathcal{Y}$. For $(x, y) \in \mathcal{X} \times \mathcal{Y}$, define $p(y|x) \stackrel{\text{def}}{=} \frac{p(x,y)}{p(x)}$ and $p(y) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} p(x, y)$. Similarly one can define $p(x|y)$ and $p(x)$. We use a random variable to also represent its distribution. Let joint random variable (X, Y) be distributed in $\mathcal{X} \times \mathcal{Y}$ and have joint distribution p. The *mutual information* between X and Y is defined as $I(X : Y) \stackrel{\text{def}}{=} H(X) + H(Y) - H(XY) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log \frac{p(x,y)}{p(x) \cdot p(y)}$.

We will use the following fact, which follows from the log sum inequality, see for example [10].

FACT 3. *Let p, q be a probability distributions on $\mathcal{X}$. Then for all $\mathcal{X}' \subseteq \mathcal{X} : \sum_{x \in \mathcal{X}'} p(x) \log \frac{p(x)}{q(x)} \geq -1$.*

3 Correlation complexity of precisely generating a quantum state and a classical distribution

In this section we show characterizations of correlation complexities for general quantum states (Theorem 1.2) and also for classical distributions (Theorem 1.1). We start with the following lemma, which, together with Lemma 2.2, implies Theorem 1.2. The proof is deferred to Appendix.

LEMMA 3.1. *Let ρ be a quantum state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $|x\rangle$, $|x'\rangle$ range over the computational basis states for $\mathcal{H}_A$, and $|y\rangle$, $|y'\rangle$ for $\mathcal{H}_B$. Then there exists a purification $|\psi\rangle$ of ρ, with $\texttt{S-rank}(|\psi\rangle) = r$, if and only if there exist matrices $\{A_x\}$ and $\{B_y\}$, each with r columns, such that*

$$\rho = \sum_{x,x';y,y'} |x\rangle\langle x'| \otimes |y\rangle\langle y'| \cdot \texttt{tr}\left((A_{x'}^{\dagger} A_x)^T (B_{y'}^{\dagger} B_y)\right).$$

We now use similar arguments as in the proof of the above lemma to show Theorem 1.1

Proof. (of Theorem 1.1) We will first show $\mathsf{Q}(\rho) \leq \lceil \log_2 \texttt{rank}_{\texttt{psd}}(P) \rceil$, where $\rho = \sum_{x,y} P(x,y)|x\rangle\langle x| \otimes |y\rangle\langle y|$ (and generating P is just the same as generating ρ). Let $r = \texttt{rank}_{\texttt{psd}}(P)$. We will exhibit a purification $|\psi\rangle$ of ρ with $\texttt{S-rank}(|\psi\rangle) = r$. This combined with Lemma 2.2 will show $\mathsf{Q}(\rho) \leq \lceil \log_2 r \rceil$. Let $C_x, D_y \in \mathbb{C}^{r\times r}$ be positive semi-definite matrices with $\texttt{tr}(C_x D_y) = P(x,y)$, $\forall x \in \mathcal{X}, y \in \mathcal{Y}$. For $i \in [r]$, let $|v_x^i\rangle$ be the i-th column of $\sqrt{C_x^T}$ and let $|w_y^i\rangle$ be the i-th column of $\sqrt{D_y}$. Define $|\psi\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$ as follows.

$$|\psi\rangle \stackrel{\text{def}}{=} \sum_{i=1}^{r} \Big(\sum_x |x\rangle \otimes |x\rangle \otimes |v_x^i\rangle \Big) \otimes \Big(\sum_y |y\rangle \otimes |y\rangle \otimes |w_y^i\rangle \Big).$$

It is clear that $\texttt{S-rank}(|\psi\rangle) \leq r$. Also,

$$\begin{aligned}
&\texttt{tr}_{\mathcal{H}_A\otimes\mathcal{H}_{A_1}\otimes\mathcal{H}_B\otimes\mathcal{H}_{B_1}} |\psi\rangle\langle\psi| \\
=&\sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \left(\sum_{i,j=1}^{r} \langle v_x^j|v_x^i\rangle \cdot \langle w_y^j|w_y^i\rangle \right) \\
=&\sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \cdot \texttt{tr}(C_x D_y) \quad = \quad \rho.
\end{aligned}$$

Note that Alice and Bob after sharing $|\psi\rangle$ can measure their first registers in their respective computational basis (discarding the second and third registers) to obtain ρ.

