A Hierarchy for Constant Communication Complexity

Andris Ambainis*[1], Hartmut Klauck†[2], and Debbie Lim‡[1]

[1]Center for Quantum Computing Science, Faculty of Sciences and Technology, University of Latvia, Latvia
[2]Centre for Quantum Technologies, Singapore

October 6, 2025

Abstract

Similarly to the Chomsky hierarchy, we offer a classification of communication complexity measures such that these measures are organized into equivalence classes. Different from previous attempts of this endeavor, we consider two communication complexity measures as equivalent, if, when one is constant, then the other is constant as well, and vice versa. Most previous considerations of similar topics have been using polylogarithmic input length as a defining characteristic of equivalence. In this paper, two measures $\mathcal{C}_1, \mathcal{C}_2$ are *constant-equivalent*, if and only if for all total Boolean (families of) functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ we have $\mathcal{C}_1(f) = O(1)$ if and only if $\mathcal{C}_2(f) = O(1)$. We identify five equivalence classes according to the above equivalence relation. Interestingly, the classification is counter-intuitive in that powerful models of communication are grouped with weak ones, and seemingly weaker models end up on the top of the hierarchy.

1 Introduction

Communication complexity, as a subfield of theoretical computer science emerged from the seminal works of [1, 2, 3, 4] more than four decades ago. Several other early works include Refs. [5, 6, 7]. In this field, we are able to show lower bound for explicit problems, with a plethora of applications to other computational models such as finite automata, Turing machines, decision trees, ordered binary decision diagrams, VLSI chips, streaming algorithms, networks of threshold gates and data structures [8, 9, 10, 2, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20].

In the theory of formal languages, the Chomsky hierarchy [21] is a central concept. This hierarchy classifies formal languages according to the type of computational model that can recognize them. The levels of the Chomsky hierarchy can be labelled by a machine model that is able to recognize the languages from the corresponding level, i.e., Finite Automata, Context-Free Grammars, Context-Sensitive Grammars, and Turing Machines (acting as the most powerful formal computation model). The Chomsky hierarchy also classifies formal languages into one of its levels. For instance, the language $\{0^n 1^n\}$ (understood as a family over all n) is in the class of context-free languages, but not the class of regular languages (as defined by DFA's). Nondeterministic and deterministic finite automata on the other hand define the same class of regular languages, even though their state-complexity can be exponentially far apart.

In this paper, we attempt a similar classification for communication complexity. Technically, communication complexity is usually established for a *family* of functions $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with all possible input lengths. The complexity of such a family is *constant* if the needed

*andris.ambainis@lu.lv
†hklauck@gmail.com
‡limhueychih@gmail.com

communication does not increase with the input length. A trivial example of this is the decision if the first bits of Alice's and Bob's input are both 1, which a deterministic protocol can decide with 1 bit of communication. A less trivial problem is the Equality function, in which Alice and Bob each receive a string of n bits and their task is to decide whether the two strings are the same. While a deterministic protocol cannot do better than communicating the whole length of a string, a public coin protocol can solve this problem with a single bit of communication and satisfactory error probability.

Prior to our work, there have been several studies on establishing hierarchies in communication complexity. In a seminal paper, Babai, Frankl, and Simon [3] proposed polylogarithmic complexity as the criteria for efficiency and used it to define communication classes analogous to the classical computational complexity classes. Their definition of communication classes provides a structured framework for comparing the powers of various communication models. While this work has undoubtedly led to the development in proving separations between different communication classes, it is also crucial to consider communication problems that have uniformly bounded complexity. In fact, the idea of constant-cost communication classes began with the works of Hambardzumyan et al. [22] and Harms et al. [23] studying constant-cost communication problems and the notion of "classes" was used in prior works such as Refs. [24] and [25]. Other works on constant-cost communication include [26, 27, 28, 29, 30, 31, 32, 33, 34]. The recent survey paper of Hatami and Hatami [35] studies the constant-cost analogues of the communication classes by Babai, Frankl, and Simon [3], where the criterion of effectiveness is a $O(1)$ complexity, independent of the input size n. Another related work is that of Göös, Pitassi and Watson [36]. In their work, the authors outlined the relationships between a wide range of models for two-party communication, clarifying which are stronger, weaker, or incomparable. By establishing new separations and inclusions, they resolves several long-standing open questions and gives a nearly complete structural picture of the complexity landscape in this area.

Main contribution. We establish a hierarchy of increasingly powerful classes of communication complexity, based on what these computational models can do with constant communication. While our investigation is based on the constant-equivalence of different measures of communication complexity (for instance nondeterministic and deterministic communication are constant-equivalent), this also leads to a hierarchy of (families of) communication problems. Namely, the problems that can be solved by an equivalence class of communication complexity measure. For instance, the Equality problem is in the class of cheap public coin protocols (see below), while the Greater-Than problem is in the class of low sign-rank protocols [29].

We identify the following levels of our hierarchy, named here by the simplest way to define them (more details in Figure 1):

- Class 1: Only a constant number of distinct rows in the communication matrix.
- Class 2: The γ_2 norm of the communication matrix is constant.
- Class 3: The approximate γ_2 norm, γ_2^α, of the communication matrix is constant which implies that information complexity and also the commonly used randomized/quantum communication complexity lower bounds such as the rectangle bound (with error) and the discrepancy bound are also constant.
- Class 4: The sign-rank and the (truly) unbounded error communication complexity are constant.
- Class 5: Measures defined in terms of product-distribution communication complexity are constant.

We provide equivalence results for all the measures in the respective classes, and separating functions for the different levels of the hierarchy. While most of this work is collecting existing results, we have the following new result:

Result 1. *Let f be a total Boolean function and denote by $R_\epsilon^{\text{pub}}(f)$ the public-coin randomized communication complexity with oblivious and exact error (see Section 3.2). Then we have*

$$\log \gamma_2 \leqslant R_{=\epsilon}^{\text{pub}}(f) \leqslant O((\gamma_2(f))^2).$$

We summarize our results in Figure 1.

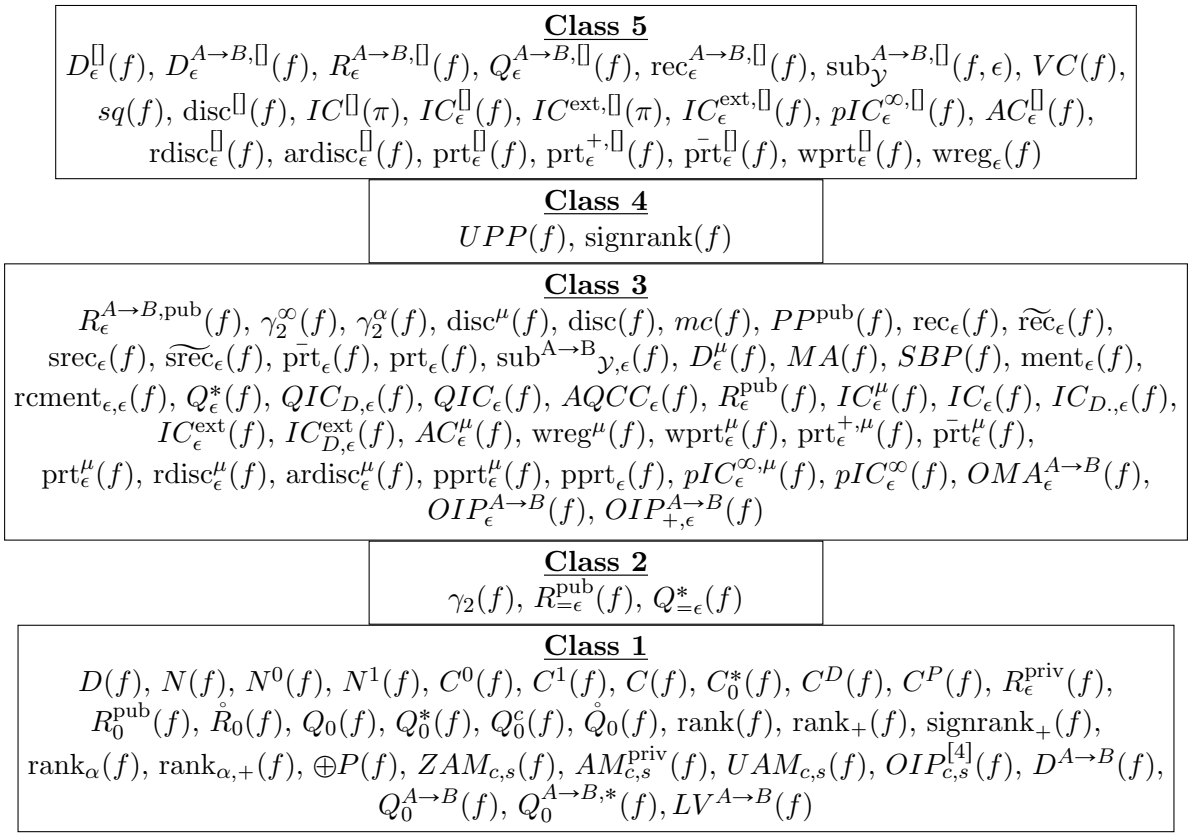


Figure 1: Summary of results. We categorize communication complexity measures in to five increasingly powerful classes.

Outline. This paper is organized as follows: In Section 2, we define the communication model, the necessary notations and results that will be used to prove our main theorems. We present our main results in Section 3, which include graphs illustrating the constant equivalence relationship among all complexity measures in every class, followed by separation results between consecutive classes in Section 4. In Section 5, we conclude our work, discuss some open problems and make several conjectures.

2 Preliminaries and notations

We consider Yao's two-party communication complexity model [1]. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be arbitrary finite sets and let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be an arbitrary function. Two players, Alice and Bob, who are assumed to have unlimited computational power, are each given inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively. Each player knows only their input and the task is to evaluate $z = f(x, y)$ by alternately send messages to each other. When $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is a relation, the task translates into determining an element $z \in \mathcal{Z}$ such that $(x, y, z) \in f$.

We let X, Y denote the random variables from $\mathcal{X}, \mathcal{Y}$ respectively. In direct sum/product theorems, we define $f^k : \mathcal{X}^k \times \mathcal{Y}^k \to \mathcal{Z}^k$ to be the concatenation of the evaluations

$$f^k(x_1, \cdots, x_k, y_1, \cdots, y_k) \coloneqq (f(x_1, y_1), \cdots, f(x_k, y_k)).$$

We write μ^k to denote the distribution on k inputs, where each is sampled according to μ independently.

All functions considered in this work are total Boolean functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, unless stated otherwise. Sometimes, we use $f : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ instead. For a function f, let M_f denote its corresponding communication matrix, i.e. $M_{x,y} = f(x,y)$ for all $x, y \in \{0,1\}^n$. We use f instead of M_f as the argument of the complexity measure when there is no ambiguity. For any total Boolean function f, we use $D(f), N^1(f), N^0(f), R_\epsilon^{\text{priv}}(f), R_\epsilon^{\text{pub}}(f)$ to denote the deterministic communication complexity, nondeterministic communication complexity, co-nondeterministic communication complexity, private coin ϵ-error randomized communication complexity and public coin ϵ-error randomized communication complexity. Readers are referred to Ref. [37] for background on classical communication complexity. For quantum communication complexity measures, we write $Q_0(f)$, $Q_\epsilon(f)$ to denote the exact (zero-error) quantum communication complexity and the ϵ-error quantum communication complexity respectively, both without prior entanglement. We denote by $*$ in superscript when players share prior entanglement. For all communication complexity measures, we write $A \to B$ in superscript to denote the one-way communication model.

To distinguish between constant and non-constant complexity in a non-uniform model of computation like communication complexity, we need to consider *families* of functions $f_1, f_2, \ldots, f_n, \ldots$ with one function $f_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ for every input size n. For example, the Equality problem is defined by taking a family of functions $EQ_1, EQ_2, \ldots, EQ_n, \ldots$ where $EQ_n(x,y)$ is a function from $\{0,1\}^n \times \{0,1\}^n$ to $\{0,1\}$ defined by $EQ_n(x,y) = 1$ iff $x = y$. Then, the public-coin randomized communication complexity of the Equality problem is $O(1)$, which means that this problem can be solved with constant communication in the described model, *no matter what the length of the input is*. On the other hand, the Disjointness problem cannot be solved in any model of communication complexity using constant communication, unless *all* problems can be (under reasonable assumptions) solved in that model at constant cost.

We give a quick review of some communication complexity measures, starting with nondeterministic communication complexity. Let π be a protocol and v be a node of the protocol tree. The set R_v of all inputs (x,y) that lead π to v is a *combinatorial rectangle*, where $R_v = A_v \times B_v$ for $A_v \subseteq \mathcal{X}$ and $B_v \subseteq \mathcal{Y}$. For every $z \in \{0,,1\}$, we say that a combinatorial rectangle R is a z-monochromatic rectangle if $f(x,y) = b$ for all $(x,y) \in R$. If v is a leaf with the label b, then R_v must be a z-monochromatic rectangle. Hence, the leaves of a deterministic protocol partitions its communication matrix into monochromatic rectangles. The protocol partition number of f, denoted as $C^P(f)$, is the minimum number of leaves in a protocol for f. The partition number of f, denoted as $C^D(f)$, is the minimum number of monochromatic rectangles in a partition of $\mathcal{X} \times \mathcal{Y}$. Moreover, the cover number of f, denoted as $C(f)$, is the minimum number of monochromatic rectangles required to cover $\mathcal{X} \times \mathcal{Y}$. For any $z \in \{0,1\}$, $C^z(f)$ is the number of monochromatic rectangles required to cover the z-inputs of f [37].

Variants of rank and factorization norm. Let $M \in \mathbb{R}^{n \times n}$ be a matrix. Denote by $\text{rank}(M)$ the linear *rank* of M over the field of reals. The *α-approximate rank* of M, denoted as $\text{rank}_\alpha(M)$, is defined as follows [38, Definition 1], [39]

$$\text{rank}_\alpha(f) := \min_{M' : 1 \leqslant M_{i,j} \cdot M'_{i,j} \leqslant \alpha} \text{rank}(M'),$$

where $\alpha \geqslant 1$. The *nonnegative rank* of a matrix $M \in \mathbb{R}_+^{n \times n}$, denoted as $\text{rank}_+(M)$, is the smallest r such that there exist matrices $A \in {\mathbb{R}_+}^{n \times r}$ and $B \in \mathbb{R}_+^{r \times n}$ such that $M = AB$. The *α-approximate nonnegative rank* of M, denoted as $\text{rank}_{\alpha,+}(M)$ is given by

$$\text{rank}_{\alpha,+}(M) := \min \left\{ \text{rank}_+(M') : M' \text{ is nonnegative, } \|M - M'\|_\infty \leqslant \alpha \right\}.$$

In certain cases, it makes sense to consider functions with ± 1 outputs, i.e. $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,1\}$ [26]. sign-rank is one of the important analytic notions in communication complexity.

The *sign-rank* of a matrix $M \in \{-1,1\}^{n\times n}$, denoted as signrank(M), is the minimal rank of a real matrix whose entries have the same sign pattern as M. More formally.

$$\text{signrank}(M) := \min\{\text{rank}(M') : M' \in \mathbb{R}^{n\times n}, M'_{ij}M_{ij} > 0 \text{ for all } i,j \in [n]\}.$$

The *nonnegative sign-rank* of a matrix $M \in \mathbb{R}_+^{m\times n}$, denoted as $signrank_+(M)$, is the smallest r such that there exists $A \in \mathbb{R}_+^{m\times r}$ and $B \in \mathbb{R}_+^{r\times n}$ such that AB has the same sign pattern as M.

The γ_2 norm is an important matrix norm initially developed in Banach Space theory. Linial and Shraibman [40] introduced this norm in the context of communication complexity and subsequently, this complexity measure and its approximate version found applications in communication complexity and other related fields such as discrepancy theory [41] and differential privacy [42, 43, 44]. Formally, the γ_2 *norm* of a matrix $M \in \mathbb{R}^{m\times n}$, denoted as $\gamma_2(M)$, is defined as

$$\gamma_2(M) = \min_{A,B:AB=M} \|A\|_r \|B\|_c, \tag{1}$$

where $\|A\|_r$ denotes the maximum ℓ_2 norm of a row of A and $\|B\|_c$ denotes the maximum ℓ_2 norm of a column of B. Alternatively, $\gamma_2(M)$ can be defined as

$$\gamma_2(M) = \max_{u,v:\|u\|_2=\|v\|_2=1} \|uv^T \circ M\|_{tr},$$

where $\circ$ and $\|\cdot\|_{tr}$ denotes the Hadamard product and trace norm respectively.

The α-*approximate* γ_2 *norm* of $M \in \{-1,1\}^{n\times n}$, where $\alpha \geqslant 1$, denoted as $\gamma_2^\alpha(M)$, is given by

$$\gamma_2^\alpha(M) = \min_{\{M':1\leqslant M_{ij}M'_{ij}\leqslant\alpha\}} \gamma_2(M').$$

In particular,

$$\gamma_2^\infty(M) = \min_{\{M':1\leqslant M_{ij}M'_{ij}\}} \gamma_2(M').$$

Information complexity. Before the notion of information complexity was formally defined, information-theoretic tools have already been used to prove communication complexity bounds. For example, Ablayev [10] used information theory to prove a lower bound for the Index problem, Nayak [45] used it in the quantum setting, and Klauck et al. [46] used it for pointer jumping. Interactive information complexity has been studied by information theorists starting with Orlitsky's work in the 1990s [47]. A paper from this line of work by Ma and Ishwar [48] gives an alternative way of looking at the interactive information complexity of problems and amortized communication.