Now we will show $\mathsf{Q}(\rho) \geq \lceil \log_2 \texttt{rank}_{\texttt{psd}}(P) \rceil$. Let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$ be a purification of ρ with $\texttt{S-rank}(|\psi\rangle) = r$ and $\mathsf{Q}(\rho) = \lceil \log_2 r \rceil$, as guaranteed by Lemma 2.2. We will show $r \geq \texttt{rank}_{\texttt{psd}}(P)$ and this will show the desired. Let

$$|\psi\rangle = \sum_{i=1}^{r} \left(\sum_x |x\rangle \otimes |v_x^i\rangle \right) \otimes \left(\sum_y |y\rangle \otimes |w_y^i\rangle \right).$$

For all x, define $r \times r$ matrices C_x such that $C_x(j,i) = \langle v_x^j|v_x^i\rangle$ for all $i,j \in [r]$. Similarly for all y, define $r \times r$ matrices D_y such that $D_y(i,j) = \langle w_y^j|w_y^i\rangle$ for all $i,j \in [r]$. Then C_x, D_y are positive semi-definite for all x and y. Consider

$$\begin{aligned}
\rho &= \texttt{tr}_{\mathcal{H}_{A_1}\otimes\mathcal{H}_{B_1}} |\psi\rangle\langle\psi| \\
&= \sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \left(\sum_{i,j=1}^{r} \langle v_x^j|v_x^i\rangle \cdot \langle w_y^j|w_y^i\rangle \right) \\
&= \sum_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y| \cdot \texttt{tr}(C_x D_y).
\end{aligned}$$

Therefore for all x and y we have $p_{x,y} = P(x,y) = \texttt{tr}(C_x D_y)$. Hence $\texttt{rank}_{\texttt{psd}}(P) \leq r$.

4 Classical complexity of approximating a distribution

In this section we prove Theorem 1.3, and we will use the lower-case p to denote the distribution P in the statement.

Proof. (of Theorem 1.3) Let W be a random variable, taking values in the set $\mathcal{W}$, that achieves the minimum in the definition of $C(X,Y)$. Let $Z \stackrel{\text{def}}{=} (X,Y)$ and $\mathcal{Z} \stackrel{\text{def}}{=} \mathcal{X} \times \mathcal{Y}$. The idea is to find a small number (in terms of $C(X,Y)$) of $w_i \in \mathcal{W}$'s such that the uniform average, over w_i's, of the conditional distributions $(Z|w_i)$ is close to Z itself. Recall that $(X,Y) \sim p$.

Lemma 4.1. *Let q be the joint distribution of (Z,W), and*

$$\begin{aligned} k &\stackrel{\text{def}}{=} C(X:Y) = I(Z:W) \\ &= \sum_{(z,w)\in\mathcal{Z}\times\mathcal{W}} q(z,w)\log\frac{q(z|w)}{q(z)}, \end{aligned}$$

where we use convention $0\log(0/0)=0$. For any $\delta\in(0,1)$, there exists an $n = O(2^{(k+1)/\delta}\cdot\delta^{-3}\log(1/\delta))$ and a set $\{w_1,...,w_n\}$ s.t. the distribution p' on $\mathcal{Z}$ defined by $p'(z) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n q(z|w_i)$ satisfies $\|p-p'\|_1 \le 8\delta$, where $w_i \in \mathcal{W}$.

Proof. Define

$$\mathsf{Good}_1 \stackrel{\text{def}}{=} \{(z,w)\in\mathcal{Z}\times\mathcal{W} : q(z|w) \le 2^{(k+1)/\delta} q(z)\}.$$

Using Fact 3 on the set $\{(z,w)\in\mathcal{Z}\times\mathcal{W} : q(z|w) > 2^{(k+1)/\delta}q(z)\}$, we have that $\sum_{(z,w)\notin\mathsf{Good}_1} q(z,w)\log\frac{q(z|w)}{q(z)} \le k+1$. Considering for any $(z,w)\notin\mathsf{Good}_1$, it holds that $\log\frac{q(z|w)}{q(z)} > (k+1)/\delta$, then we get that $q(\mathsf{Good}_1) \ge 1-\delta$. Define

$$\mathsf{Good}_2 \stackrel{\text{def}}{=} \{z : \sum_{w:(z,w)\in\mathsf{Good}_1} q(z,w) \ge \delta\sum_w q(z,w)\}.$$

Define a function q' on $\mathcal{Z}\times\mathcal{W}$ as follows.

$$q'(z,w) \stackrel{\text{def}}{=} \begin{cases} q(z,w) & \text{if } (z,w)\in\mathsf{Good}_1 \text{ and } z\in\mathsf{Good}_2 \\ 0 & \text{otherwise} \end{cases}$$

and define $q'(z|w) \stackrel{\text{def}}{=} q'(z,w)/q(w)$ and $q'(z) \stackrel{\text{def}}{=} \sum_w q'(z,w) = \sum_w q(w)q'(z|w)$. Note that $q'(z,w)$ is not a distribution, but it can be seen that q' deviates from q by a small amount:

$$\begin{aligned} &\sum_{(z,w)\in\mathcal{Z}\times\mathcal{W}} \left|q'(z,w)-q(z,w)\right| && (4.4)\\ &= \sum_{(z,w)\notin\mathsf{Good}_1} q(z,w) + \sum_{(z,w)\in\mathsf{Good}_1:\ z\notin\mathsf{Good}_2} q(z,w) \\ &\le\ 1-q(\mathsf{Good}_1) + \sum_{z\notin\mathsf{Good}_2} \delta\cdot q(z) \le 2\delta. && (4.5) \end{aligned}$$

Let us sample $\{w_1,w_2,\dots,w_n\}$, where each $w_i\in\mathcal{W}$ is sampled independently according to W. For all $z\in\mathcal{Z}$ we have, $\mathbf{E}_{w_i\leftarrow W}\left[q'(z|w_i)\right] = q'(z)$ by the definitions of $q'(z|w_i)$ and $q'(z)$. Also for all $(z,w)\in\mathcal{Z}\times\mathcal{W}$

$$q'(z|w) \le \frac{1}{\delta}q'(z)2^{(k+1)/\delta}. \qquad (4.6)$$

The inequality above is easily seen for $z\notin\mathsf{Good}_2$ or $(z,w)\notin\mathsf{Good}_1$. Now assume that $z\in\mathsf{Good}_2$ and $(z,w)\in\mathsf{Good}_1$, then by definition,

$$q'(z|w_i) = q(z|w_i) \le 2^{(k+1)/\delta}q(z) \le \frac{1}{\delta}2^{(k+1)/\delta}q'(z),$$

where the first inequality is because $(z,w) \in \mathsf{Good}_1$ and the second inequality is because $z \in \mathsf{Good}_2$.

With these bounds set up, we are able to apply Chernoff's bound and get that for all z with $q'(z) \neq 0$,

$$\begin{aligned}
&\mathbf{Pr}_{\{w_i\}}\Big[\Big|\frac{q'(z|w_1)+\cdots+q'(z|w_n)}{n} - q'(z)\Big| > \delta q'(z)\Big] \\
=&\mathbf{Pr}_{\{w_i\}}\Big[\Big|\frac{\sum_i q'(z|w_i)}{q'(z)2^{(k+1)/\delta}/\delta} - \frac{n}{2^{(k+1)/\delta}/\delta}\Big| > \frac{\delta n}{2^{(k+1)/\delta}/\delta}\Big] \\
\leq& 2e^{-\delta^3 n/(3\cdot 2^{(k+1)/\delta})} \leq \delta.
\end{aligned} \tag{4.7}$$

In the second step we used Eq. (4.6). Since the bound is trivially true if $q'(z) = 0$, we have the bound regardless of the value of $q'(z)$. Therefore,

$$\begin{aligned}
&\mathbf{E}_{\{w_i\}}\left[\sum_z |p'(z) - p(z)|\right] \\
&= \mathbf{E}_{\{w_i\}}\left[\sum_z \Big|\frac{\sum_i q(z|w_i)}{n} - p(z)\Big|\right] \\
&\leq \mathbf{E}_{\{w_i\}}\Big[\sum_z \Big(\Big|\frac{\sum_i q(z|w_i)}{n} - \frac{\sum_i q'(z|w_i)}{n}\Big| \\
&\qquad + \Big|\frac{\sum_i q'(z|w_i)}{n} - q'(z)\Big| + |q'(z) - p(z)|\Big)\Big] \\
&\leq \mathbf{E}_{\{w_i\}}\Big[\frac{1}{n}\sum_i \sum_z |q(z|w_i) - q'(z|w_i)|\Big] \\
&\quad + \sum_z \mathbf{E}_{\{w_i\}}\Big[\Big|\frac{\sum_i q'(z|w_i)}{n} - q'(z)\Big|\Big] + \sum_z |q'(z) - p(z)| \\
&= \mathbf{E}_{w\leftarrow W}\Big[\sum_z |q(z|w) - q'(z|w)|\Big] \\
&\quad + \sum_z \mathbf{E}_{\{w_i\}}\Big[\Big|\frac{\sum_i q'(z|w_i)}{n} - q'(z)\Big|\Big] + \sum_z |q'(z) - p(z)|.
\end{aligned}$$

The first summand is $\sum_{w,z} |q(z,w) - q'(z,w)| \leq 2\delta$, by Eq.(4.4). Using the same relationship, one can also prove that the third summand is at most $\sum_{z,w} |q'(z,w) - q(z,w)| \leq 2\delta$. We now show that the second summand is at most 4δ. In fact, for an arbitrary set $\{w_i\}$, we denote those z which do not make the event in Eq.(4.7) happen by $\mathsf{Good}_3(\{w_i\})$, then we have that