From the perspective of information complexity, communication can be thought of as players wanting to reveal the minimum amount of information about their inputs in order to solve some pre-determined task. The internal information complexity measures the amount of information revealed to each player about the other player's input, whereas, the external information complexity measures the amount of information revealed about the inputs x, y to an external observer who only observes the transcript of the protocol. Intuitively, both of these measures should be lower bounds on the communication complexity of the protocol.

Given a protocol π, $T(X,Y)$ denotes the random variable of the transcript on inputs X, Y, which are also random variables. The *(internal) information cost* of a protocol π over all inputs from $\mathcal{X} \times \mathcal{Y}$ is given by

$$IC^\mu(\pi) := I(T(X,Y); X|Y) + I(T(X,Y); Y|X),$$

where $I(A; B|C)$ denotes the mutual information between random variables A, B, conditioned on C. A second measure of the information complexity of a communication protocol is its external

information cost. The *external information cost* of a protocol π over all inputs from $\mathcal{X} \times \mathcal{Y}$ is given by

$$IC_{\mu}^{\text{ext}}(\pi) := I(XY; T(X,Y)).$$

For a function f, the external information cost of f with respect to μ is defined to be the infimum of $IC_{\mu}^{\text{ext}}(\pi)$ over all protocols π that compute f on inputs from $\mathcal{X} \times \mathcal{Y}$ with probability larger than $2/3$ [49]. Clearly, this implies that the external information cost is always smaller than the communication complexity . Moreover, the external information cost is always greater than or equal to the internal information cost of a protocol [50].

3 Simulation results

In this section, we state our main results. We divide this section into five subsections, one subsection for every class. Each subsection starts with a graph showing the network of simulation results for the corresponding class, following by a main theorem and a list of facts and lemmas that are used in the main theorem.

The graphs in the subsequent subsections use the following notations. For any two complexity measures $\mathcal{C}_1, \mathcal{C}_2$, we use the arrow $\to$ to denote the "$\leqslant$" relation, i.e.,$\mathcal{C}_1(f) \to \mathcal{C}_2(f)$ denotes $\mathcal{C}_1(f) \leqslant O(g(\mathcal{C}_2(f)))$ for some function $g : \mathbb{R} \to \mathbb{R}$. Moreover, we use $\leftrightarrow$ between $\mathcal{C}_1, \mathcal{C}_2$ to denote that both complexity measures are equivalent, i.e. $\mathcal{C}_1 \leftrightarrow \mathcal{C}_2$ denotes $\mathcal{C}_1(f) = \Theta(h(\mathcal{C}_2(f)))$ for some function $h : \mathbb{R} \to \mathbb{R}$.

3.1 Class 1

This class relates the constant equivalence between deterministic communication complexity, nondeterministic communication complexity, several variants of randomized communication complexity, quantum communication complexity, variants of Arthur-Merlin complexities, rank, parity complexity and sampling complexity. The constant equivalence relationship between complexity measures in Class 1 is summarized in Figure 2.

Theorem 1. *Class 1 includes the following complexity measures:*

- $D(f)$
- $N(f)$
- $N^0(f)$
- $N^1(f)$
- $C^0(f)$
- $C^1(f)$
- $C(f)$
- $C_0^*(f)$
- $C^D(f)$
- $C^P(f)$
- $R_{\epsilon}^{\text{priv}}(f)$
- $R_0^{\text{pub}}(f)$
- $\mathring{R}_0(f)$
- $Q_0(f)$
- $Q_0^*(f)$
- $Q_0^c(f)$
- $\mathring{Q}_0(f)$
- $\text{rank}(f)$
- $\text{rank}_+(f)$
- $\text{signrank}_+(f)$
- $\text{rank}_{\alpha}(f)$
- $\text{rank}_{\alpha,+}(f)$
- $\oplus P(f)$
- $ZAM_{c,s}(f)$
- $AM_{c,s}^{\text{priv}}(f)$
- $UAM_{c,s}(f)$
- $OIP_{c,s}^{[4]}(f)$
- $D^{A\to B}(f)$
- $LV^{A\to B}(f)$
- $Q_0^{A\to B}(f)$
- $Q_0^{A\to B,*}(f)$

Proof. We prove the cyclic relationship of the form $\mathcal{C}_1 \to \mathcal{C}_2 \to \mathcal{C}_3 \to \cdots \to \mathcal{C}_1$ for all complexity measures in this class.

We have $C(f) \to C^D(f)$ by Fact 11; $C^D(f) \to C^P(f)$ by Fact 11; $C^P(f) \to D(f)$ by Fact 11; $D(f) \to Q_0^*(f)$ by Facts 1, 3; $Q_0^*(f) \to Q_0(f)$ by Fact 2; $Q_0(f) \to \text{rank}(f)$ by Facts 2, 1; $\text{rank}(f) \to \text{rank}_{\alpha}(f)$ by Fact 22; $\text{rank}_{\alpha}(f) \to \text{rank}_{\alpha,+}(f)$ by Fact 23; $\text{rank}_{\alpha,+}(f) \to R_{\epsilon}^{\text{priv}}(f)$ by Fact 24.

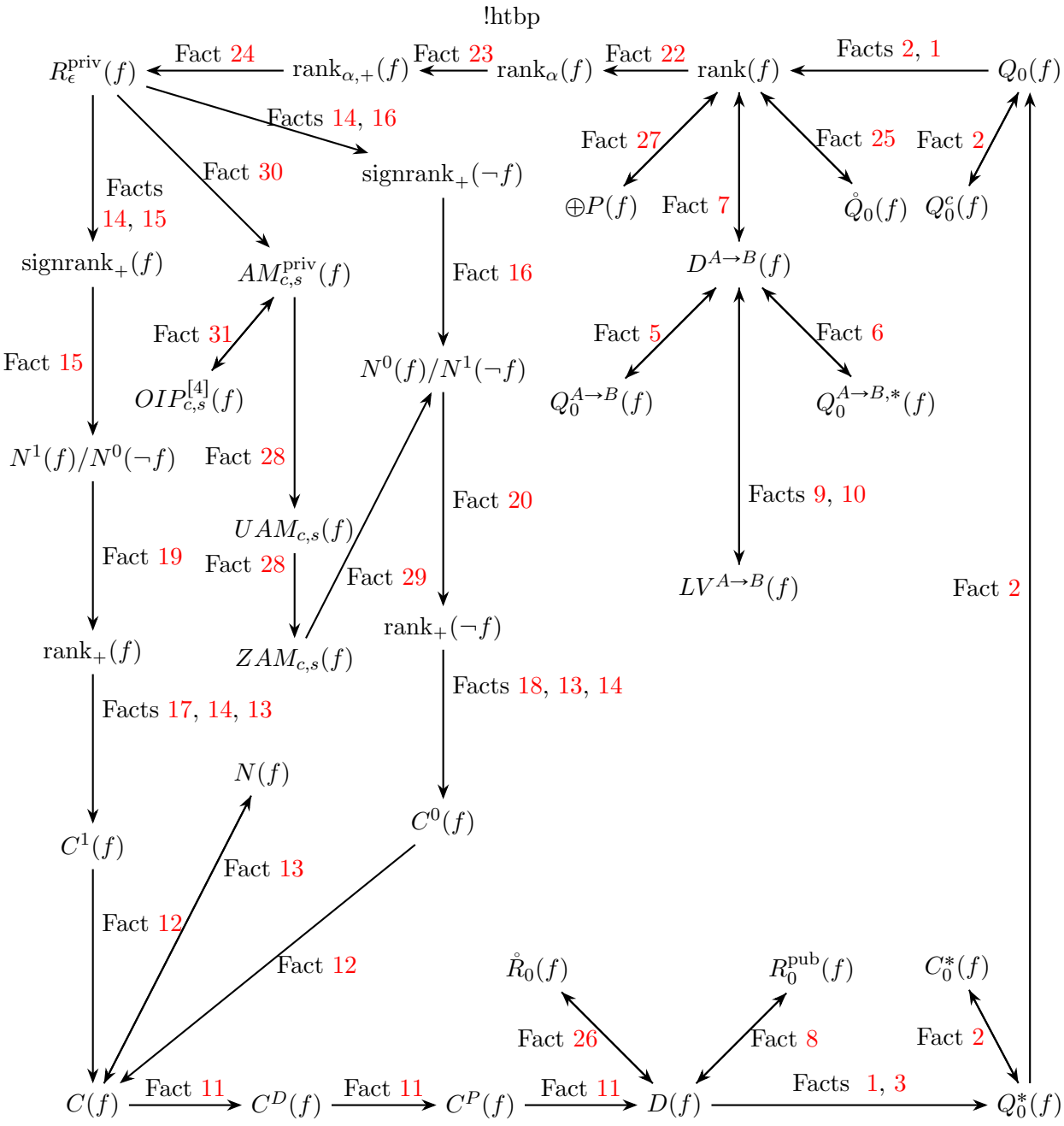


Figure 2: Relationship between complexity measure in Class 1.

The relation $R_\epsilon^{\text{priv}}(f) \to C(f)$ can be detailed separately in terms of 0- and 1-inputs. First, note that a 0-input for the function f is equivalent to a 1-input for the function $\neg f$. Using the fact that $R_\epsilon^{\text{priv}}(f) \leqslant D(f)$ and Facts 14, 15, 16, we jointly obtain $R^{\text{priv}_\epsilon}(f) \to \text{signrank}_+(\neg f)$ and $R^{\text{priv}_\epsilon}(f) \to \text{signrank}_+(f)$. From $\text{signrank}_+(\neg f)$, we have $\text{signrank}_+(\neg f) \to N^0(f)$ (in other words, $\text{signrank}_+(\neg f) \to N^1(\neg f)$) by Fact 16; $N^0(f) \to \text{rank}_+(\neg f)$ by Fact 20; $\text{rank}_+(f) \to C^0(f)$ by Facts 18, 13 ,14 (the direct relation $N^0(f) \to C^0(f)$ due to Fact 13). On the other hand, from $\text{signrank}_+(f)$, we have $\text{signrank}_+(f) \to N^1(f)$ (in other words, $\text{signrank}_+(f) \to N^0(\neg f)$)by Fact 15; $N^1(f) \to \text{rank}_+(f)$ by Fact 19; $\text{rank}_+(f) \to C^1(f)$ by Fact 13, 14, 17 (the direct relation $N^1(f) \to C^1(f)$ due to Fact 13). Finally, $C^0(f) \to C(f)$ and $C^1(f) \to C(f)$ due to Fact 12 and also $N(f) \leftrightarrow C(f)$ due to Fact 13.

Furthermore, $R_\epsilon^{\text{priv}}(f) \to AM_{c,s}^{\text{priv}}(f)$ by Fact 30; $AM_{c,s}^{\text{priv}}(f) \to UAM_{c,s}(f)$ by Fact 28; $UAM_{c,s}(f) \to ZAM_{c,s}(f)$ by Fact 28; $ZAM_{c,s}(f) \to N^0(f)$ by Fact 29.

Additionally , $AM_{c,s}(f) \leftrightarrow OIP^{[4]}_{c,s}(f)$ by Fact 31; $\mathrm{rank}(f) \leftrightarrow \oplus P(f)$ by Fact 27; $\mathrm{rank}(f) \leftrightarrow \mathring{Q}_0(f)$ by Fact 25; $\mathrm{rank}(f) \leftrightarrow D^{A\to B}(f)$ by Fact 7; $D^{A\to B}(f) \leftrightarrow Q_0^{A\to B}(f)$ by Fact 5; $D^{A\to B}(f) \leftrightarrow Q_0^{A\to B,*}(f)$ by Fact 6, $Q_0(f) \leftrightarrow Q_0^c(f)$ by Fact 2; $Q_0^*(f) \leftrightarrow C_0^*(f)$ by Fact 2; $D(f) \leftrightarrow R_0^{\mathrm{pub}}(f)$ by Fact 8; $D(f) \leftrightarrow \mathring{R}_0(f)$ by Fact 26; $D^{A\to B}(f) \leftrightarrow LV^{A\to B}(f)$ by Facts 9, 10. □

We now describe the facts used in the proof of Theorem 1. We start with an upper bound result on $D(f)$ in terms of $\mathrm{rank}(f)$ by Lovett.

Fact 1 ([51, Theorem 1.1]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$ be a Boolean function. Then, $D(f) \leqslant O(\sqrt{\mathrm{rank}(f)} \log \mathrm{rank}(f))$.*

For a function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, let $C_0^*(f)$ denote the (worst-case) cost of a protocol that computes f exactly, which communicates classical bits but is allowed to make use of an unlimited (but finite) number of shared EPR-pairs. $Q_0^c(f)$ denotes the (worst-case) cost of a clean qubit protocol without prior entanglement that compute f exactly, i.e. a protocol that starts with $|0\rangle|0\rangle|0\rangle$ and ends with $|0\rangle|f(x, y)\rangle|0\rangle$ [52]. Some simple relations that hold between these measures are as follows:

Fact 2 ([52], [53]). *For any function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$,*

1. $Q_0^*(f) \leqslant Q_0(f) \leqslant D(f) \leqslant D^{A\to B}(f)$,
2. $Q_0(f) \leqslant Q_0^c(f) \leqslant 2Q_0(f)$ *and*
3. $Q_0^*(f) \leqslant C_0^*(f) \leqslant 2Q_0^*(f)$.

As first mentioned by Buhrman, and Wigderson [54] and Ambainis *et al.* [55], the methods from Refs.[56, 57] imply that $Q(f) \in \Omega(\log \mathrm{rank}(f))$. Buhman and de Wolf [52] then proved that the $\log \mathrm{rank}(f)$ lower bound holds for exact protocols.

Fact 3 ([52, Theorem 2]). *For any function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, $Q_0^*(f) \geqslant \frac{\log \mathrm{rank}(f))}{2}$.*

Other relations relating zero error quantum communication complexity with deterministic communication complexity are listed below.

Fact 4 ([58, Theorem 5.11]). *For all total functions f, we have $Q_0(f) = D(f)$.*

Fact 5 ([59, Theorem 4]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$. Then, $Q_0^{A\to B}(f) = D^{A\to B}(f)$.*

Fact 6 ([59, Theorem 5]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$. Then, $Q_0^{A\to B,*}(f) = \lceil D^{A\to B}(f)/2 \rceil$.*

Fact 7 ([60, Equation 3]). *Let $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$. Then, $\log \mathrm{rank}(f) \leqslant D^{A\to B}(f) \leqslant \mathrm{rank}(f)$.*

A result of David *et al.* [61] shows that the largest gap between public-coin zero-error randomized communication complexity and $D(f)$ is quadratic. This, together the trivial bound of $R_0^{\mathrm{pub}}(f) \leqslant D(f)$ gives the fact below.

Fact 8 ([61, Theorem 2.1]). *For every $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, we have $\Omega\left((D(f))^{1/4}\right) \leqslant R_0^{\mathrm{pub}}(f) \leqslant D(f)$.*

The Las Vegas communication complexity, denoted by $LV(f)$, is the communication complexity of a public-coin randomized protocol that does not err and either accept, rejects or says "don't know". The one-way Las Vegas communication complexity $LV^{A\to B}(f)$ is trivially upper bounded by $D^{A\to B}(f)$ and lower bounded by $D^{A\to B}(f)/2$.

Fact 9 ([62, Theorem 2.1]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a Boolean function. Then, $LV^{A\to B}(f) \geqslant D^{A\to B}(f)/2$.*

Fact 10. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. Then, $LV^{A\to B}(f) \leqslant D^{A\to B}(f)$.*

Turning our attention to partition, cover numbers and nondeterministic communication complexity, we have a list of bounds among them.

Fact 11 ([37, Proposition 2.2])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, then $C(f) \leqslant C^D(f) \leqslant C^P(f) \leqslant 2^{D(f)}$.*

Fact 12 ([37, Proposition 2.2])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, then $C(f) = C^0(f) + C^1(f)$.*

Fact 13 ([37, Definition 2.3])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. For $z \in \{0,1\}$, $N^z(f) = \log C^z(f)$. Moreover, $N(f) = \log C(f)$.*

Fact 14 ([37, Theorem 2.11])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, then $D(f) \leqslant O(N^0(f)N^1(f))$.*

Deaett and Srinivasan [63] showed that $N^1(f)$ can be tightly bounded by the nonnegative sign-rank, and this bound is tight up to constant factors.

Fact 15 ([63, Corollary 3.9])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $\lceil \log \operatorname{signrank}_+(f) \rceil \leqslant N^1(f) \leqslant \lceil \log \operatorname{signrank}_+(f) \rceil + 2$.*

An analogous result naturally holds between $N^0(f)$ and $\operatorname{signrank}_+(\neg f)$. We will be stating the $\neg f$ versions of certain results for cross-referencing purposes in our main results.

Fact 16. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $\lceil \log \operatorname{signrank}_+(\neg f) \rceil \leqslant N^0(f) \leqslant \lceil \log \operatorname{signrank}_+(\neg f) \rceil + 2$.*

Another variant of rank is the nonnegative rank, $\operatorname{rank}_+(f)$. It is clear that $\operatorname{rank}_+(f)$ is at least as large as $\operatorname{rank}(f)$. The nonnegative rank can be arbitrarily larger than the rank in the context of non-Boolean matrices. It was shown in that for every $k \in \mathbb{N}$, there exists a matrix M such that $\operatorname{rank}(M) = 3$ and $\operatorname{rank}_+(M) \geqslant k$ [64]. However, no such separation between $\operatorname{rank}(M)$ and $\operatorname{rank}_+(f)$ is known when the matrix is restricted to being Boolean. For a Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, we have $D(f) \leqslant O((\log \operatorname{rank}_+(f) + 1)(\log \operatorname{rank}_+(\neg f) + 1))$ and $D(f) \leqslant O(\log^2 r \operatorname{ank}_+(f))$ [6],[65, Theorem 2] as upper bounds, while the lower bound is as below.