$$\begin{aligned}
&\sum_z \mathbf{E}_{\{w_i\}}\Big[\Big|\frac{\sum_i q'(z|w_i)}{n} - q'(z)\Big|\Big] \\
=&\sum_z \sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot \Big|\frac{\sum_i q'(z|\overline{w_i})}{n} - q'(z)\Big| \\
=&\sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}]\Big(\sum_{z\in\mathsf{Good}_3(\{\overline{w_i}\})} \Big|\frac{\sum_i q'(z|\overline{w_i})}{n} - q'(z)\Big| + \sum_{z\notin\mathsf{Good}_3(\{\overline{w_i}\})} \Big|\frac{\sum_i q'(z|\overline{w_i})}{n} - q'(z)\Big|\Big) \\
\leq&\sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}]\Big(\sum_{z\in\mathsf{Good}_3(\{\overline{w_i}\})} \delta q'(z) + \sum_{z\notin\mathsf{Good}_3(\{\overline{w_i}\})} \Big(\frac{\sum_i q'(z|\overline{w_i})}{n} + q'(z)\Big)\Big) \\
\leq&\delta \sum_z q'(z) + \sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot \sum_{z\notin\mathsf{Good}_3(\{\overline{w_i}\})} \Big(\frac{\sum_i q'(z|\overline{w_i})}{n}\Big) \qquad (4.8) \\
&+ \sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot \sum_{z\notin\mathsf{Good}_3(\{\overline{w_i}\})} q'(z).
\end{aligned}$$

We now bound the three terms in Eq.(4.8) separately. Apparently, the first term $\delta \sum_z q'(z) \le \delta$. The third term

$$\begin{aligned}(4.9)\qquad &\sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot \sum_{z \notin \mathsf{Good}_3(\{\overline{w_i}\})} q'(z) \\ =&\sum_z \sum_{\{\overline{w_i}\}: z \notin \mathsf{Good}_3(\{\overline{w_i}\})} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot q'(z) \\ \le&\sum_z \delta \cdot q'(z) \\ \le&\delta,\end{aligned}$$

where we have used the result in Eq.(4.7). In order to bound the second term of Eq.(4.8), note that for any i,

$$\sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] q'(z|\overline{w_i}) = \sum_{\overline{w_i}} \mathbf{Pr}[w_i = \overline{w_i}] q'(z|\overline{w_i}) = q'(z),$$

which indicates that

$$\sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot \Big(\frac{\sum_i q'(z|\overline{w_i})}{n} \Big) = q'(z).$$

In this way, we have that

$$(4.10)\qquad \sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot \sum_z \Big(\frac{\sum_i q'(z|\overline{w_i})}{n} \Big) = \sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \sum_z q'(z).$$

On the other hand, it holds that

$$\begin{aligned}(4.11)\qquad &\sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \Big| \sum_{z \in \mathsf{Good}_3(\{\overline{w_i}\})} \frac{\sum_i q'(z|\overline{w_i})}{n} - \sum_{z \in \mathsf{Good}_3(\{\overline{w_i}\})} q'(z) \Big| \\ \le& \sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \sum_{z \in \mathsf{Good}_3(\{\overline{w_i}\})} \Big| \frac{\sum_i q'(z|\overline{w_i})}{n} - q'(z) \Big| \\ \le&\delta,\end{aligned}$$

where the second inequality has been proved above. Combining Eq.(4.10) and Eq.(4.11), we obtain that

$$(4.12)\qquad \sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \Big| \sum_{z \notin \mathsf{Good}_3(\{\overline{w_i}\})} \frac{\sum_i q'(z|\overline{w_i})}{n} - \sum_{z \notin \mathsf{Good}_3(\{\overline{w_i}\})} q'(z) \Big| \le \delta.$$

Substituting Eq.(4.9) into Eq.(4.12), we eventually get that the second term in Eq.(4.8)

$$\sum_{\{\overline{w_i}\}} \mathbf{Pr}[\{w_i\} = \{\overline{w_i}\}] \cdot \sum_{z \notin \mathsf{Good}_3(\{\overline{w_i}\})} \Big(\frac{\sum_i q'(z|\overline{w_i})}{n} \Big) \le 2\delta.$$

Putting these bounds together, we have $\mathbf{E}_{\{w_i\}} \left[\sum_z |p'(z) - p(z)| \right] \le 8\delta$. Thus there exists $\{w_i\}$ to make $\|p' - p\|_1 \le 8\delta$ as well. This completes the proof of Lemma 4.1.

We now continue the proof of Theorem 1.3. Consider the set $\{w_i\}$ given by the above lemma. Consider the protocol in which Alice and Bob sample $i \in [n]$ uniformly using public-coins and generate $X|(W = w_i)$ and $Y|(W = w_i)$, respectively locally (since $X|(W = w_i)$ and $Y|(W = w_i)$ are independent). Then by the lemma above, their joint distribution p' (on $\mathcal{X} \times \mathcal{Y}$) satisfies $\|p - p'\|_1 \le 8\delta$. The correlation cost of the protocol is $\lceil \log n \rceil = (k+1)/\delta + O(\log(1/\delta))$.

5 Correlation complexity of approximating a pure state

In this section we prove Theorem 1.4. We start by characterizing the approximate Schmidt rank in the following lemma.

LEMMA 5.1. *Let $\epsilon > 0$. Let $|\psi\rangle$ be a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$ with a Schmidt decomposition $|\psi\rangle = \sum_{i=1}^{r} \sqrt{p_i} \cdot |v_i\rangle \otimes |w_i\rangle$ (with $p_1 \geq p_2 \geq \ldots \geq p_r > 0$ and $\sum_{i=1}^{r} p_i = 1$). Let r' be the minimum number such that $\sum_{i=1}^{r'} p_i \geq (1-\epsilon)^2$. Then $r' = \texttt{S-rank}_\epsilon(|\psi\rangle)$.*

Proof. We will first show that $r' \geq \texttt{S-rank}_\epsilon(|\psi\rangle)$. Let $q = \sum_{i=1}^{r'} p_i$. Define

$$|\phi\rangle \stackrel{\text{def}}{=} \frac{1}{\sqrt{q}} \cdot \sum_{i=1}^{r'} \sqrt{p_i} \cdot |v_i\rangle \otimes |w_i\rangle.$$

Then $\mathsf{F}(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = |\langle\psi|\phi\rangle| = \sqrt{q} \geq 1 - \epsilon$. Clearly $\texttt{S-rank}(|\phi\rangle) = r'$ and hence $r' \geq \texttt{S-rank}_\epsilon(|\psi\rangle)$.

Now we will show that $r' \leq \texttt{S-rank}_\epsilon(|\psi\rangle)$. Let $s = \texttt{S-rank}_\epsilon(|\psi\rangle)$. Let $|\theta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a pure state such that $|\langle\theta|\psi\rangle| = \mathsf{F}(|\psi\rangle\langle\psi|, |\theta\rangle\langle\theta|) \geq 1 - \epsilon$ and $\texttt{S-rank}(|\theta\rangle) = s$. Without loss of generality (by multiplying $|\theta\rangle$ by an appropriate phase) let us assume that $\beta \stackrel{\text{def}}{=} \langle\psi|\theta\rangle$ is real. Let $|\theta\rangle = \sum_{j=1}^{s} \sqrt{q_i} \cdot |v'_j\rangle \otimes |w'_j\rangle$ be a Schmidt decomposition of $|\theta\rangle$. Define

$$A \stackrel{\text{def}}{=} \sum_{i=1}^{r} \sqrt{p_i} \cdot |v_i\rangle\langle w_i| \quad \text{and} \quad B \stackrel{\text{def}}{=} \beta \cdot \sum_{i=1}^{s} \sqrt{q_i} \cdot |v'_i\rangle\langle w'_i|.$$

Note that $A = \mathsf{vecinv}(|\psi\rangle)$ and $B = \mathsf{vecinv}(|\theta'\rangle)$, where $|\theta'\rangle = \beta|\theta\rangle$. Since $\{v_i\}$ and $\{w_i\}$ are orthonormal, $\{\sqrt{p_i}\}$ form the singular values of A. Similarly $\{\beta \cdot \sqrt{q_i}\}$ form the singular values of B. Now,

$$\begin{aligned} 1 - (1-\epsilon)^2 \geq 1 - \beta^2 &= \||\psi\rangle\|^2 + \||\theta'\rangle\|^2 - 2\langle\theta'|\psi\rangle \\ &= \||\psi\rangle - |\theta'\rangle\|^2 = \|\mathsf{vecinv}(|\psi\rangle - |\theta'\rangle)\|_2^2 \\ &= \|\mathsf{vecinv}(|\psi\rangle) - \mathsf{vecinv}(|\theta'\rangle)\|_2^2 = \|A - B\|_2^2. \end{aligned}$$

Hence from Lemma 2.1, $\texttt{S-rank}(|\theta\rangle) = \texttt{rank}(B) \geq r'$.

We can now get the desired characterization for $\mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|)$.

THEOREM 5.1. *Let $\epsilon > 0$. Let $|\psi\rangle$ be a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$ with a Schmidt decomposition $|\psi\rangle = \sum_{i=1}^{r} \sqrt{p_i} \cdot |v_i\rangle \otimes |w_i\rangle$ (with $p_1 \geq p_2 \geq \ldots \geq p_r > 0$ and $\sum_{i=1}^{r} p_i = 1$). Let $A = \sum_{i=1}^{r} \sqrt{p_i} \cdot |v_i\rangle\langle w_i| = \mathsf{vecinv}(|\psi\rangle)$. Then, $\mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|) = \lceil \log_2 \texttt{rank}_{2\epsilon-\epsilon^2}(A) \rceil$.*

Proof. From the definitions and Lemma 2.2 it is clear that $\mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|) = \lceil \log_2 \texttt{S-rank}_\epsilon(|\psi\rangle) \rceil$. Also from Lemma 2.1 and Lemma 5.1 it follows that $\texttt{S-rank}_\epsilon(|\psi\rangle) = \texttt{rank}_{2\epsilon-\epsilon^2}(A)$ (by noting that $\{\sqrt{p_i}\}$ form singular values of A).