Fact 17 ([66])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $\log \operatorname{rank}_+(f) \leqslant D(f)$.*

Fact 18. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $\log \operatorname{rank}_+(\neg f) \leqslant D(\neg f)$.*

Fact 19 ([67, Corollary], [66, Theorem 3.4])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $N^1(f) \leqslant \log \operatorname{rank}_+(f)$.*

Fact 20. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $N^0(f) \leqslant \log \operatorname{rank}_+(\neg f)$.*

In the following lemma, we show that if the nonnegative rank of a matrix is small, then it follows that its nonnegative sign-rank is small.

Lemma 1. *Let $M \in R_+^{m\times n}$ be a nonnegative matrix. Then $\operatorname{rank}_+(M) \geqslant \operatorname{signrank}_+(M)$.*

Proof. Suppose that $M \in \mathbb{R}^{m\times n}$ is a nonnegative matrix. Let $k = \operatorname{rank}_+(M)$. Then, there exist nonnegative matrices $A \in \mathbb{R}_+^{m\times p}$ and $B \in \mathbb{R}_+^{p\times n}$ such that $M = AB$. It follows that AB has the same sign pattern as M and hence, $\operatorname{signrank}_+(M) \leqslant k$. □

Fact 21. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $\operatorname{rank}_+(f) \geqslant \operatorname{signrank}_+(f)$ and $\operatorname{rank}_+(\neg f) \geqslant \operatorname{signrank}_+(\neg f)$.*

Recall the definition of nonnegative approximate sign-rank from Section 2. Similarly to the separation between $\operatorname{rank}(f)$ and $\operatorname{rank}_+(f)$ in non-Boolean matrices, one can show that for every $k \in \mathbb{N}$, there exists a matrix M and $\alpha > 0$ such that $\operatorname{rank}_\alpha(M) \leqslant 3$ and $\operatorname{rank}_{\alpha,+}(M) \geqslant k$ [65]. The following shows how this complexity measure relates to different variants of $\operatorname{rank}(f)$ and private-coin randomized communication complexity.

Fact 22 ([38, Corollary 1]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$ be a Boolean function and let $\alpha > 1$ be a constant. Then,* $\mathrm{rank}_\alpha(f) \geqslant \Omega(\log \mathrm{rank}_+(f)) \geqslant \Omega(\log \mathrm{rank}(f))$

Fact 23 ([65]). *For any nonegative matrix M, we have* $\mathrm{rank}_\alpha(M) \leqslant \mathrm{rank}_{\alpha,+}(M)$.

Fact 24 ([39]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$. Then, $R_\epsilon^{\mathrm{priv}}(f) \geqslant \log \mathrm{rank}_{\alpha,+}(f)$.*

Ambainis, Schulman, Ta-Shma, Vazirani and Wigderson introduced the classical and quantum sampling complexity [68]. Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ and let $\mathcal{D}$ be any distribution on $\mathcal{X} \times \mathcal{Y}$. We say that a protocol samples f according to $\mathcal{D}$ with error ϵ if the distribution the protocol induces on $\{(x, y, z)\}$ is ϵ-close in the total variation distance to the distribution $(\mathcal{D}, f(\mathcal{D}))$ obtained by first choosing (x, y) according to $\mathcal{D}$ and then evaluating $f(x, y)$. The number of bits (resp. qubits) needed for a randomized (resp. quantum) protocol to sample f according to $\mathcal{D}$ with ϵ error is denoted as $\mathring{R}_\epsilon(f, \mathcal{D})$ (resp. $\mathring{Q}_\epsilon(f, \mathcal{D})$). When $\mathcal{D}$ is the uniform distribution, we write $\mathring{R}_\epsilon(f)$ (resp. $\mathring{Q}_\epsilon(f)$) [68].

Fact 25 ([68, Corollary 8.3]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, Then, $\mathring{Q}_0(f) = \Theta(log\, \mathrm{rank}(f))$.*

Fact 26 ([68, Theorem 8.4]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, Then, $\sqrt{D(f)} \leqslant \mathring{R}_0(f) \leqslant D(f)$.*

The $\oplus P$ notion was first introduced by Papadimitriou and Zachos [69]. In the context of communication complexity, $\oplus P$ is defined as follows: Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a Boolean function. Given a cost-k nondeterministic protocol for f and its corresponding collection of rectangles $\{R_v : v \in \{0, 1\}^k\}$, where R_v denotes the rectangle of inputs arriving at v in the protocol. The protocol outputs 1 if the number of (x, y)'s such that $(x, y) \in R_v$ is odd. The parity complexity of a function f, denoted as $\oplus P(f)$, is the total length k of the accepting path in the protocol tree. A few results are known about the non-inclusion of $\oplus P(f)$ in communication complexity classes [70, 36].

Fact 27 ([36, Observation B.33]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, Then, $\oplus P(f) \in \log rank(f) \pm O(1)$.*

Now we shall turn our attention to interactive proofs. In the interactive proof model of communication complexity, there are three computationally unbounded parties: Merlin, Alice and Bob. Both Alice and Bob can see only their input, while Merlin sees both Alice's and Bob's input. Merlin is the prover, who wants to convince the verifier, consisting of Alice and Bob together, that $f(x, y) = 1$. We formally define the Arthur-Merlin complexity. Consider a function $f : \mathcal{X} \times \mathcal{Y} \to Z$. In a private coin Arthur-Merlin (AM) protocol, Alice is given a uniform sample from some finite set $\mathcal{R}$, Bob is given a uniform sample from some other independent finite set $\mathcal{Q}$, and there is a collection of proofs $R_1, \cdots, R_m \subseteq (\mathcal{X} \times \mathcal{R}) \times (\mathcal{Y} \times \mathcal{Q})$. The acceptance probability of the protocol on input (x, y) is defined to be $\mathbb{P}_{r \in \mathcal{R}, q \in \mathcal{Q}}[\exists i : ((x, r), y, q) \in R_i]$. The index i of a rectangle R_i represents a message sent from Merlin to Alice and Bob, who then separately decides whether to accept or not. The output of the protocol is 1 if and only if both Alice and Bob accept. The communication cost of the protocol is the length of Merlin's proof. The protocol has completeness c and soundness s if it accepts 1-input with probability at least c and accepts 0-inputs with probability at most s, We define $AM_{c,s}^{\mathrm{priv}}(f)$ to be the minimum cost over all AM protocols for f with completeness c and soundness s. In short, an Arthur-Merlin (AM) protocol is a probability distribution over nondeterministic protocols, together with a bounded-error acceptance condition [71].

An AM protocol is *unambiguous* if, for every 1-input and every outcome of the randomness, there is at most one proof of Merlin that causes the players to accept. More formally, if $(x, y \in f^{-1}(1))$, $r \in \mathcal{R}$, $q \in \mathcal{Q}$ and $i \neq j$, it holds that $((x, r), (y, q)) \notin R_i \cap R_j$ and $i \neq j$. We write $UAM_{c,s}(f)$ to denote the minimum cost over all UAM protocols for f with completeness c and soundness s.

A UAM protocol is a zero-information AM (ZAM) protocol when the distribution of Merlin's unique proof is identical across all 1-inputs. In other words, an unambiguous AM protocol

can be viewed as a mapping from an input $((x,r),(y,q))$ where $(x,y) \in f^{-1}(1)$ to the unique $i \in \{1,\cdots,m\}$ such that $((x,r),(y,q)) \in R_i$, or to $\perp$ if no such i exists [71]. We use $ZAM_\epsilon(f)$ to denote the minimum cost over all ZAM protocols with completeness c and soundness s as $ZAM_{c,s}(f)$.

Fact 28 ([71]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ then $AM^{\text{priv}}_{1,1/2}(f) \leqslant UAM_{1,1/2}(f) \leqslant ZAM_{1,1/2}(f)$.*

Fact 29 ([71, Theorems 2 & 6]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ then $\Omega\left(N^0(f)\right) \leqslant ZAM_{1,1/2}(f) \leqslant O\left(2^{N^0(f)}\right)$.*

Proving lower bounds on $AM(f)$ has been notoriously known to be hard [72, 73, 74]. The only known $AM(f)$ lower bounds follow from the observation that $AM(f) \geqslant \Omega(\log R^{\text{priv}}_\epsilon(f))$ for all f [36].

Fact 30 ([36]). *For all f, we have $AM^{\text{priv}}_{1,1/2}(f) \geqslant \Omega(\log R^{\text{priv}}_\epsilon(f))$.*

Chakrabarti et al. [75] gave a hierarchy of communication models, called Online Interactive Proofs (OIPs). These models consist of $OMA^{[k]}$, $OIP^{[k]}$ and $OIP^{[k]}_+$ for $k \geqslant 1$, which are described as follows: In each model, Alice and Bob first toss some hidden coins (these coins are shared between Alice and Bob, but are not known to Merlin). Upon receiving the input,

(i) Merlin and Bob communicate for k rounds, with Merlin being the last to communicate at the end of the k rounds.

(ii) Alice sends Bob a message, randomized using public coins that are hidden from Merlin.

Then, Bob produces an output in $\{0,1\}$. The difference between the three models are as follows.

- $OMA^{[k]}$: (i) happens before (ii) and Bob can only look at his input after communicating with Merlin.
- $OIP^{[k]}$: (i) happens before (ii) and Bob may choose to look at his input before communicating with Merlin.
- $OIP^{[k]}_+$: Similar to $OIP^{[k]}$, with the difference being that (ii) happens before (i). In this case, Bob's message may depend on Alice's actual message to Bob, not just on his own input and the public coins.

The soundness and completeness for all three models are defined similarly as that of the AM model. The communication cost of a protocol that computes a function f is the sum of the maximum number of bits sent by Alice, the private coin and the number of bits communicated between Bob and Merlin. The corresponding communication complexity is the cost of the cheapest protocol that computes f with soundness s and completeness c. We shall denote them as $OMA^{[k]}_{c,s}(f)$, $OIP^{[k]}_{c,s}(f)$ and $OIP^{[k]}_{+,c,s}(f)$.

Fact 31 ([75, Corollary 5.14]). *For all $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, we have $\Omega(AM_{2/3,1/3}(f)^c) \leqslant OIP^{[4]}_{2/3,1/3}(f) \leqslant O(AM_{2/3,1/3}(f) \log AM_{2/3,1/3}(f))$.*

3.2 Class 2

This class contains three complexity measures. Namely, the γ_2 norm, the randomized public coin communication complexity with exact error and the quantum communication complexity with prior entanglement and exact error. The equivalence relationship between these complexity measures is depicted in Figure 3.

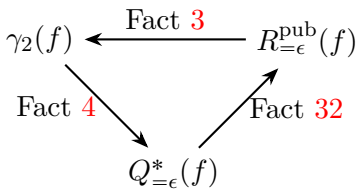


Figure 3: Relationship between complexity measures in Class 2.

Theorem 2. *Class 2 includes the following complexity measures:*

- $\gamma_2(f)$
- $R^{\text{pub}}_{=\epsilon}(f)$
- $Q^*_{=\epsilon}(f)$

Proof. We have $Q^*_{=\epsilon}(f) \to R^{\text{pub}}_{=\epsilon}(f)$ by Fact 32; $R^{\text{pub}}_{=\epsilon}(f) \to \gamma_2(f)$ by Fact 3; $\gamma_2(f) \to Q^*_{=\epsilon}(f)$ by Fact 4. □

We now describe the facts used in the proof of Theorem 2. Define oblivious-error public-coin randomized communication complexity as the usual public-coin randomized communication complexity with the additional requirement that the error is exactly the same for all inputs x, y. In other words, let $\epsilon > 0$ be the global error such that all 1-inputs are accepted with probability exactly $1-\epsilon$ and all 0-inputs are accepted with probability exactly ϵ. For more flexibility, one can also define different errors for 1- and 0-inputs[1]. For instance, the standard public coin protocol for the Equality problem has error 0 on 1-inputs and error $1/2$ on 0-inputs. We denote the oblivious-error randomized public coin communication complexity of a function f as $R^{\text{pub}}_{=\epsilon}(f)$.

We first state an obvious fact. This fact holds since we can replace classical transformations with their corresponding quantum counterpart. The public coin feature in randomized communication complexity corresponds to both players measuring the quantum state $\sum_i \sqrt{p(i)}\,|i\rangle$ in the same basis.

Fact 32. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $Q^*_{=\epsilon}(f) \leqslant R^{\text{pub}}_{=\epsilon}(f)$.*

In the following, we prove that $R^{\text{pub}}_{=\epsilon}(f)$ is at least as large as the log of $\gamma_2(f)$.

Lemma 2. $R^{\text{pub}}_{=\epsilon}(f) \geqslant \log \gamma_2(f)$.

Proof. Let $\epsilon > 0$ be a small constant. Consider a c-round public-coin randomized communication protocol that accepts 0-inputs with probability ϵ and accepts 1-inputs with probability $1-\epsilon$. Let $M \in \{\epsilon, 1-\epsilon\}^{|\mathcal{X}|\times|\mathcal{Y}|}$ be the matrix whose entries $M_{x,y}$ correspond to the acceptance probabilities of the protocol on input (x,y) for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. On the other hand, consider an optimal quantum protocol with shared entanglement that has cost T, such that the protocol accepts 0-inputs with probability at most ϵ and accepts 1-inputs with probability at least $1-\epsilon$. Let $M' \in [0,1]^{|\mathcal{X}|\times|\mathcal{Y}|}$ be the corresponding matrix of acceptance probabilities. Notice that for two sets of vectors $U \in \{\epsilon, 1-\epsilon\}^n$ and $V = \{v \in [0,1]^n | v(i) \in [0,\epsilon] \cup [1-\epsilon, 1], \forall i \in [n]\}$, the following is true:

$$\max_{u\in U} \|u\|_2 \leqslant \max_{v\in V} \|v\|_2.$$

Therefore,

$$\gamma_2(M) = \min_{A,B:AB=M} \|A\|_r \|B\|_c \leqslant \min_{A',B':A'B'=M'} \|A'\|_r \|B'\|_c \leqslant 2^T \leqslant 2^c,$$

where the first inequality is due to [40, Lemma 12] and the second inequality is due to Fact 32. □

[1] More specifically, there exists $\epsilon_0, \epsilon_1 > 0$ such that 1-inputs are accepted with probability exactly $1-\epsilon_1$ and 0-inputs are accepted with probability exactly ϵ_0. Nevertheless, both measures are equivalent.

We also prove that $R^{\text{pub}}_{=\epsilon}(f)$ is polynomially bounded from above by $\gamma_2(f)$.

Lemma 3. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $R^{\text{pub}}_{=\epsilon}(f) \leqslant O((\gamma_2(f))^2)$.*

Proof. We consider the definition of $\gamma_2(f)$ as in Equation (1). Let u_x, v_y be vectors of the optimal γ_2 decomposition of the communication matrix M_f such that $\|u_x\|_2, \|v_y\|_2 \leqslant \gamma_2(M_f)$. Given $x \in \mathcal{X}, y \in \mathcal{Y}$, Alice and Bob normalize their inputs to obtain u_x, v_y and use a randomized public coin protocol [76, Theorem 4.4] with error at most $1/(10\gamma_2(f))$ to compute the inner product $\langle u_x, v_y \rangle$. This inner product is either -1 or 1 before normalization[2], so Alice and Bob can decide which is the case. Note that in this protocol, the error only depends on the absolute value of the inner product, since

$$|\langle u_x, v_y \rangle - \widetilde{\langle u_x, v_y \rangle}| \leqslant \frac{1}{10\gamma_2(f)} \leqslant \frac{1}{\|u_x\|_2 \|v_y\|_2} \leqslant \frac{1}{|\langle u_x, v_y \rangle|^2}$$

by Cauchy-Schwartz inequality. Hence, the error is the same for all inputs. □

Lastly, we draw the connection between $\gamma_2(f)$ and $Q^*_\epsilon(f)$.