The following lemma shows a monotonicity property for the approximate Schmidt rank.

LEMMA 5.2. *Let $|\psi\rangle$ be a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$ and $|\theta\rangle$ a pure state in $\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}$. Then,*

$$\texttt{S-rank}_\epsilon(|\psi\rangle \otimes |\theta\rangle) \geq \texttt{S-rank}_\epsilon(|\psi\rangle).$$

Hence from Lemma 2.2,

$$\mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi| \otimes |\theta\rangle\langle\theta|) \geq \mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|).$$

Proof. Let $|\psi\rangle = \sum_{i=1}^{r} \sqrt{p_i} \cdot |u_i^1\rangle \otimes |v_i^1\rangle$ (with $p_1 \geq \cdots \geq p_r > 0$) and $|\theta\rangle = \sum_{i=1}^{s} \sqrt{q_i} \cdot |u_i^2\rangle \otimes |v_i^2\rangle$ be some Schmidt decompositions of $|\psi\rangle$ and $|\theta\rangle$ respectively. Then

$$|\psi\rangle \otimes |\theta\rangle = \sum_{i,j} \sqrt{p_i q_j} \cdot |u_i^1\rangle \otimes |u_i^2\rangle \otimes |v_i^1\rangle \otimes |v_i^2\rangle.$$

Fix a minimal set $S \subseteq [r] \times [s]$ with $\sum_{(i,j)\in S} p_i q_j \geq (1-\epsilon)^2$. Let $r' \stackrel{\text{def}}{=} \texttt{S-rank}_\epsilon(|\psi\rangle)$. We will show $|S| \geq r'$. Assume for contradiction $|S| \leq r'-1$. Let $S_1 = \{i : \exists j \text{ such that } (i,j) \in S\}$, then $|S_1| \leq |S| \leq r'-1$. We have

$$\sum_{(i,j)\in S} p_i q_j \leq \sum_{i\in S_1} p_i \leq p_1 + \cdots + p_{|S_1|} < (1-\epsilon)^2,$$

where the first inequality is because $\sum_{j:(i,j)\in S} q_j \leq 1$ for all i, the second inequality is because p_i's are in the non-increasing order, and the last one is by the definition of $\texttt{S-rank}_\epsilon(|\psi\rangle) = r'$, the smallest number such that $p_1 + \cdots + p_{r'} \geq (1-\epsilon)^2$ (from Lemma 5.1). This contradicts the way we picked S and hence

$$\texttt{S-rank}_\epsilon(|\psi\rangle \otimes |\theta\rangle) = |S| \geq r' = \texttt{S-rank}_\epsilon(|\psi\rangle),$$

where we use Lemma 5.1 again.

THEOREM 5.2. *Let $\epsilon > 0$. Let $|\psi\rangle$ be a pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Then, $\mathsf{Q}_\epsilon(|\psi\rangle\langle\psi|) = \mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|)$.*

Proof. By definition, we have $\mathsf{Q}_\epsilon(|\psi\rangle\langle\psi|) \leq \mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|)$. Now consider the other direction. By the definition of $\mathsf{Q}_\epsilon(|\psi\rangle\langle\psi|)$, there exists a $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$ such that

$$\mathsf{Q}_\epsilon(|\psi\rangle\langle\psi|) = \mathsf{Q}(\rho) \text{ and } \mathsf{F}(\rho, |\psi\rangle\langle\psi|) \geq 1-\epsilon. \tag{5.13}$$

By Lemma 2.2, there exists a purification $|\phi\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$ of ρ with

$$\mathsf{Q}(\rho) = \lceil \log_2 \texttt{S-rank}(|\phi\rangle\langle\phi|) \rceil = \mathsf{Q}(|\phi\rangle\langle\phi|). \tag{5.14}$$

Without loss of generality, we can assume that $\dim(\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}) \geq \dim(\mathcal{H}_A \otimes \mathcal{H}_B)$ (otherwise we can attach $|0\rangle$ to $|\phi\rangle$ appropriately). Now by Uhlmann's Theorem, there exists a pure state $|\psi'\rangle = |\psi\rangle \otimes |\theta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$ such that $|\theta\rangle \in \mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}$ and $|\langle\phi|\psi'\rangle| \geq 1-\epsilon$. Therefore,

$$\begin{aligned}
& \mathsf{Q}_\epsilon^{pure}(|\psi\rangle\langle\psi|) && \\
\leq\ & \mathsf{Q}_\epsilon^{pure}(|\psi'\rangle\langle\psi'|) && \text{(from Lemma 5.2)} \\
\leq\ & \mathsf{Q}(|\phi\rangle\langle\phi|) && \text{(from def. of } \mathsf{Q}_\epsilon^{pure}(|\psi'\rangle\langle\psi'|)) \\
=\ & \mathsf{Q}(\rho) && \text{(from Eq. (5.14))} \\
=\ & \mathsf{Q}_\epsilon(|\psi\rangle\langle\psi|). && \text{(from Eq. (5.13))}
\end{aligned}$$

Theorem 1.4 now follows immediately by combining Theorem 5.1 and Theorem 5.2 and noting that the matrix A as defined in the statement of Theorem 1.4 is $\mathsf{vecinv}(|\psi\rangle)$.

Acknowledgments R.J. is supported by the internal grants of Centre for Quantum Technologies. Y.S. and S.Z. were partially supported by China Basic Research Grant 2011CBA00300 (sub-project 2011CBA00301). Y.S. was also supported in part by US NSF (1017335 and 1216729). Z.W. would like to acknowledge the WBS grant under contract no. R-710-000-007-271. S.Z. was also supported by Research Grants Council of the Hong Kong S.A.R. (Project no. CUHK419309, CUHK418710, CUHK419011).

References

[1] Andris Ambainis, Leonard Schulman, Amnon Ta-Shma, Umesh Vazirani, and Avi Wigderson. The quantum communication complexity of sampling. *SIAM Journal on Computing*, 32(6):1570–1585, 2003.

[2] Dorit Aharonov, Itai Arad, Umesh Vazirani and Zeph Landau. The detectability lemma and its applications to quantum Hamiltonian complexity. *New Journal of Physics*, 13(11):113043, 2011.

[3] Alexander Barvinok. Approximations of convex bodies by polytopes and by projections of spectrahedra. *arXiv:1204.0471*, 2012.

[4] Leroy Beasley and Thomas Laffey. Real rank versus nonnegative rank. *Linear Algebra and its Applications*, 431:2330–2335, 2009.

[5] Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Review of Modern Physics*, 82:665–698, 2010.

[6] Moody Chu and Robert Plemmons. Nonnegative matrix factorization and applications. *Linear Algebra and its Applications*, 34:2–7, 2005.

[7] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[8] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Linear vs. semidefinite extended formulations: Exponential separation and strong lower bounds. In *Proceedings of the 44th ACM Symposium on Theory of Computing*, pages 95–106, 2012.

[9] João Gouveia, Pablo A. Parrilo, and Rekha Thomas. Lifts of convex sets and cone factorizations. *arXiv:1111.3164*, 2011.

[10] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2009.

[11] Hartmut Klauck, Troy Lee, and Shengyu Zhang. An explicit and exponential separation between randomized and quantum correlation complexities. *manuscript*, 2011.

[12] László Lovász. *Communication complexity: A survey. In book Paths, flows, and VLSI layout edited by B. Korte, L. Lovász, H. J. Prömel, and A. Schrijver.* Springer-Verlag, 1990.

[13] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, UK, 2000.

[14] Noam Nisan. Lower bounds for non-commutative computation. In *Proceedings of the 23th Annual ACM Symposium on Theory of Computing*, pages 410–418, 1991.

[15] Andreas Winter. Secret, public and quantum correlation cost of triples of random variables. In *Proceedings of the 2005 IEEE International Symposium on Information Theory*, pages 2270–2274, 2005.

[16] Aaron Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975.

[17] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pages 223–228, 1988.

[18] Shengyu Zhang. Quantum strategic game theory. In *Proceedings of the 3rd Innovations in Theoretical Computer Science*, pages 39–59, 2012. Earlier at arXiv:1012.5141 and QIP'11.

A Deferred proofs

A.1 Proof of Lemma 2.3

Proof. Clearly $\mathsf{Q}(\rho) \geq \mathsf{QComm}(\rho)$. For the other direction let $r \stackrel{\text{def}}{=} \mathsf{QComm}(\rho)$. Let Alice and Bob start with the state $\sigma_A \otimes \sigma_B \in \mathcal{H}_A \otimes \mathcal{H}_B$, do local quantum operations, communicate r qubits and at the end output ρ. This protocol can be converted into another protocol where Alice and Bob start with a purification $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$ of $\sigma_A \otimes \sigma_B$ (with $\texttt{S-rank}(|\phi\rangle) = 1$), do local unitaries, exchange r qubits and at the end output a purification $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$ of ρ. Since local unitaries do not increase the Schmidt rank of the shared state and exchanging r qubits increases the Schmidt rank by a factor at most 2^r (since the $\texttt{rank}$ of the marginal state possessed by Alice increases by at most a factor 2 on receiving a qubit from Bob, and similarly for Bob on receiving a qubit from Alice), we have $\texttt{S-rank}(|\psi\rangle) \leq 2^r$. Hence from Lemma 2.2, $\mathsf{Q}(\rho) \leq r$.