Lemma 4. *Fix an $\epsilon > 0$. Let $f : \mathcal{X} \times \mathcal{Y} \to \{\epsilon, 1-\epsilon\}$. Then, $\gamma_2(f) \leqslant 2^{Q^*_{=\epsilon}(f)}$.*

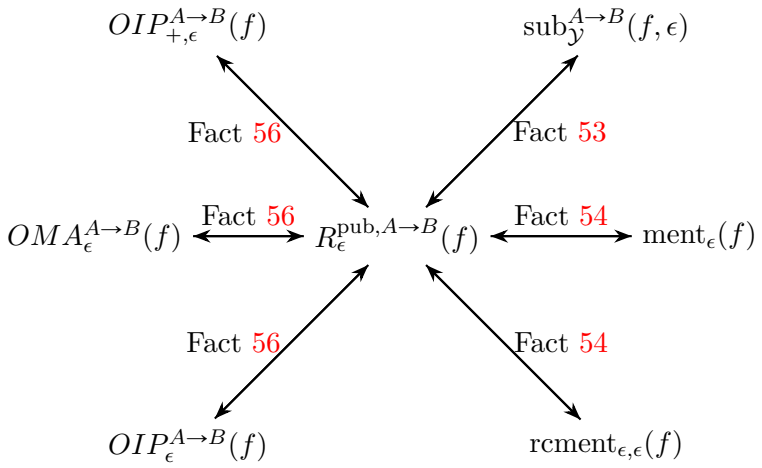
Proof. Consider a quantum protocol that solves f with shared entanglement, that accepts ϵ-inputs with probability ϵ and accepts $(1-\epsilon)$-inputs with probability $1-\epsilon$. Let M_f be the communication matrix of f and P be its corresponding matrix of acceptance probabilities. Clearly, $P = M_f$. Then,

$$\gamma_2(M_f) = \gamma_2(P) \leqslant 2^{Q^*_{=\epsilon}(f)},$$

where the last inequality is due to [40, Lemmas 12]. □

3.3 Class 3

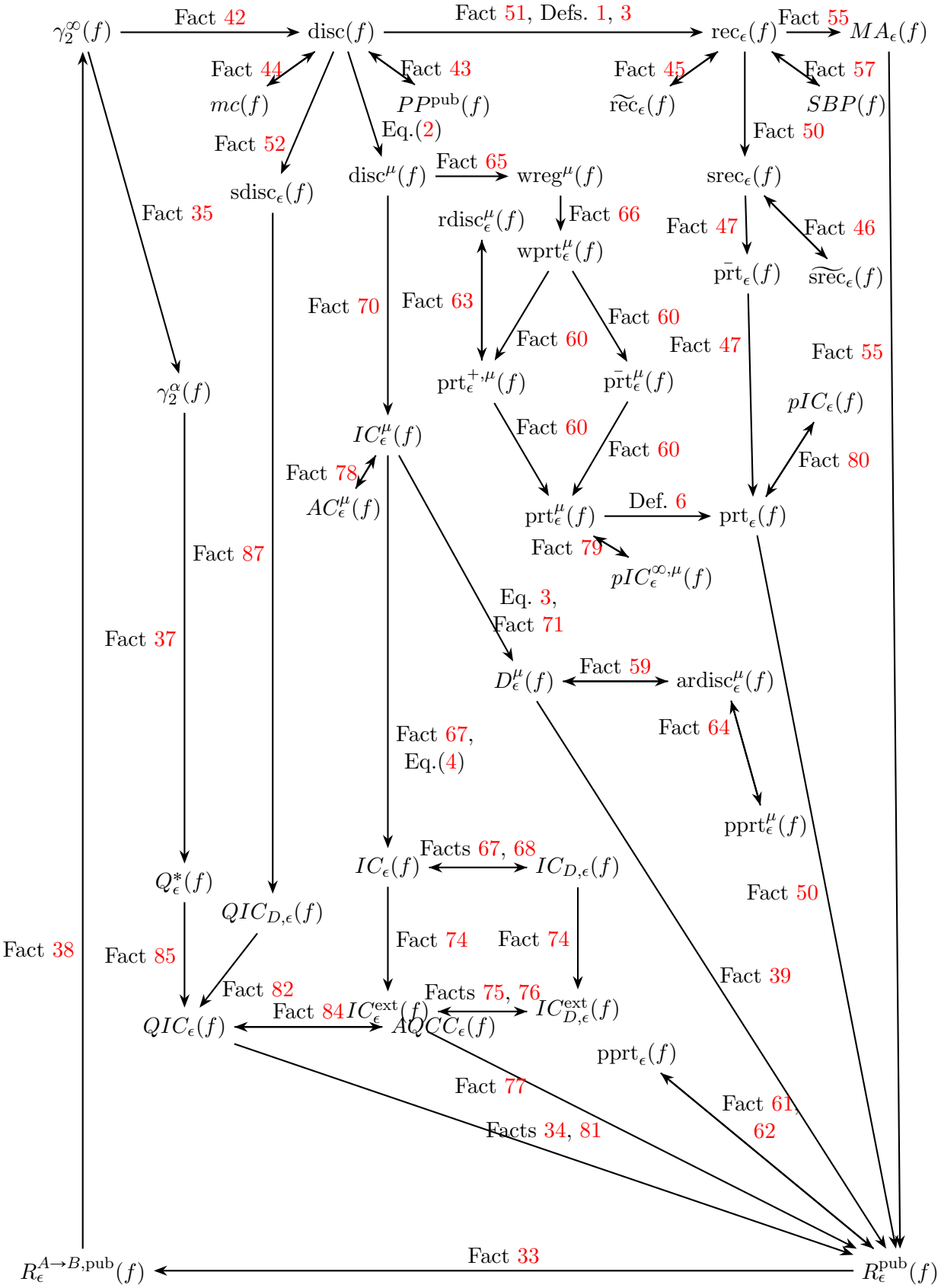
This class consists of variants of the approximate γ_2 norm, various lower bounds measures, distributional complexity, weakly unbounded error communication complexity, quantum communication complexity with prior entanglement, public-coin randomized communication complexity, Merlin-Arthur communication complexity, variants of internal and external information complexity, pseudotranscript complexity and amortized communication complexity. The constant equivalence relationship between complexity measures in Class 3 is summarized in Figures 4a and 5a.



(a) Constant equivalence relationship between complexity measures with $R^{A\to B,\text{pub}}_{\epsilon}(f)$.

Theorem 3. *Class 3 includes the following complexity measures:*

[2]This is due to the definition of $\gamma_2(f)$ and the fact that M_f has entries -1 and 1.



(a) Relationship between complexity measure in Class 3.

Figure 5

- $R_\epsilon^{A\to B,\mathrm{pub}}(f)$
- $\gamma_2^\infty(f)$
- $\gamma_2^\alpha(f)$
- $\mathrm{disc}^\mu(f)$
- $\mathrm{disc}(f)$
- $\mathrm{sdisc}_\epsilon(f)$
- $mc(f)$
- $PP^{\mathrm{pub}}(f)$
- $\mathrm{rec}_\epsilon(f)$
- $\widetilde{\mathrm{rec}}_\epsilon(f)$
- $\mathrm{srec}_\epsilon(f)$
- $\widetilde{\mathrm{srec}}_\epsilon(f)$
- $\bar{\mathrm{prt}}_\epsilon(f)$
- $\mathrm{prt}_\epsilon(f)$
- $\mathrm{sub}^{\mathrm{A}\to\mathrm{B}}{}_{\mathcal{Y},\epsilon}(f)$
- $D_\epsilon^\mu(f)$
- $MA_\epsilon(f)$
- $SBP(f)$
- $\mathrm{ment}_\epsilon(f)$
- $\mathrm{rcment}_{\epsilon,\epsilon}(f)$
- $Q_\epsilon^*(f)$
- $QIC_{D,\epsilon}(f)$
- $QIC_\epsilon(f)$
- $AQCC_\epsilon(f)$
- $R_\epsilon^{\mathrm{pub}}(f)$
- $IC_\epsilon^\mu(f)$
- $IC_\epsilon(f)$
- $IC_{D,\epsilon}(f)$
- $IC_\epsilon^{\mathrm{ext}}(f)$
- $IC_{D,\epsilon}^{\mathrm{ext}}(f)$
- $AC_\epsilon^\mu(f)$
- $\mathrm{wreg}^\mu(f)$
- $\mathrm{wprt}_\epsilon^\mu(f)$
- $\mathrm{prt}_\epsilon^{+,\mu}(f)$
- $\bar{\mathrm{prt}}_\epsilon^\mu(f)$
- $\mathrm{prt}_\epsilon^\mu(f)$
- $\mathrm{rdisc}_\epsilon^\mu(f)$
- $\mathrm{ardisc}_\epsilon^\mu(f)$
- $\mathrm{pprt}_\epsilon^\mu(f)$
- $\mathrm{pprt}_\epsilon(f)$
- $pIC_\epsilon^{\infty,\mu}(f)$
- $pIC_\epsilon^\infty(f)$
- $OMA_\epsilon^{A\to B}(f)$
- $OIP_\epsilon^{A\to B}(f)$
- $OIP_{+,\epsilon}^{A\to B}(f)$

Proof. First, note that $R_\epsilon^{A\to B,\mathrm{pub}}(f) \to \gamma_2^\infty(f)$ by Fact 38; $\gamma_2^\infty(f) \to \mathrm{disc}(f)$ by Fact 42; $\mathrm{disc}(f) \to \mathrm{rec}_\epsilon(f)$ by Fact 51, Definitions 1 and 3; $\mathrm{rec}_\epsilon(f) \to MA_\epsilon(f)$ by Fact 55; $MA_\epsilon(f) \to R_\epsilon^{\mathrm{pub}}(f)$ by Fact 55; $R_\epsilon^{\mathrm{pub}}(f) \to R_\epsilon^{A\to B,\mathrm{pub}}(f)$ by Fact 33.

From $\gamma_2^\infty(f)$, we have $\gamma_2^\infty(f) \to \gamma_2^\alpha(f)$ by Fact 35; $\gamma_2^\alpha(f) \to Q_\epsilon^*(f)$ by Fact 37; $Q_\epsilon^*(f) \to QIC_\epsilon(f)$ by Fact 85; $QIC_\epsilon(f) \to R^{\mathrm{pub}_\epsilon(f)}$ by Facts 34 and 81.

From $\mathrm{disc}(f)$, we have $\mathrm{disc}(f) \to \mathrm{sdisc}_\epsilon(f)$ by Fact 52; $\mathrm{sdisc}_\epsilon(f) \to QIC_{D,\epsilon}(f)$ by Fact 87; $QIC_{D,\epsilon}(f) \to QIC_\epsilon(f)$ by Facts 82 and 83. Furthermore, $\mathrm{disc}(f) \to \mathrm{disc}^\mu(f)$ by Equation 2; $\mathrm{disc}^\mu(f) \to IC_\epsilon^\mu(f)$ by Fact 70; $IC_\epsilon^\mu(f) \to IC_\epsilon(f)$ by Fact 67 and Equation 4; $IC_\epsilon(f) \to IC_\epsilon^{\mathrm{ext}}(f)$ by Fact 74; $IC_\epsilon^{\mathrm{ext}}(f) \to R_\epsilon^{\mathrm{pub}}(f)$ by Fact 77. At the same time, $IC_\epsilon(f) \leftrightarrow IC_{D,\epsilon}(f)$ by Facts 67, 68; $IC_{D,\epsilon}(f) \to IC_{D,\epsilon}^{\mathrm{ext}}(f)$ by Fact 74; $IC_\epsilon^{\mathrm{ext}}(f) \to IC_{D,\epsilon}^{\mathrm{ext}}(f)$ by Facts 75, 76. From $IC_\epsilon^\mu(f)$, we have $IC_\epsilon^\mu(f) \to D_\epsilon^\mu(f)$ by Equation 3 and Fact 71; $D_\epsilon^\mu(f) \to R_\epsilon^{\mathrm{pub}}(f)$ by Fact 39.

From $\mathrm{disc}^\mu(f)$, we have $\mathrm{disc}^\mu(f) \to \mathrm{wreg}^\mu(f)$ by Fact 65; $\mathrm{wreg}^\mu(f) \to \mathrm{wprt}^\mu(f)$ by Fact 66; $\mathrm{wprt}^\mu(f) \to \mathrm{prt}_\epsilon^{+,\mu}(f), \bar{\mathrm{prt}}_\epsilon^\mu(f)$ by Fact 60; $\mathrm{prt}_\epsilon^{+,\mu}(f), \bar{\mathrm{prt}}_\epsilon^\mu(f) \to \mathrm{prt}_\epsilon^\mu(f)$ by Fact 60.

Moreover, $\mathrm{rec}_\epsilon(f) \to \mathrm{srec}_\epsilon(f)$ by Fact 50; $\mathrm{srec}_\epsilon(f) \to \bar{\mathrm{prt}}_\epsilon(f)$ by Fact 47; $\bar{\mathrm{prt}}_\epsilon(f) \to \mathrm{prt}_\epsilon(f)$ by Fact 47; $\mathrm{prt}_\epsilon(f) \to R_\epsilon^{\mathrm{pub}}(f)$ by Fact 50.

Lastly, $\mathrm{prt}_\epsilon^\mu(f) \to \mathrm{prt}_\epsilon(f)$ by Definition 6. For $\leftrightarrow$ relations, we have $QIC_\epsilon(f) \leftrightarrow AQCC_\epsilon(f)$ by Fact 84; $\mathrm{mc}(f) \leftrightarrow \mathrm{disc}(f)$ by Fact 44; $PP^{\mathrm{pub}}(f) \leftrightarrow \mathrm{disc}(f)$ by Fact 43; $AC_\epsilon^\mu(f) \leftrightarrow IC_\epsilon^\mu(f)$ by Fact 78; $\mathrm{prt}_\epsilon^{+,\mu}(f) \leftrightarrow \mathrm{rdisc}_\epsilon^\mu(f)$ by Fact 63; $\mathrm{prt}_\epsilon^\mu(f) \leftrightarrow pIC_\epsilon^{\infty,\mu}(f)$ by Fact 79; $\mathrm{prt}_\epsilon(f) \leftrightarrow pIC_\epsilon^\infty(f)$ by Fact 80; $\mathrm{rec}_\epsilon(f) \leftrightarrow \widetilde{\mathrm{rec}}_\epsilon(f)$ by Fact 45; $\mathrm{rec}_\epsilon(f) \leftrightarrow SBP(f)$ by Fact 57; $\mathrm{srec}_\epsilon(f) \leftrightarrow \widetilde{\mathrm{srec}}_\epsilon(f)$ by Fact 46; $R_\epsilon^{\mathrm{pub}}(f) \leftrightarrow \mathrm{pprt}_\epsilon(f)$ by Facts 61 and 62.

For $\leftrightarrow$ relations with $R_\epsilon^{A\to B,\mathrm{pub}}(f)$, we have $R_\epsilon^{A\to B,\mathrm{pub}}(f) \leftrightarrow \mathrm{ment}_\epsilon(f)$ by Fact 54; $R_\epsilon^{A\to B,\mathrm{pub}}(f) \leftrightarrow \mathrm{rcment}_{\epsilon,\epsilon}(f)$ by Fact 54; $R_\epsilon^{A\to B,\mathrm{pub}}(f) \leftrightarrow OIP_\epsilon^{A\to B}(f)$ by Fact 56; $R_\epsilon^{A\to B,\mathrm{pub}}(f) \leftrightarrow OMA_\epsilon^{A\to B}(f)$ by Fact 56; $R_\epsilon^{A\to B,\mathrm{pub}}(f) \leftrightarrow OIP_{+,\epsilon}^{A\to B}(f)$ by Fact 56; $R_\epsilon^{A\to B,\mathrm{pub}}(f) \leftrightarrow \mathrm{sub}_{\mathcal{Y},\epsilon}^{A\to B}(f)$ by Fact 53. □

Below, we state the results used in the proof of Theorem 3. We first state the fact that one-way public-coin randomized communication complexity is at least exponentially larger than its two-way counterpart and quantum communication complexity with prior entanglement is smaller than public-coin randomized communication complexity.

Fact 33 ([37, Exercise 4.2.1]). *For $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, $R_\epsilon^{\mathrm{pub}}(f) \geqslant \log R_\epsilon^{A\to B,\mathrm{pub}}(f)$.*

Fact 34. *For $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, $Q_\epsilon^*(f) \leqslant R_\epsilon^{\mathrm{pub}}(f)$.*

Proof. In order to simulate a public-coin randomized protocol using a quantum protocol with shared entanglement, Alice and Bob share the state $\sum_i \sqrt{p(i)}\,|i\rangle$, where $p(i) \in [0,1]$ and $\sum_i \sqrt{p(i)} = 1$. Each of them then measures their sides of the entangled state to get the randomness. The quantum protocol then proceeds as per the randomized protocol. □

The next three results from Linial and Shraibman [40] show that $\gamma_2^\infty(f) \leqslant \gamma_2^\alpha(f)$ and both $R_\epsilon^{\text{pub}}(M_f)$ and $Q_\epsilon^*(M_f)$ can be lower bounded by $\gamma_2^\alpha(f)$ when $\alpha = \frac{1}{1-2\epsilon}$. Furthermore, Lee and Shraibman showed that $R_\epsilon^{A\to B,\text{pub}}(f)$ is upper bounded by $\gamma_2^\infty(f)$ [77].

Fact 35 ([40, Proposition 3])**.** *For every $m \times n$ sign matrix M_f and for every $\alpha \geqslant 1$, $\gamma_2^\infty(M_f) \leqslant \gamma_2^\alpha(M_f) \leqslant \gamma_2(M_f) \leqslant \sqrt{\text{rank}(M_f)}$ for all $1 \leqslant \alpha < \infty$.*

Fact 36 ([40, Theorem 1])**.** *For every sign matrix M_f and every $\epsilon > 0$, $R_\epsilon^{\text{pub}}(M_f) \geqslant 2\log \gamma_2^{\frac{1}{1-2\epsilon}}(M_f) - 2\log\left(\frac{1}{1-2\epsilon}\right)$.*

Fact 37 ([40, Theorem 1])**.** *For every sign matrix M_f and every $\epsilon > 0$,, $Q_\epsilon^*(M_f) \geqslant \log \gamma_2^{\frac{1}{1-2\epsilon}}(M_f) - \log\left(\frac{1}{1-2\epsilon}\right) - 2$.*

Fact 38 ([40, Claim 2])**.** *For any function f, we have $R_\epsilon^{A\to B,\text{pub}}(f) \leqslant O((\gamma_2^\infty(f))^2)$.*

Consider a probability distribution μ over the set of inputs $\mathcal{X} \times \mathcal{Y}$. The (μ,ϵ)-*distributional complexity* of $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, denoted as $D_\epsilon^\mu(f)$, is the cost of the best deterministic protocol that errs on at most a ϵ fraction of the inputs, weighted according to the distribution μ [37]. Distributional complexity is a useful complexity measure when obtaining lower bounds on randomized communication complexity, in both the one-way and two-way models.

Fact 39 ([4], [37, Theorem 3.20])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$,. Then, $R_\epsilon^{\text{pub}}(f) = \max_\mu D_\epsilon^\mu(f)$ over distributions μ.*

Introduced by Chor and Goldreich [78], discrepancy has become one of the most commonly used measures in communication complexity to prove lower bounds for randomized protocols. Moreover, the inverse of this measure is rightly characterized by $\gamma_2^\infty(f)$ up to constant factors. For a Boolean matrix, M_f, the discrepancy of M_f is the minimum over all input distributions of the maximum correlation that M_f has with a rectangle.

Definition 1 ([37, Definition 3.27])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,l\}$ be a function, R be any rectangle, and μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. Denote*

$$\text{disc}_\mu(R,f) = \left|\Pr_\mu[f(x,y) = 0 \text{ and } (x,y) \in R] - \Pr_\mu[f(x,y) = 1 \text{ and } (x,y) \in R]\right|.$$

The discrepancy of f with respect to μ is given by

$$\text{disc}_\mu(f) = \max_R \text{disc}_\mu(R,f),$$

where the maximum is taken over all rectangles. Finally, the discrepancy of f is

$$\text{disc}(f) = \min_\mu \text{disc}_\mu(f). \tag{2}$$

Fact 40 ([79, Theorem 7])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a two-party function and let $\epsilon \in [0,1/2)$. Then, $Q_\epsilon^*(f) = \Omega\left(\log \frac{1-2\epsilon}{\text{disc}(f)}\right)$.*

Fact 41 ([37, Proposition 3.28])**.** *For every function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, every probability distribution μ on $\mathcal{X} \times \mathcal{Y}$ and every $\epsilon \geqslant 0$, $D_{1/2-\epsilon}^\mu(f) \geqslant \log\left(\frac{2\epsilon}{\text{disc}_\mu(f)}\right)$.*

Fact 42 ([40, Theorem 18]). *For every sign matrix M_f, $\frac{1}{8}\gamma_2^\infty(M_f) \leqslant \frac{1}{\mathrm{disc}(M_f)} \leqslant 8\gamma_2^\infty(M_f)$,*

Define the weakly unbounded error (public-coin) communication complexity as follows [80, Definition 2.3.5]: $PP^{\mathrm{pub}}(f) = \inf_{0\leqslant\epsilon<1/2}\left\{R_\epsilon^{\mathrm{pub}}(f) - \log\left(\frac{1}{1/2-\epsilon}\right)\right\}$. The result below shows the equivalence relation between $PP^{\mathrm{pub}}(f)$ and $\mathrm{disc}(f)$.

Fact 43 ([80, Theorem 2.3.8], [81, Corollary 1.4]). *For all $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, we have $PP^{\mathrm{pub}}(f) = \Theta\left(\log \frac{1}{\mathrm{disc}(f)}\right)$.*

Given a sign matrix $M \in \{-1,1\}^{n\times n}$, a (Euclidean) embedding of M_f is a collection of vectors $u_1, \cdots, u_n, v_1, \cdots, v_n \in \mathbb{R}^k$ such that $\langle u_i, \cdot v_j\rangle \cdot M_{i,j} > 0$ for all $i, j \in [n]$. The margin of the embedding is defined as

$$\zeta = \min_{i,j} \frac{|\langle u_i, v_j\rangle|}{\|u_i\|_2 \cdot \|v_j\|_2}.$$

The margin complexity of M, denoted as $\mathrm{mc}(M)$ is the minimum $1/\zeta$ over all embeddings of M [82]. The fact below shows a tight characterization of $\frac{1}{\mathrm{disc(f)}}$ in terms of $\mathrm{mc}(f)$ up to some constant factor.

Fact 44 ([82, Theorem 2.7], [27, Theorem 3.1]). *Let M_f be a sign matrix. Then, $\frac{1}{8}\,\mathrm{mc}(M_f) \leqslant \frac{1}{\mathrm{disc}(M_f)} \leqslant 8\,\mathrm{nc}(M_f)$.*

We now review a few other lower bound methods in communication complexity. In the beginning of this section, we briefly introduced the concept of rectangles. We shall proceed with a more detailed description here.

Definition 2 (Rectangle bound, [83, Definition 5]). *Let $f : \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}$. The ϵ-rectangle bound of f, denoted as $\widetilde{\mathrm{rec}}_\epsilon(f)$ is defined as $\widetilde{\mathrm{rec}}_\epsilon(f) := \max\{\widetilde{\mathrm{rec}}_\epsilon^z(f) : z \in \mathcal{Z}\}$, where $\widetilde{\mathrm{rec}}_\epsilon^z(f)$ is the optimal value of the following linear program:*

Primal

$$\begin{aligned} \mathit{Min}\quad & \sum_R w_R \\ \mathit{st}\quad & \forall (x,y) \in f^{-1}(z) : \sum_{R:(x,y)\in R} w_R \geqslant 1-\epsilon, \\ & \forall (x,y) \in f^{-1} - f^{-1}(z) : \sum_{R:(x,y)\in R} w_R \leqslant \epsilon, \\ & \forall R : w_R \geqslant 0. \end{aligned}$$

Dual

$$\begin{aligned} \mathit{Max}\quad & \sum_{(x,y)\in f^{-1}(z)} (1-\epsilon)\cdot\mu_{x,y} - \sum_{(x,y)\in f^{-1}-f^{-1}(z)} \epsilon\cdot\mu_{x,y} \\ \mathit{st}\quad & \forall R : \sum_{(x,y)\in f^{-1}(z)\cap R} \mu_{x,y} - \sum_{(x,y)\in (R\cap f^{-1})-f^{-1}(z)} \mu_{x,y} \leqslant 1, \\ & \forall (x,y) : \mu_{x,y} \geqslant 0. \end{aligned}$$

Definition 3 (Rectangle bound: conventional definition [72, Definition 3], [83, Definition 6]). *Let $f : \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}$ be a function and let $\mathcal{R}$ be the set of all rectangles in $\mathcal{X}\times\mathcal{Y}$. The ϵ-rectangle bound of f i, denoted as $\mathrm{rec}_\epsilon(f)$, is defined as:*

$$\begin{aligned} &\mathrm{rec}_\epsilon(f) := \max\{\mathrm{rec}_\epsilon^z(f) : z \in \mathcal{Z}\} \\ &\mathrm{rec}_\epsilon^z(f) := \max\{\mathrm{rec}_\epsilon^{z,\lambda}(f) : \lambda \textit{ is a distribution on } \mathcal{X}\times\mathcal{Y} \cap f^{-1} \textit{ with } \lambda(f^{-1}(z)) \geqslant 0.5\} \\ &\mathrm{rec}_\epsilon^{z,\lambda}(f) := \min\left\{\frac{1}{\lambda(f^{-1}(z)\cap R)} : R \in \mathcal{R} \textit{ with } \epsilon\cdot\lambda(f^{-1}(z)\cap R) > \lambda(R - f^{-1}(z))\right\}, \end{aligned}$$

The two definitions of the rectangle bound has been shown to be equivalent.

Fact 45 ([83, Lemma 1]). *Let $f : \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}$ be a function and let $\epsilon > 0$. Then,*

1. $\widetilde{\mathrm{rec}}_\epsilon^z(f) \leqslant \mathrm{rec}_{\epsilon/2}^z(f)$.
2. $\widetilde{\mathrm{rec}}_\epsilon^z(f) \geqslant \frac{1}{2}\cdot\left(\frac{1}{2}-\epsilon\right)\cdot \mathrm{rect}_{2\epsilon}^z(f)$.

In the one-way communication model, a *one-way rectangle* $R^{A\to B}$ is a set $S \times \mathcal{Y}$, where $S \subseteq \mathcal{X}$. For a distribution μ over $\mathcal{X} \times \mathcal{Y}$, let $\mu_{R^{A\to B}}$ be the distribution that arises from μ conditioned on the event $R^{A\to B}$ and let $\mu(R^{A\to B})$ represent the probability, under μ, of event $R^{A\to B}$ [84, Definition 4]. The *one-way rectangle bound*, denoted as $\mathrm{rec}_\epsilon^{A\to B}(f)$, is defined as [84, Definition 6]

$$\mathrm{rec}_\epsilon^{A\to B}(f) := \max_{\mu} \min_{R^{A\to B}} \log \frac{1}{\mu(R^{A\to B})}.$$

Definition 4 (Smooth-rectangle bound [83, Definition 2])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. The ϵ-smooth rectangle bound of f, denoted as $\widetilde{\mathrm{srec}}_\epsilon(f)$, is defined to be $\max\{\widetilde{\mathrm{srec}}_\epsilon^z(f) : z \in \mathcal{Z}\}$, where $\widetilde{\mathrm{srec}}_\epsilon^z(f)$ is given by the optimal value of the linear program below:*

Primal

$$\begin{aligned} \textit{Min} \quad & \sum_{R\in\mathcal{R}} w_R \\ \textit{st} \quad & \forall (x,y) \in f^{-1}(z) : \sum_{R:(x,y)\in R} w_R \geqslant 1-\epsilon, \\ & \forall (x,y) \in f^{-1}(z) : \sum_{R:(x,y)\in R} w_R \leqslant 1, \\ & \forall (x,y) \in f^{-1} - f^{-1}(z) : \sum_{R:(x,y)\in R} w_R \leqslant \epsilon, \\ & \forall R : w_R \geqslant 0. \end{aligned}$$

Dual

$$\begin{aligned} \textit{Max} \quad & \sum_{(x,y)\in f^{-1}(z)} ((1-\epsilon)\mu_{x,y} - \phi_{x,y}) - \sum_{(x,y)\in f^{-1}-f^{-1}(z)} \epsilon \cdot \mu_{x,y} \\ \textit{st} \quad & \forall R : \sum_{(x,y)\in f^{-1}(z)\cap R} (\mu_{x,y} - \phi_{x,y}) - \sum_{(x,y)\in (R\cap f^{-1})-f^{-1}(z)} \mu_{x,y} \leqslant 1, \\ & \forall (x,y) : \mu_{x,y} \geqslant 0, \phi_{x,y} \geqslant 0. \end{aligned}$$

Definition 5 (Smooth-rectangle bound: natural definition [83, Definition 3])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. The (ϵ,δ)-smooth-rectangle bound of f, denoted by $\mathrm{srec}_{\epsilon,\delta}(f)$, is defined as (refer to the definition of $\mathrm{rec}_\epsilon^{z,\lambda}$ in Definition 3)*

$$\begin{aligned} &\mathrm{srec}_{\epsilon,\delta}(f) := \max\{\mathrm{srec}_{\epsilon,\delta}^z(f) : z \in \mathcal{Z}\} \\ &\mathrm{srec}_{\epsilon,\delta}^z(f) := \max\{\mathrm{srec}_{\epsilon,\delta}^{z,\lambda}(f) : \lambda \textit{ is a probability distribution on } (\mathcal{X}\times\mathcal{Y}) \cap f^{-1}\} \\ &\mathrm{srec}_{\epsilon,\lambda}^{z,\lambda}(f) := \max\left\{\mathrm{rec}_\epsilon^{z,\lambda}(g) : g : \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}; \Pr_{(x,y)\leftarrow\lambda}[f(x,y)\neq g(x,y)] \leqslant \delta; \lambda(g^{-1}(z)) \geqslant 0.5\right\}. \end{aligned}$$

Definitions 4 and 5 are also equivalent.

Fact 46 ([83, Lemma 2])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function and let $\epsilon > 0$. Then, for all $z \in \mathcal{Z}$,*

1. $\widetilde{\mathrm{srec}}_\epsilon^z(f) \leqslant \mathrm{srec}_{\epsilon/2,(1-\epsilon)/2}^z(f)$.
2. $\widetilde{\mathrm{srec}}_\epsilon^z(f) \geqslant \frac{1}{2}\cdot\left(\frac{1}{4}-\epsilon\right)\cdot \mathrm{srec}_{2\epsilon,\epsilon/2}^z(f)$.

Definition 6 (Partition bound [83, Definition 1])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. The ϵ-partition bound of f, denoted $\mathrm{prt}_\epsilon(f)$, is given by the optimal value of the following linear program:*

Primal

$$\begin{aligned} \textit{Min} \quad & \sum_z \sum_R w_{z,R} \\ \textit{st} \quad & \forall (x,y) \in f^{-1} : \sum_{R:(x,y)\in R} w_{f(x,y),R} \geqslant 1-\epsilon, \\ & \forall (x,y) : \sum_{R:(x,y)\in R} \sum_z w_{z,R} = 1, \\ & \forall z, \forall R : w_{z,R} \geqslant 0. \end{aligned}$$

Dual

$$\begin{aligned} \textit{Max} \quad & \sum_{(x,y)\in f^{-1}} (1-\epsilon)\mu_{x,y} + \sum_{(x,y)} \phi_{x,y}, \\ \textit{st} \quad & \forall z, \forall R : \sum_{(x,y)\in f^{-1}(z)\cap R} \mu_{x,y} + \sum_{(x,y)\in R} \phi_{x,y} \leqslant 1, \\ & \forall (x,y) : \mu_{x,y} \geqslant 0, \phi_{x,y} \in \mathbb{R}. \end{aligned}$$

Definition 7 (Relaxed partition bound [85, Definition 3.2])**.** *Let $f : \mathcal{I} \to \mathcal{Z}$, where $\mathcal{I} \subseteq \mathcal{X}\times\mathcal{Y}$ The distributional relaxed partition bound, denoted as $\bar{\mathrm{prt}}_\epsilon^\mu(f)$, is the value of the following linear program. (The value of z ranges over $\mathcal{Z}$ and R over all rectangles, including the empty rectangle.)*

$$\bar{\mathrm{prt}}_\epsilon^\mu(f) = \min_{\eta,p,R,z\geqslant 0} \quad \frac{1}{\eta}$$

$$
\begin{aligned}
st \quad & \sum_{(x,y)\in\mathcal{I}} \mu_{x,y} \sum_{R:(x,y)\in R} p_{R,f(x,y)} + \sum_{(x,y)\notin\mathcal{I}} \mu_{x,y} \sum_{z,R:(x,y)\in R} p_{R,z} \geqslant (1-\epsilon)\eta, \\
& \forall (x,y) \in \mathcal{X}\times\mathcal{Y}, \sum_{z,R:(x,y)\in R} p_{R,z} \leqslant \eta, \\
& \sum_{R,z} p_{R,z} = 1.
\end{aligned}
$$

The relaxed partition bound is given by $\bar{\text{prt}}_\epsilon(f) = \max_\mu \bar{\text{prt}}^\mu_\epsilon(f)$.

The relaxed partition bound subsumes the norm-based methods (for example, the γ_2 method) and the rectangle-based methods (for example, the rectangle/corruption bound, the smooth rectangle bound and the discrepancy bound).

Fact 47 ([85, Lemma 3.3]). *For all functions $f : \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$, ϵ and $z\in\mathcal{Z}$, we have* $\text{srec}^z_\epsilon(f) \leqslant \bar{\text{prt}}_\epsilon(f) \leqslant \text{prt}_\epsilon(f)$.

Furthermore, the same authors showed that the distributional complexity is an upper bound on the logarithm of the relaxed partition bound.

Fact 48 ([85, Corollary 1.4]). *For every ϵ, μ and every $f : \mathcal{X}\times\mathcal{Y}\to\{0,1\}$, $D^\mu_\epsilon(f) \geqslant \log \bar{\text{prt}}^\mu_\epsilon(f)$.*

Definition 8 (Smooth discrepancy [83, Definition 4]). *Let $f : \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ be a Boolean function. The smooth discrepancy of f, denoted by $\widetilde{\text{sdisc}}_\epsilon(f)$, is given by the optimal value of the linear program below*

Primal

$$
\begin{aligned}
Min \quad & \sum_{R\in\mathcal{R}} w_R + v_R \\
st \quad & \forall (x,y)\in f^{-1}(1) : 1+\epsilon \geqslant \sum_{R:(x,y)\in R} w_R - v_R \geqslant 1, \\
& \forall (x,y)\in f^{-1}(0) : 1+\epsilon \geqslant \sum_{R:(x,y)\in R} v_R - w_R \geqslant 1, \\
& \forall R : w_F, v_R \geqslant 0.
\end{aligned}
$$

Dual

$$
\begin{aligned}
Max \quad & \sum_{(x,y)\in f^{-1}} \mu_{x,y} - (1+\epsilon)\phi_{x,y}, \\
st \quad & \forall R : \sum_{(x,y)\in f^{-1}(1)\cap R} (\mu_{x,y} - \phi_{x,y}) - \sum_{(x,y)\in R\cap f^{-1}(0)} (\mu x, y - \phi_{x,y}) \leqslant 1, \\
& \forall R : \sum_{(x,y)\in f^{-1}(0)\cap R} (\mu_{x,y} - \phi_{x,y}) - \sum_{(x,y)\in R\cap f^{-1}(1)} (\mu_{x,y} - \phi_{x,y}) \leqslant 1, \\
& \forall (x,y) : \mu_{x,y} \geqslant 0; \phi_{x,y} \geqslant 0.
\end{aligned}
$$

Definition 9 (Smooth discrepancy: natural definition [83, Definition 9]). *Let $f : \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ be a Boolean function. The δ-smooth discrepancy of f, denoted as* $\text{sdisc}_\delta(f)$, *is defined as follows*

$$
\begin{aligned}
&\text{sdisc}_\delta(f) := \max\{\text{sdisc}^\lambda_\delta(f) : \lambda \text{ is a distribution on } \mathcal{X}\times\mathcal{Y}\cap f^{-1}\}, \\
&\text{sdisc}^\lambda_\delta(f) :=: \max\left\{\text{disc}^\lambda(g) : g : \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}; \Pr_{(x,y)\leftarrow\lambda}[f(x,y)\neq g(x,y)] < \delta\right\}.
\end{aligned}
$$

Definitions 8 and 9 are equivalent.