A.2 Proof of Lemma 3.1

Proof. We first show the 'only if' implication. Let $|\psi\rangle$ be a purification of ρ in $\mathcal{H}_A \otimes \mathcal{H}_{A_1} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B_1}$. Let $\texttt{S-rank}(|\psi\rangle) = r$. Consider a Schmidt decomposition of $|\psi\rangle$.

$$\begin{aligned}|\psi\rangle &= \sum_{i=1}^{r} |v^i\rangle \otimes |w^i\rangle \\ &= \sum_{i=1}^{r} \Big(\sum_x |x\rangle \otimes |v_x^i\rangle \Big) \otimes \Big(\sum_y |y\rangle \otimes |w_y^i\rangle \Big).\end{aligned}$$

Above for any i, x, y, the vectors $|v^i\rangle, |w^i\rangle, |v^i_x\rangle, |w^i_y\rangle$ are not necessarily unit vectors. Consider

$$\begin{aligned}
\rho &= \mathtt{tr}_{\mathcal{H}_{A_1}\otimes\mathcal{H}_{B_1}} |\psi\rangle\langle\psi| \\
&= \mathtt{tr}_{\mathcal{H}_{A_1}\otimes\mathcal{H}_{B_1}} \Big(\sum_{i=1}^{r} \big(\sum_x |x\rangle \otimes |v^i_x\rangle \big) \otimes \big(\sum_y |y\rangle \otimes |w^i_y\rangle \big) \big) \\
&\qquad \big(\sum_{j=1}^{r} \big(\sum_{x'} \langle x'| \otimes \langle v^j_{x'}| \big) \otimes \big(\sum_{y'} \langle y'| \otimes \langle w^j_{y'}| \big) \Big) \\
&= \mathtt{tr}_{\mathcal{H}_{A_1}\otimes\mathcal{H}_{B_1}} \sum_{i,j} \Big(\sum_{x,x'} |x\rangle\langle x'| \otimes |v^i_x\rangle\langle v^j_{x'}| \Big) \\
&\qquad \otimes \Big(\sum_{y,y'} |y\rangle\langle y'| \otimes |w^i_y\rangle\langle w^j_{y'}| \Big) \\
&= \sum_{i,j} \Big(\sum_{x,x'} |x\rangle\langle x'| \cdot \langle v^j_{x'}|v^i_x\rangle \Big) \otimes \Big(\sum_{y,y'} |y\rangle\langle y'| \cdot \langle w^j_{y'}|w^i_y\rangle \Big) \\
&= \sum_{x,x';y,y'} |x\rangle\langle x'| \otimes |y\rangle\langle y'| \Big(\sum_{i,j} \langle v^j_{x'}|v^i_x\rangle \cdot \langle w^j_{y'}|w^i_y\rangle \Big).
\end{aligned}$$

For each x, let us define matrices $A_x \stackrel{\text{def}}{=} (|v^1_x\rangle, |v^2_x\rangle, \ldots, |v^r_x\rangle)$. Similarly for each y, let us define matrices $B_y \stackrel{\text{def}}{=} (|w^1_y\rangle, |w^2_y\rangle, \ldots, |w^r_y\rangle)$. Then from above,

$$\rho = \sum_{x,x';y,y'} |x\rangle\langle x'| \otimes |y\rangle\langle y'| \cdot \mathtt{tr}\Big((A^\dagger_{x'} A_x)^T (B^\dagger_{y'} B_y) \Big).$$

Next we show the 'if' implication. Let there exist matrices $\{A_x : x \in [\dim(\mathcal{H}_A)]\}$ and $\{B_y : y \in [\dim(\mathcal{H}_B)]\}$, each with r columns, such that

$$\rho = \sum_{x,x';y,y'} |x\rangle\langle x'| \otimes |y\rangle\langle y'| \cdot \mathtt{tr}\left((A^\dagger_{x'} A_x)^T (B^\dagger_{y'} B_y) \right).$$

For $i \in [r]$, let $|v^i_x\rangle$ be the i-th column of A_x and let $|w^i_y\rangle$ be the i-th column of B_y. Define

$$|\psi\rangle \stackrel{\text{def}}{=} \sum_{i=1}^{r} \left(\sum_x |x\rangle \otimes |v^i_x\rangle \right) \otimes \left(\sum_y |y\rangle \otimes |w^i_y\rangle \right)$$

It is clear that $\mathtt{S\text{-}rank}(|\psi\rangle) = r$. We can check, by analogous calculations as above, that

$$\rho = \mathtt{tr}_{\mathcal{H}_{A_1}\otimes\mathcal{H}_{B_1}} |\psi\rangle\langle\psi|.$$