Fact 49 ([83, Lemma 3]). *Let $f : \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ be a function and let $\epsilon > 0$. Then,*

1. $\text{sdisc}_{1/2-\epsilon/8}(f) \geqslant \widetilde{\text{sdisc}}_\epsilon(f)$,
2. $\frac{1}{2}\cdot \text{sdisc}_{\frac{1}{4+2\epsilon}}(f) \leqslant \widetilde{\text{sdisc}}_\epsilon(f)$.

Definition 10 (One-way subdistribution bounds, [86, Dfinitino 3.2]). *For a distribution μ over $\mathcal{X}\times\mathcal{Y}$, define* $\text{sub}^{A\to B,\mu}_{\mathcal{Y},\epsilon}(f) := \min_\lambda S_\infty(\lambda\|\mu)$[3], *where λ is taken over all distributions that are one-message-like*[4] *for μ (with respect to $\mathcal{X}$) and one-way ϵ-monochromatic for f. The one-way subdistribution bound is defined as such* $\text{sub}^{A\to B}_{\mathcal{Y},\epsilon}(f) = \max_\mu \text{sub}^{A\to B,\mu}_{\mathcal{Y},\epsilon}(f)$, *where μ is taken over all distributions on $\mathcal{X}\times\mathcal{Y}$. When the maximization is restricted to product distributions μ, we refer to the quantity as the one-way product subdistribution bound* $\text{sub}^{A\to B,[]}_{\mathcal{Y},\epsilon}(f)$.

[3] For distributions $\mathcal{D}_1, \mathcal{D}_2$, with respect to set $\mathcal{X}$, the relative co-min-entropy of $\mathcal{D}_1$ with respect to $\mathcal{D}_2$ is defined as $S_\infty(\mathcal{D}_1\|\mathcal{D}_2) = \inf\{c : \mathcal{D}_2 \geqslant \mathcal{D}_1/2^c\}$. When $\mathcal{X}$ is finite, we have $S_\infty(\mathcal{D}_1\|\mathcal{D}_2) = \max_{x\in\mathcal{X}} \log \frac{\mathcal{D}_1(x)}{\mathcal{D}_2(x)}$ [86].

[4] We say that λ is one-message-like for μ with respect to $\mathcal{X}$ if for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$, whenever $\lambda_\mathcal{X}(x) > 0$, we have $\mu_\mathcal{X}(x) > 0$ and $\lambda_\mathcal{Y}(y|x) = \mu_\mathcal{Y}(y|x)$. Here, $P_\mathcal{X}(x) = \sum_{y\in\mathcal{Y}} P(x,y)$ and $P_\mathcal{Y}(y|x) = \frac{P(x,y)}{P_\mathcal{X}(x)}$ for some distribution P.

In the definition above, we say that a distribution λ is ϵ-monochromatic for f if there exists a function $g : \mathcal{Y} \to \mathcal{Z}$ such that $\Pr_{XY\sim\lambda}[(X, Y, g(Y)) \in f] \geqslant 1 - \epsilon$ [86, Definition 3.1].

The subsequent results show how different lower bound measures relate to each other. Since we only consider total Boolean functions in this work, results that hold for partial functions are useful for us in drawing the equivalence relation between complexity measures. Since the set of total functions is a subset of partial functions, lower bound results for partial functions also hold true for total functions.

Fact 50 ([83, Theorem 1]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a (partial) function. Then,*

1. $R_\epsilon^{\text{pub}}(f) \geqslant \log \text{prt}_\epsilon(f)$.
2. $\text{prt}_\epsilon(f) \geqslant \text{srec}_\epsilon(f)$.
3. $\text{srec}_\epsilon(f) \geqslant \text{rec}_\epsilon(f)$.

Fact 51 ([83, Lemma 4]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a function; let $z \in \{0, 1\}$ and let λ be a distribution on $\mathcal{X} \times \mathcal{Y} \cap f^{-1}$. Let $\epsilon, \delta > 0$, then $\text{rec}_\epsilon^z(f) \geqslant \left(\frac{1}{2} - \epsilon\right) \text{disc}^\lambda(f) - \frac{1}{2}$.*

Fact 52 ([87, Equation 29]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$. Then, $\left(\frac{1}{\text{disc}(f)}\right)^{O(1)} \leqslant 2^{O(\text{sdisc}_{1/5}(f))}$.*

We shall see how these lower bound measures associates with different variants of public-coin randomized communication complexity. In the one-way communication model, Jain, Klauck and Nayak [86] showed that for any relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, the subdistribution bound characterizes the public-coin randomized one-way communication complexity.

Fact 53 ([86, Theorem 4.4]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $0 \leqslant \epsilon \leqslant 1/6$. There exist universal constants c_1, c_2 such that $\text{sub}_{\mathcal{Y},\epsilon}^{A\to B}(f) - 1 \leqslant c_1 \cdot R_\epsilon^{A\to B,\text{pub}}(f) \leqslant c_2 \left(\text{sub}_{\mathcal{Y},\epsilon}^{A\to B}(f) + \log \frac{1}{\epsilon} + 2\right)$.*

The tight characterization of $R_\epsilon^{A\to B,\text{pub}}(f)$ in terms of the one-way subdistribution bound by Jain, Klauck and Nayak [86] was subsequently strengthened by Jain [88] via the robust conditional relative min-entropy and relative min-entropy bounds.

Definition 11 (Robust conditional relative min-entropy bound [88, Definition 3.4]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. The ϵ-error δ-robust conditional relative min-entropy bound of f with respect to distribution μ, denoted as $\text{rcment}_{\epsilon,\delta}^\mu(f)$, is defined as*

$$\text{rcment}_{\epsilon,\delta}^\mu(f) := \min\{\text{rcment}_\delta^\mu(\lambda) | \lambda \text{ is one-way for } \mu \text{ and } \text{err}_f(\lambda) \leqslant \epsilon\},$$

where $\text{err}_f(\lambda) := \min\left\{\Pr_{(x,y)\leftarrow\lambda}[(x, y, h(y)) \notin f | h : \mathcal{Y} \to \mathcal{Z}]\right\}$. The ϵ-error δ-robust conditional relative min-entropy bound of f, denoted as $\text{rcment}_{\epsilon,\delta}(f)$, is

$$\text{rcment}_{\epsilon,\delta}(f) := \max\{\text{rcment}_{\epsilon,\delta}^\mu(f) | \mu \text{ is a distribution over } \mathcal{X} \times \mathcal{Y}\}.$$

Definition 12 (Relative min-entropy bound [88, Definition 3.5]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. The ϵ-error relative min-entropy bound of f with respect to distribution μ is*

$$\text{ment}_\epsilon^\mu(f) := \min\{S_\infty(\lambda\|\mu) | \lambda \text{ is one-way for } \mu \text{ and } \text{err}_f(\lambda) \leqslant \epsilon\},$$

where $S_\infty(\lambda\|\mu) = \max_{x\in\mathcal{X}} \log \frac{\lambda(x)}{\mu(x)}$[5]. The ϵ-error relative min-entropy bound of f, denoted as $\text{ment}_\epsilon(f)$, is defined as

$$\text{ment}_\epsilon(f) := \max\{\text{ment}_\epsilon^\mu(f) | \mu \text{ is a distribution over } \mathcal{X} \times \mathcal{Y}\}.$$

[5]We use $\lambda(x)$ and $\mu(x)$ to denote de probability of x under λ and μ respectively.

Fact 54 ([88, Theorem 4.3]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $\epsilon > 0$. Then,* $\mathrm{ment}_{2\epsilon}(f) - 1 \leqslant R_{\epsilon}^{A\to B,\mathrm{pub}}(f) \leqslant \mathrm{rcment}_{\epsilon/5,\epsilon/5}(f) + O\left(\log \frac{1}{\epsilon}\right)$. *Therefore,* $R_{\epsilon}^{A\to B,\mathrm{pub}}(f) = \Theta(\mathrm{ment}_{\epsilon}(f)) = \Theta(\mathrm{rcment}_{\epsilon,\epsilon}(f))$.

We formally define the Merlin-Arthur complexity below.

Definition 13 (Merlin-Arthur complexity [72, Definition 8], [73, Definition 1]). *In a Merlin-Arthur (MA) protocol for a Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, Alice receives a message (also known as the proof) from Merlin. Then, Alice and Bob communicate using public key randomness until they compute an output (the proof cannot depend on the randomness). The cost of an MA protocol is the sum of the length a of the proof, and the length c of the overall communication between Alice and Bob.*

We say that the protocol computes f if for all inputs x, y with $f(x,y) = 1$, there exists a proof such that x, y is accepted with probability $1 - \epsilon$ and for all inputs x, y with $f(x,y) = 0$ and all proofs, the probability that x, y is accepted is at most ϵ, for some $\epsilon < 1/2$. The Merlin-Arthur complexity of f, denoted $MA_{\epsilon}(f)$, is the smallest cost of an MA protocol for f.

Klauck [72] drew the relation between public coin interactive proofs, the rectangle bound and randomized communication complexity.

Fact 55 ([72, Corollary 2]). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Then,*

1. $R_{1/4}^{\mathrm{pub}}(f) \geqslant \Omega(\mathrm{rec}_{1/4}(f))$,
2. $R_{1/4}^{\mathrm{pub}}(f) \geqslant \max\{MA_{1/4}(f), MA_{1/4}(\neg f)\} \geqslant \Omega(\sqrt{\mathrm{rec}_{1/4}(f)})$,
3. $\mathrm{rec}_{1/4}(f) \geqslant \Omega(\max\{AM_{1/4}(f), AM_{1/4}(\neg f)\})$.

Recall the definition of online interactive proofs from Section 3.1. Similarly to how communication complexity classes are defined in the work of Babai et al. [3], Chakrabarti et al. showed that the following communication complexity classes are equivalent: $\mathsf{OMA}^{[1]} = \mathsf{OIP}^{[1]} = \mathsf{OIP}_{+}^{[1]} = \mathsf{MA}^{[1,A]} = \mathsf{R}^{[1,A]}$[6] [75, Theorem 5.1]. We note here that by changing the criteria for efficiency from polylogarithmic to constant, the result of Chakrabarti et al. remains almost the same in our setting.

Fact 56 ([75, Theorem 5.1]). *For all f, and $\epsilon > 0$, we have $OMA_{\epsilon}^{A\to B}(f) = \Theta(OIP_{\epsilon}^{A\to B}(f)) = \Theta(OIP_{+,\epsilon}^{A\to B}(f)) = \Theta(MA_{\epsilon}^{A\to B}(f))$. Moreover, $MA_{\epsilon}^{A\to B}(f) \leqslant R_{\epsilon}^{A\to B,\mathrm{pub}}(f) \leqslant (MA_{\epsilon}^{A\to B}(f))^2$.*

Introduced by Göös and Watson [89], the small bounded-error probabilities communication complexity, denoted as $SBP(f)$, is given by $SBP(f) \coloneqq \min_{\alpha(n)>0} R_{\alpha,\alpha/2}^{\mathrm{pub}}(f) + \log(1/\alpha)$, where $R_{\alpha,\beta}^{\mathrm{pub}}(f)$ is the public-coin randomized communication complexity of a protocol that computes f with probability of acceptance of at least $\alpha(n)$ on all 1-inputs and at most $\beta(n)$ on all 0-inputs of length n. The authors showed that this complexity measure is exactly characterized by the rectangle bound.

Fact 57 ([89, Theorem 1]). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $0 < \epsilon \leqslant 1/8$. Then, $SBP(f) = \Theta(\mathrm{rec}_{\epsilon}(f))$*[7] *for all f.*

[6] $\mathsf{MA}^{[1,A]}$ denotes the one-round (or one-way) MA complexity where Alice is the first to send a message, whereas $\mathsf{R}^{[1,A]}$ denotes the one-round (or one-way) public-coin randomized communication complexity with Alice being the first to send a message.

[7] Although it is written in [89] that $SBP(f) = \Theta(\mathrm{rec}_{1/8}(f))$, one can replace the constant factor of 1/8 by any positive constant at most 1/8 while affecting the rectangle bound by only a constant factor [72].

Other lower bound techniques such as the relative discrepancy bound, the positive partition bound, the weak partition bound, the public-coin partition bound, the adaptive relative discrepancy and the weak regularity are also shown to be related to each other as well as to the public-coin randomized communication co,complexity. We note that although the partition bound was mentioned earlier in Definition 6, we shall restate it below as a fixed-distribution version, equivalent to the partition bound in Definition 6 in the worst case distribution.

Definition 14 (Partition bound (fixed-distribution variant) [90, Definition 5]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and let $\epsilon > 0$. For a distribution, μ, the partition bound of f with respect to μ, denoted as $\mathrm{prt}^{\mu}_{\epsilon}(f)$,is the optimal value of the following problem:*

$$\begin{aligned} \max_{\kappa,\phi_{x,y}} \quad & \phi - \epsilon\kappa \\ st \quad & \phi(R) - \kappa\mu(R \cap f^{-1}(z)) \leqslant 1 && \forall R, z \\ & \kappa \geqslant 0. \end{aligned}$$

Furthermore, $\mathrm{prt}_{\epsilon}(f) = \max_{\mu} \mathrm{prt}^{\mu}_{\epsilon}(f)$.

Definition 15 (Relative discrepancy bound [90, Definition 3], [91]). *Let μ be a distribution over $\mathcal{X} \times \mathcal{Y}$ and let $f : \mathrm{supp}(\mu) \to \{0,1\}$ be a function[8]. The distributional relative discrepancy bound of f, denoted as $\mathrm{rdisc}^{\mu}_{\epsilon}(f)$, is the optimal value of the following problem:*

$$\begin{aligned} \sup_{\alpha,\delta,\phi_{xy}} \quad & \frac{1}{\delta}\left(\frac{1}{2} - \alpha - \epsilon\right) \\ st \quad & \left(\frac{1}{2} - \alpha\right) \cdot \phi(R) \leqslant \mu(R \cap f^{-1}(z)) \quad \forall R, z \text{ such that } \phi(R) \geqslant \delta \\ & \sum_{xy} \phi_{xy} = 1 \\ & 0 \leqslant \alpha < \frac{1}{2}, 0 < \delta < 1, \phi_{xy} \geqslant 0 \quad \forall (x,y). \end{aligned}$$

For the non-distributional case, the relative discrepancy bound $\mathrm{rdisc}_{\epsilon}(f) = \max_{\mu} \mathrm{rdisc}^{\mu}_{\epsilon}(f)$, where the maximum is taken over all distribution μ over $\mathcal{X} \times \mathcal{Y}$, implicitly adding nonnegativity and normalization constraints on μ.

Definition 16 (Positive partition bound [90, Definition 6]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a (possibly partial) Boolean function and let $\epsilon > 0$. For a distribution μ, the positive partition bound of f, denoted as $\mathrm{prt}^{+,\mu}_{\epsilon}(f)$, is the optimal value of the following maximization problem:*

$$\begin{aligned} \max_{\kappa,\phi_{xy}} \quad & \phi - \epsilon\kappa \\ st \quad & \phi(R) - \kappa\mu(R \cap f^{-1}(z)) \leqslant 1 && \forall R, z \\ & \kappa \geqslant 0, \quad \phi_{xy} \geqslant 0 && \forall (x,y). \end{aligned}$$

The positive partition bound is defined as $\mathrm{prt}^{+}_{\epsilon}(f) = \max_{\mu} \mathrm{prt}^{+,\mu}_{\epsilon}(f)$.

Definition 17 (Weak partition bound [90, Definition 7]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ denote a (possibly partial) Boolean function and let $\epsilon > 0$. For a distribution μ, the weak partition bound, denoted as $\mathrm{wprt}^{\mu}_{\epsilon}(f)$, is the optimal value of the problem below:*

$$\begin{aligned} \max_{\kappa,\phi_{xy}} \quad & \phi - \epsilon\kappa \\ st \quad & \phi(R) - \kappa\mu(R \cap f^{-1}(z)) \leqslant 1 && \forall R, z \\ & \kappa \geqslant 0, \quad \phi_{xy} \geqslant 0, \quad \kappa\mu_{xy} - \phi_{xy} \geqslant 0 && \forall (x,y). \end{aligned}$$

The weak partition bound is defined as $\mathrm{wprt}_{\epsilon}(f) = \max_{\mu} \mathrm{wprt}^{\mu}_{\epsilon}(f)$.

[8] As in [90], $\mathrm{supp}(\mu) = \mathrm{supp}(f)$ is assumed.

Definition 18 (Public-coin partition bound [92, Definition 2]). *Let $\subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $\epsilon > 0$. The ϵ-public-coin partition bound pf f, denoted as $\mathrm{pprt}_\epsilon(f)$, is the optimal value of the linear program below.*

Primal

$$\begin{aligned} \textit{Min} \quad & \sum_z \sum_R w_{z,R} \\ \textit{st} \quad & \forall (x,y): \sum_{z:(z,y,z)\in f} \sum_{R:(x,y)\in R} w_{z,R} \geqslant 1-\epsilon, \\ & \forall (x,y): \sum_{R:(x,y)\in R} \sum_z w_{z,R} = 1 \\ & \forall (z,R): w_{z,R} = \sum_{P:(z,R)\in P} a_P, \\ & \sum_P a_P = 1, \\ & \forall (z,R): w_{z,R} \geqslant 0; \quad \forall P: a_P \geqslant 0. \end{aligned}$$

Dual

$$\begin{aligned} \textit{Max} \quad & (1-\epsilon)\sum_{(x,y)} \mu_{x,y} + \sum_{(x,y)} \phi_{x,y} + \lambda \\ \textit{st} \quad & \forall (z,R): \sum_{(x,y)\in R:(x,y,z)\in f} \mu_{x,y} + \sum_{(x,y)\in R} \phi_{x,y} + v_{z,R} \leqslant 1, \\ & \forall P: \sum_{(z,R)\in P} v_{z,R} \geqslant \lambda, \\ & \forall (x,y): \mu_{x,y} \geqslant 0, \phi_{x,y} \in \mathbb{R}; \quad \forall (z,R): v_{z,R} \in \mathbb{R} \\ & \lambda \in \mathbb{R}. \end{aligned}$$

Here, R represents a rectangle in $\mathcal{X} \times \mathcal{Y}$ and P is a partition along with outputs $z \in \mathcal{Z}$, i.e. $P = \{(z_1, R_1), \cdots, (z_m, R_m)\}$ such that $\{R_1, \cdots, R_m\}$ form a partition of $\mathcal{X} \times \mathcal{Y}$ into rectangles and $z_i \in \mathcal{Z}$ for all $i \in [m]$.

Fact 58 (Adaptive relative discrepancy [90, Definition 9], [91]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a (possibly partial) Boolean function and let $\epsilon > 0$. For a distribution μ, the adaptive relative discrepancy of f with respect to μ, denoted as $\mathrm{ardisc}^\mu_\epsilon(f)$, is the optimal value of the optimization problem below:*

$$\begin{aligned} \sup_{\alpha,\delta,\phi^P_{x,y}} \quad & \frac{1}{\delta}\left(\frac{1}{2} - \alpha - \epsilon\right) & \\ \textit{st} \quad & \left(\frac{1}{2} - \alpha\right)\phi^P(R) \leqslant \mu(R \cap f^{-1}(z)) & \forall P, \forall (z,R) \in P \ \textit{st}\ \phi^P(R) \geqslant \delta \\ & \phi^P = 1 & \forall P \\ & 0 \leqslant \alpha < \frac{1}{2}, 0 < \delta < 1, \phi^P_{x,y} \geqslant 0 & \forall P, \forall (x,y) \end{aligned}$$

Furthermore, $\mathrm{ardisc}_\epsilon(f) = \max_\mu \mathrm{ardisc}^\mu_\epsilon(f)$.

Definition 19 (Weak regularity [93, Definition 2.14] [94]). *Let $f : \mathcal{X} \times \mathcal{Y} \to Z$ be a total function and μ be a distribution over $\mathcal{X} \times \mathcal{Y}$. The weak regularity of f with respect to μ, denoted by $\mathrm{wreg}^\mu_\epsilon(\mathrm{f})$, is the optimal value of the following optimization problem:*

$$\begin{aligned} \min \quad & \delta & \\ \textit{st} \quad & \mu(R \cap f^{-1}(z)) \geqslant \frac{1}{|\mathcal{Z}|}(\mu(R) - \delta) & \forall R, z. \end{aligned}$$

Alternatively,

$$\mathrm{wreg}^\mu_\epsilon(f) = \max_{R,z} \mu(R) - |\mathcal{Z}|\mu(R \cap f^{-1}(z)).$$

We say that f is δ-weakly regular with respect to μ for any $\delta \geqslant \mathrm{wreg}^\mu(f)$.

Fact 59 ([90, Corollary 2]). *For any μ, $f : \mathrm{supp}(\mu) \to \{0,1\}$ and $\epsilon \in (0, 1/8)$, we have* $\log\left(\mathrm{ardisc}^\mu_\epsilon(f)\right) \leqslant D^\mu_\epsilon(f) \leqslant \left(\log \mathrm{ardisc}^\mu_{\epsilon/8}(f) + 2\log\frac{1}{\epsilon} + 6\right)^2$.

Fact 60 ([90, Proposition 1]). *For all f, μ, ϵ, we have* $\mathrm{wprt}^\mu_\epsilon(f) \leqslant \mathrm{prt}^{+,\mu}_\epsilon(f) \leqslant \mathrm{prt}^\mu_\epsilon(f)$ *and* $\mathrm{wprt}^\mu_\epsilon(f) \leqslant \overline{\mathrm{prt}}^\mu_\epsilon(f) \leqslant \mathrm{prt}^\mu_\epsilon(f)$.

Fact 61 ([92, Lemma 1]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $\epsilon > 0$. Then,* $\log\left(\text{pprt}_\epsilon(f)\right) \leqslant R_\epsilon^{\text{pub}}(f)$.

Fact 62 ([92, Theorem 1]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $\epsilon > 0$. Then* $R_{2\epsilon}^{\text{pub}}(f) \leqslant \left(\log \text{pprt}_\epsilon(f) + \log \frac{1}{\epsilon} + 1\right)^2$.

Fact 63 ([90, Theorem 5]). *Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$ and let f be a Boolean function on the support of μ such that either* $\text{rdisc}_\epsilon^\mu(f) \geqslant 1$ *or* $\text{prt}_{4\epsilon}^{+,\mu}(f) > 2$ *. Then, for any $\epsilon \in (0, 1/4)$, we have* $\frac{\epsilon}{2}\,\text{prt}_{4\epsilon}^{+,\mu}(f) \leqslant \text{rdisc}_\epsilon^\mu(f) \leqslant \text{prt}_\epsilon^{+,\mu}(f)$.

Fact 64 ([90, Theorem 8]). *For any distribution μ, for any function $f : \text{supp}(f) \to \{0,1\}$ and $\epsilon \in (0, 1/4)$ such that either* $\text{ardisc}_\epsilon^\mu(f) \geqslant 1$ *or* $\text{pprt}_{4\epsilon}^\mu(f) > 2$*, we have* $\frac{\epsilon}{2}\,\text{pprt}_{4\epsilon}^\mu \leqslant \text{ardisc}_\epsilon^\mu(f) \leqslant \text{pprt}_\epsilon^\mu(f)$.

Fact 65 ([93, Proposition 2.16]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then,* $\text{wereg}^\mu(f) = \text{disc}^\mu(f)$.

Fact 66 ([93, Proposition 2.18]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a total function. Then,* $\text{wprt}_\epsilon^\mu(f) \geqslant \frac{1-\epsilon\cdot\frac{|\mathcal{Z}|}{|\mathcal{Z}|-1}}{\text{wreg}^\mu(f)}$.

In the model of zero-communication protocols and inspired by the physics setup of Bell experiments, Laplante, Lerays and Roland [95] introduced a new lower bound on quantum communication complexity that subsumes the factorization norm [40, Theorem 1] when the players share a quantum sate. With slight modifications from [96, Definition 4], we define the quantum partition bound below.

Definition 20 (Quantum partition bound [96, Definition 4]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and let $\epsilon > 0$. Let $\perp$ be a special character. The quantum partition bound of f with error ϵ, denoted by* $\text{eff}_\epsilon^*(f)$*, is given by the optimal value of the following non-linear program:*

$$
\begin{aligned}
\max \quad & \frac{1}{\eta} \\
st \quad & \sum_{(x,y):f(x,y)=1} q(f(x,y)|x,y) \geqslant (1-\epsilon)\eta, && \forall x,y \in \mathcal{X}\times\mathcal{Y} \\
& \sum_{f(x,y)\in\{0,1\}} q(f(x,y)|x,y) = \eta && \forall x,y \in \mathcal{X}\times\mathcal{Y} \\
& q(f(x,y)|x,y) \in \mathcal{Q}(\{0,1\}\cup\perp, \mathcal{X}, \mathcal{Y})
\end{aligned}
$$

Based on the fact that the communication complexity of predicates generalizes the communication complexity of (total or partial) functions, the next result, originally stated for predicates, holds true.

Lemma 5. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{-1,1\}$ and let $\epsilon > 0$. When $\alpha = \frac{1+2\epsilon}{1-2\epsilon}$, we have* $\text{eff}_\epsilon^*(f) \geqslant \gamma_2^\alpha(f)$.

Proof. For a total function $f : \mathcal{X}\times\mathcal{Y} \to \{-1,1\}$, let $\vee_f$ denote the predicate on $(\{-1,1\})^2 \times (\mathcal{X}\times\mathcal{Y})$ given by $\vee(ab, xy) = 1 \Leftrightarrow a \cdot b = f(x,y)$. Note that for a function f, one can express $\gamma_2^\alpha(f)$ as $\max_\mu \gamma_2^{\alpha,\mu}(f)$, where $\gamma_2^{\alpha,\mu}(f)$ is a distributional version of $\gamma_2^\alpha(f)$. Then, we have

$$\text{eff}_\epsilon^*(f) \geqslant \max_\mu \text{eff}_\epsilon^{*,\mu}(\vee_f) \geqslant (1-2\epsilon)\max_\mu \gamma_2^{\alpha,\mu}(f) = (1-2\epsilon)\gamma_2^\alpha(f),$$

where the second inequality is due to [96, Theorem 3]. □

Given a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, and error parameter ϵ, the *distributional information complexity* of f with error ϵ, denoted as $IC_\epsilon^\mu(f)$, is defined to be the infimum of the information

cost (see Section 2) over all (randomized) protocols π that achieve an error of at most ϵ with respect to the distribution μ [50, Definition 3.2]. Mathematically speaking,

$$IC^{\mu}_{\epsilon}(f) := \inf_{\pi:\Pr_{(x,y)\sim\mu}[\pi(x,y)\neq f(x,y)]\leqslant\epsilon} IC^{\mu}(\pi). \tag{3}$$

For a function f with error ϵ, its *max-distributional information complexity* is given by [50, Definition 3.3]

$$IC_{D,\epsilon}(f) := \max_{\mu \text{ a distribution on } \mathcal{X}\times\mathcal{Y}} IC^{\mu}_{\epsilon}(f), \tag{4}$$

whereas its *information complexity* is defined as [50, Definition 3.4]]

$$IC_{\epsilon}(f) := \inf_{\pi \text{ is a protocol with } \Pr[\pi(x,y)\neq f(x,y)]\leqslant\epsilon \text{ for all } (x,y)} \max_{\mu} IC^{\mu}(\pi). \tag{5}$$

It is easy to see that the information complexity of a function lower bounds its max-distributional information complexity.

Fact 67 ([50])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function and let $\epsilon > 0$. Then, $IC_{\epsilon}(f) \geqslant IC_{D,\epsilon}(f)$.*

Proof. First, observe that for any distribution μ, we have $IC^{\mu}(\pi) \geqslant IC^{\mu}_{\epsilon}(f)$. Then, taking the maximum over μ on both sides gives $\max_{\mu} IC^{\mu}(\pi) \geqslant \max_{\mu} IC^{\mu}_{\epsilon}(f)$. By Equations (3) and (5),we have

$$IC_{\epsilon}(f) = \inf_{\pi}\max_{\mu} IC^{\mu}(\pi) = \inf_{\pi} IC^{\mu_1}(\pi) = IC^{\mu_1}_{\epsilon}(f),$$

where μ_1 is the distribution that maximizes $IC^{\mu}(\pi)$. From the observation made earlier, we can write

$$IC^{\mu_1}_{\epsilon}(f) \geqslant IC^{\mu_2}_{\epsilon}(f),$$

where μ_2 is the distribution that maximizes $IC^{\mu}_{\epsilon}(f)$. By Equation (4), we get

$$C_{\mu_2}(f,\epsilon) = \max_{\mu} C_{\mu}(f,\epsilon) = IC_D(f,\epsilon).$$

□

Together with the result below, max-distributional information complexity characterizes the information complexity.

Fact 68 ([50, Theorem 3.5])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be any function, and let $\epsilon \geqslant 0$ be an error parameter. For all $0 < \alpha < 1$, we have $IC_{\epsilon/\alpha}(f) \leqslant \frac{IC_{D,\epsilon}(f)}{1-\alpha}$.*

Kerenidis et al. proved that the information complexity of a function f is bounded from below by the relaxed partition bound [85].

Fact 69 ([85, Theorem 1])**.** *There exists a positive constant C such that for all functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. all $\epsilon, \delta \in (0, \frac{1}{2}]$ and all distributions μ, we have $IC^{\mu}_{\epsilon}(f) \geqslant \frac{\delta^2}{C} \cdot \left(\log \bar{\text{prt}}^{\mu}_{\epsilon+3\delta}(f) - \log 2\right) - \delta$.*

On the other hand, Braverman and Weinstein showed that information complexity can also be lower bounded by the discrepancy bound [97]. A direct product theorem implicit in the work of [98] gives an upper bound on information complexity in terms of the distributional communication complexity.

Fact 70 ([97, Theorem 1.1]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function and let μ be any distribution on $\mathcal{X} \times \mathcal{Y}$. Then, $IC^{\mu}_{1/10}(f) \geqslant \Omega\left(\log\left(\frac{1}{\mathrm{disc}^{\mu}(f)}\right)\right)$.*

Fact 71 ([49, Theorem 2.4], [98]). *For every f, ϵ, μ, there exists a protocol π that computes f on inputs drawn from μ with probability of error at most ϵ and communication at most $D^{\mu^k}_{\rho}(f^k)$ such that $IC^{\mu}(\pi) \leqslant \frac{2D^{\mu^k}_{\epsilon}(f^k)}{k}$.*

Intuitively, the communication cost $CC(\pi)$ of a protocol π is an upper bound on its information cost over any distribution μ, since there cannot be more information leaked than the length of the messages exchanged. In other words, $IC^{\mu}(\pi) \leqslant CC(\pi)$. Another way of seeing this is by the observation that the information cost of a protocol π is always bounded by its length $|\pi|$, and therefore communication complexity is always an upper bound on information complexity.

Fact 72. *For any function f and error parameter $\epsilon > 0$, $IC_{\epsilon}(f) \leqslant R^{\mathrm{pub}}_{\epsilon}(f)$.*

Proof. The public-coin randomized communication complexity of a function f with error ϵ is defined as

$$R^{\mathrm{pub}}_{\epsilon}(f) = \min_{\pi} \max_{(x,y)\sim\mu} CC(\pi(x,y)). \tag{6}$$

According to Equation (4), its corresponding information complexity is

$$IC_{\epsilon}(f) = \inf_{\pi} \max_{\mu} IC^{\mu}(\pi)). \tag{7}$$

Comparing Equations (6) and (7), we get $IC_{\epsilon}(f) \leqslant R^{\mathrm{pub}}_{\epsilon}(f)$ using the fact that information cost of a protocol lower bounds its communication cost. □

Chakrabarti et al. [99] were the first to define the external information cost explicitly for the purpose of studying communication complexity. The external information cost is the amount of information an observer necessarily has to learn when the parties perform the computation, and hence is always smaller than the communication complexity. The natural analogues of Equations (3), (4) and (5) for external information complexity are as follows: given a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, and error parameter ϵ, the *distributional external information complexity* is given by [50, Definition 3.13]

$$IC^{\mathrm{ext},\mu}_{\epsilon}(f) := \inf_{\pi:\Pr_{(x,y)\sim\mu}[\pi(x,y)\neq f(x,y)]\leqslant\epsilon} IC^{\mathrm{ext},\mu}(\pi), \tag{8}$$

while its *max-distributional external information complexity* is defined as [50, Definition 3.14]

$$IC^{\mathrm{ext}}_{D,\epsilon}(f) := \max_{\mu \text{ a distribution on } \mathcal{X}\times\mathcal{Y}} IC^{\mathrm{ext},\mu}_{\epsilon}(f). \tag{9}$$

Finally, the *external information complexity* of f with error ϵ is [50, Definition 3.15]

$$IC^{\mathrm{etc}}_{\epsilon}(f) := \inf_{\pi\pi \text{ is a protocol with } \Pr[\pi(x,y)\neq f(x,y)]\leqslant\epsilon \text{ for all } (x,y)} \max_{\mu} IC^{\mathrm{ext},\mu}(\pi). \tag{10}$$

Braverman [50] showed that the (internal) information complexity is always upper bounded by the external information complexity.

Fact 73 ([50, Proposition 3.12]). *For every protocol π and distribution μ, $IC^{\mathrm{ext},\mu}(\pi) \geqslant IC^{\mu}(\pi)$.*

Fact 73 implies the following relationship between internal and external information complexity.

Fact 74 ([50]). *Let f be a function. For all distributions μ , $IC_{\epsilon}^{\text{ext},\mu}(f) \geqslant IC_{\epsilon}^{\mu}(f)$, $IC_{D,\epsilon}^{\text{ext}}(f) \geqslant IC_{D,\epsilon}(f)$, $IC_{\epsilon}^{\text{ext}}(f) \geqslant IC_{\epsilon}(f)$.*

The external information analogues of Facts 67 and 68 are as below.

Fact 75 ([50]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and let $\epsilon > 0$. Then, $IC_{\epsilon}^{\text{ext}}(f) \geqslant IC_{D,\epsilon}^{\text{ext}}(f)$.*

Fact 76 ([50, Theorem 3.16]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be any function, and let $\epsilon \geqslant 0$ be an error parameter. For each value of the parameter $0 < \alpha < 1$, we have $IC_{\epsilon/\alpha}^{\text{ext}}(f) \leqslant \frac{IC_{D,\epsilon}^{\text{ext}}(f)}{1-\alpha}$.*

Analogously to how the internal information cost of a protocol is upper bounded by its communication cost over any distribution on the inputs, the external information cost of a protocol is also upper bounded by its information cost, i.e., $IC_{\mu}^{\text{ext}}(\pi) \leqslant CC(\pi)$. It follows that external information complexity is always upper bounded by communication complexity [49]. The proof is similar to that of Fact 72, where one replaces the definition of internal information complexity with its external analogue.

Fact 77. *For any function f and error parameter $\epsilon > 0$, $IC_{\epsilon}^{\text{ext}}(f) \leqslant R_{\epsilon}^{\text{pub}}(f)$.*

The amortized communication complexity of a function f with respect to distribution μ, denoted as $AC_{\epsilon}^{\mu}(f)$, is defined as $AC_{\epsilon}^{\mu}(f) := \lim_{k\to\infty} \frac{D_{\epsilon}^{\mu^k}(f^k)}{k}$ when the limit exists. In a recent paper, Braverman and Roo showed that information complexity exactly equals to amortized communication complexity [100].

Fact 78 ([100, Theorem 6.3]). *For any function f, distribution μ and error ϵ, $AC_{\epsilon}^{\mu}(f) = IC_{\epsilon}^{\mu}(f)$.*

The partition bound $\text{prt}_{\epsilon}(f)$ was shown to have an information-theoretic interpretation in terms of Rényi entropy of order ∞ [101]. Let f be a relation, μ be a distribution over $\mathcal{X} \times \mathcal{Y}$ and $\epsilon > 0$. The Rényi information cost of f over μ with error ϵ, denoted as $IC_{\epsilon}^{\infty,\mu}(f)$, is the minimum external information cost of a protocol π that computes f over μ with error at most ϵ [93, Definition 2.32], i.e.

$$IC_{\epsilon}^{\infty,\mu}(f) = \min_{\pi \text{ computing } f \text{ with error } \epsilon} I_{\infty}(X, Y : \pi, R).$$

Here, $I_{\infty}(U : V) = \log\left(\sum_{v\in\mathcal{V}} \sum_{u\in\mathcal{U}} p_{V|U}(v|u)\right)$ is the Rényi mutual information, where U, V are random variables over unitaries $\mathcal{U}, \mathcal{V}$ and $p_{V|U}(v|u) = \frac{\Pr[U=u \wedge V=v]}{\Pr[U=u]}$ [93, Definition 2.31].

The pseudotranscript complexity introduced by [101] is a complexity measure based on combinatorial properties of a function's truth table. Consider a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Define a random variable Q over a space $\mathcal{Q}$ to be a *pseudotranscript* for f if there exist two functions $g : \mathcal{Q} \times \mathcal{X} \to \mathbb{R}^+$ and $h : \mathcal{Q} \times \mathcal{Y} \to \mathbb{R}^+$ such that $\Pr[Q = q | X = x, Y = y] = g(q, x)h(q, y)$, for all $q \in \mathcal{Q}, x \in \mathcal{X}, y \in \mathcal{Y}$. Specifically, the transcript of a deterministic or private coin protocol is a pseudotranscript, and the transcript of a public coin protocol concatenated with the public coins is also a pseudotranscript. We say that a pseudotranscript computes a relation f if there exists a mapping $\mathcal{M} : \mathcal{Q} \to \mathcal{Z}$ such that $\Pr[\mathcal{M}(\mathcal{Q}) \in f(X, Y)] \geqslant 1 - \epsilon$. The Rényi information cost of a pseudotranscript is defined as a relaxation of $IC_{\epsilon}^{\infty,\mu}(f)$ by replacing the communication protocol with a pseudotranscript in its definition.

Definition 21 (Rényi pseudoinformation cost of a relation [93, Definition 2.36]). *Let f be a relation, μ be a distribution and $\epsilon > 0$. Let inputs X, Y be random variables and let Q be a pseudotranscript relative to X, Y. The Rényi pseudoinformation cost of f, denoted by $pIC_{\epsilon}^{\infty}(f, \mu)$, is given by*

$$pIC_{\epsilon}^{\infty,\mu}(f) = \min_{Q \text{ computing } f \text{ with error } \epsilon} IC_{\epsilon}^{\infty,\mu}(Q).$$

Prabhakaran and Prabhakaran showed the equivalence between $pIC^{\infty}(f)$ and the partition bound [101].

Fact 79 ([93, Theorem 2.37], [101, Theorem 3]). *Let $f : \mathcal{X} \times \mathcal{Y} \to 2^{\mathcal{Z}}$ be any relation, μ be a distribution and $\epsilon \in [0, 1/2)$. Then, $pIC_{\epsilon}^{\infty,\mu}(f) = \log \mathrm{prt}_{\epsilon}^{\mu}(f)$.*

Fact 80 ([101, Theorem 2]). *Let $f : \mathcal{X} \times \mathcal{Y} \to 2^{\mathcal{Z}}$ be any relation and let $\epsilon \in [0, 1/2)$. Then, $pIC_{\epsilon}^{\infty}(f) = \log \mathrm{prt}_{\epsilon}(f)$.*

Touchette [102] introduced quantum information complexity as an extension of (internal) information complexity to the quantum setting. This is inspired by quantum state-redistribution protocols [103, 104]. Quantum information complexity has been shown to satisfy many of the natural properties possessed by information complexity. In particular, it is equal to amortized quantum communication complexity. We first review the definitions of a few measures relating to quantum information complexity before stating the results.

We follow the quantum communication model defined in the work of Anshu et al. [105] , which is close to the one defined by Cleve and Buhrman [106], where players share prior entanglement and are allowed to send quantum messages to each other. In particular, a k-round protocol π for a classical problem from input registers $A_{in} = X, B_{in} = Y$ to output registers A_{out}, B_{out} is defined by a sequence of isometries $U_1, \cdots, U_{k+1}$ with a pure state ψ shared between Alice and Bob. After the application of U_i, Alice holds register A_i, Bob holds register B_i and the communication register is C_i. Readers are referred to [105] for more details on the model.

Let π be a protocol and let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$. The *quantum communication cost* and *quantum information complexity* of π on input μ, denoted by $QCC^{\mu}(\pi)$ and $QIC^{\mu}(\pi)$ respectively, are given by

$$QCC^{\mu}(\pi) = \sum_i |C_i|$$
$$QIC^{\mu}(\pi) = \sum_{i \geqslant 1, \text{ odd}} I(C_i : R_X, R_Y | B_i) + \sum_{i \geqslant 1, \text{ even}} I(C_i :: R_X, R_Y | A_i),$$

where R_X, R_Y are the canonical purification of input X, Y, $I(X : Z|Y) \coloneqq \mathbb{E}_{y \leftarrow Y}[I(X : Z|Y = y)] = S(X|Y) + S(Z|Y) - S(XZ|Y)$ is the condition mutual information [105]. For any function f, any distribution μ and any error $\epsilon > 0$,

$$QCC_{\epsilon}^{\mu}(f) = \min_{\pi} QCC^{\mu}(\pi)$$
$$QIC_{\epsilon}^{\mu}(f) = \min_{\pi} QIC^{\mu}(\pi).$$

The *max-distributional quantum information complexity* of f error ϵ is [87, Definition 4.1]

$$QIC_{D,\epsilon}(f) = \max_{\mu \text{ is a distribution on } \mathcal{X} \times \mathcal{Y}} QIC_{\epsilon}^{\mu}(f). \tag{11}$$

The quantum analogue of Fact 72 is as below.

Fact 81. *For any function f, $QIC_{\epsilon}(f) \leqslant Q_{\epsilon}^{*}(f)$.*

Let f be a function, μ be a distribution on the inputs $\mathcal{X} \times \mathcal{Y}$ and $\epsilon > 0$. The *amortized quantum communication complexity* of a function f under the input distribution μ and with error $\epsilon > 0$, denoted by $AQCC_{\epsilon}^{\mu}(f)$ is given by

$$AQCC_{\epsilon}^{\mu}(f) \coloneqq \lim_{k \to \infty} QCC_{\epsilon}^{\mu^k}(f^k).$$

Fact 82 ([87]). *For any f and $\epsilon > 0$, we have $QIC_{D,\epsilon}(f) \leqslant QIC_{\epsilon}(f)$.*

Fact 83 ([87, Theorem 4.13]). *Let f be a relation and $\alpha, \epsilon \in (0,1)$. Then, $QIC_{\epsilon/\alpha}(f) \leqslant \frac{QIC_{D,\epsilon}(f)}{1-\alpha}$.*

Fact 84 ([102, Theorem 2]). *For any function f, distribution μ and error $\epsilon > 0$, we have $AQCC^{\mu}_{\epsilon}(f) = QIC^{\mu}_{\epsilon}(f)$.*

Fact 85 ([107, Corollary 5.8]). *For all Boolean functions f, $QCC_{1/3}(f) \leqslant 2^{O(QIC_{1/3}(f)+1)}$.*

Fact 86 ([105]). *For any function f, distribution μ and error $\epsilon > 0$, we have $AQCC^{\mu}_{\epsilon}(f) \leqslant AC^{\mu}_{\epsilon}(f)$.*

Fact 87 ([87, Theorem 5.7]). *There exists an absolute constant $\delta > 0$ such that for any Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, we have $QIC_{D,\delta}(f) \geqslant \Omega(\mathrm{sdisc}_{1/5}(f) - O(1))$.*

3.4 Class 4

This class consists of only two complexity measures. Namely, the unbounded error probabilistic communication complexity and the sign-rank. This constant equivalence relationship is shown in Figure 6.

$$UPP(f) \xleftrightarrow{\text{Fact 88}} \mathrm{signrank}(f)$$

Figure 6: Relationship between complexity measures in Class 4.

Theorem 4. *Class 4 includes the following complexity measures:*

- $UPP(f)$
- $\mathrm{signrank}(f)$

Proof. We have $UPP(f) \leftrightarrow \mathrm{signrank}(f)$ by Fact 88. □

Before stating the fact used in the proof of Theorem 4, we give definitions of the sign-rank and the unbounded error probabilistic communication complexity.

Sign-rank is one of the important analytic notions in communication complexity. The sign-rank of a matrix M is the minimal rank of a real matrix whose entries have the same sign pattern as M.

Definition 22 (sign-rank [108, Definition 1]). *For a real matrix M with nonzero entries, let $\mathrm{sign}(M)$ denote its sign matrix such that $(\mathrm{sign}(M))_{i,j} = \mathrm{sign}(M_{i,j})$ for all i, j. The sign-rank of a matrix S, denoted by $\mathrm{signrank}(S)$, is defined as $\mathrm{signrank}(S) := \min\{\mathrm{rank}(M) : \mathrm{sign}(M) = S\}$.*

The unbounded error probabilistic communication complexity was introduced by Paturi and Simon [109]. Given a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, an unbounded error protocol for f is a protocol π such that $\Pr[\pi(x,y) = f(x,y)] > 1/2$, where the probability is over the coin tosses of the players. The worst-case cost of π is the maximum number of bits transmitted during an execution of π, taken over all choices of x and y. The unbounded error communication complexity of f, denoted as $UPP(f)$, is the minimum cost of an unbounded error protocol for f. In their work, Paturi and Simon showed the equivalence relation between $UPP(f)$ and the sign-rank of a matrix.

Fact 88 ([109, Theorem 3]). *Let $f : \mathcal{X} \times \mathcal{U} \to \{0,1\}$. Then, we have $\lceil \log \mathrm{signrank}(f) \rceil \leqslant UPP(f) \leqslant \lceil \log \mathrm{signrank}(f) \rceil + 1$*

3.5 Class 5

This class consists of measures from learning complexity and complexity measures under the product distribution. A product distribution μ on $\mathcal{X} \times \mathcal{Y}$ is a distribution that can be expressed as $\mu(x, y) = \mu_{\mathcal{X}}(\mathcal{X}) \times \mu_{\mathcal{Y}}(\mathcal{Y})$, where $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$ are distributions over $\mathcal{X}$ and $\mathcal{Y}$, respectively. The elements in this class consist of complexity measures under product distribution. We use [] at the superscript to denote the corresponding ding complexity measure under product distributions. For instance, $\mathcal{C}^{[]}(f)$ denotes the complexity measure $\mathcal{C}$ under product distributions. The constant equivalence relationship between these complexity measures are presented in Figure 7.

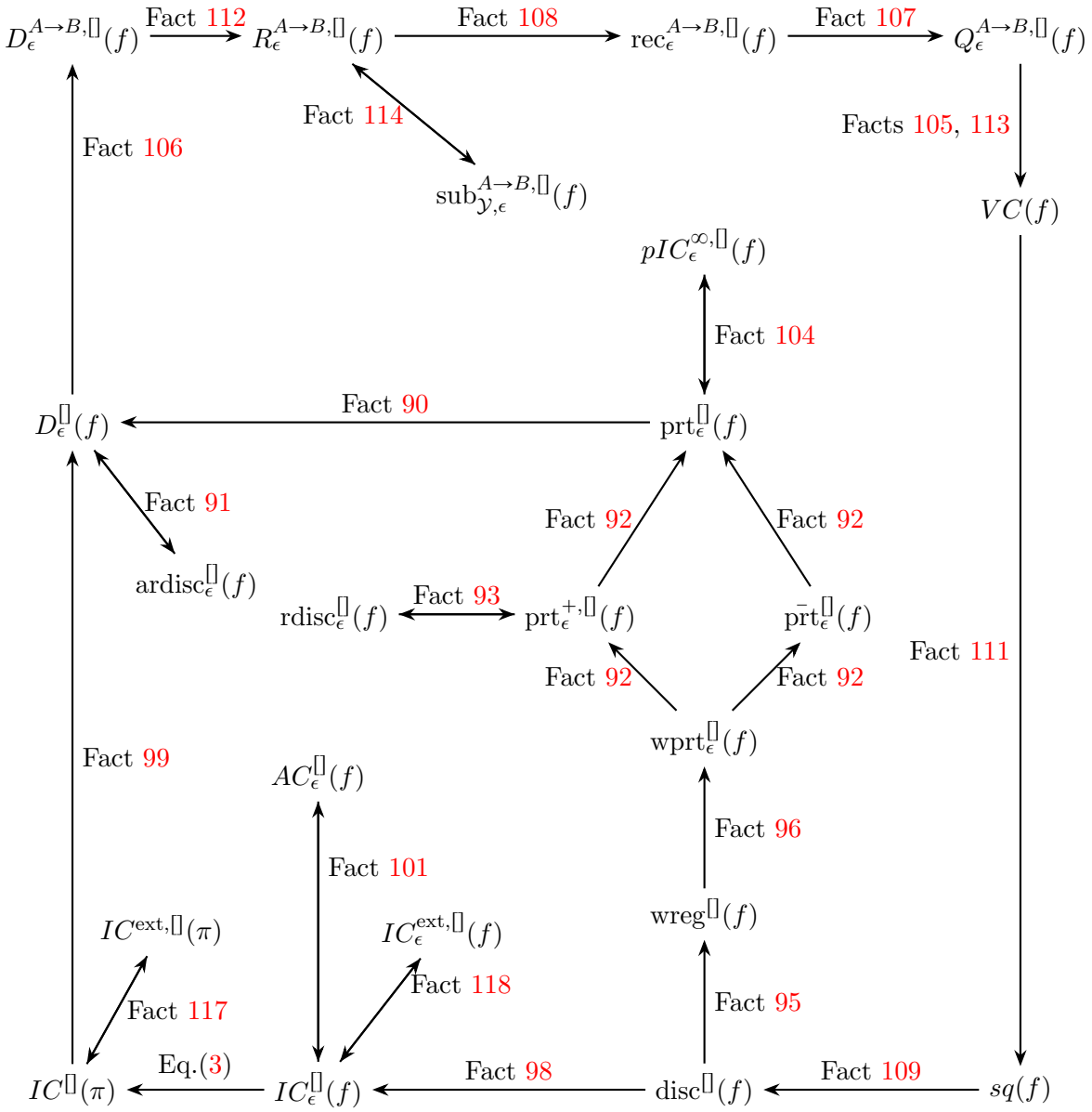


Figure 7: Relationship between complexity measure in Class 5.

Theorem 5. *Class 5 includes the following complexity measures:*

- $D_\epsilon^{[]}(f)$
- $D_\epsilon^{A\to B,[]}(f)$
- $R_\epsilon^{A\to B,[]}(f)$
- $Q_\epsilon^{A\to B,[]}(f)$
- $\mathrm{rec}_\epsilon^{A\to B,[]}(f)$
- $\mathrm{sub}_{\mathcal{Y}}^{A\to B,[]}(f,\epsilon)$
- $VC(f)$
- $sq(f)$
- $\mathrm{disc}^{[]}(f)$
- $IC^{[]}(\pi)$
- $IC_\epsilon^{[]}(f)$
- $IC^{\mathrm{ext},[]}(\pi)$

- $IC_\epsilon^{\mathrm{ext},\square}(f)$
- $pIC_\epsilon^{\infty,\square}(f)$
- $AC_\epsilon^{\square}(f)$
- $\mathrm{rdisc}_\epsilon^{\square}(f)$
- $\mathrm{ardisc}_\epsilon^{\square}(f)$
- $\mathrm{prt}_\epsilon^{\square}(f)$
- $\mathrm{prt}_\epsilon^{+,\square}(f)$
- $\bar{\mathrm{prt}}_\epsilon^{\square}(f)$
- $\mathrm{wprt}_\epsilon^{\square}(f)$
- $\mathrm{wreg}_\epsilon(f)$

Proof. First, we have $D_\epsilon^{A\to B,\square}(f) \to R_\epsilon^{A\to B,\square}(f)$ by Fact 112; $R_\epsilon^{A\to B,\square}(f) \to \mathrm{rec}_\epsilon^{A\to B,\square}(f)$ by Fact 108; $\mathrm{rec}_\epsilon^{A\to B,\square}(f) \to Q_\epsilon^{A\to B,\square}(f)$ by Fact 107; $Q_\epsilon^{A\to B,\square}(f) \to VC(f)$ by Facts 105 and 113; $VC(f) \to sq(f)$ by Fact 111; $sq(f) \to \mathrm{disc}^{\square}(f)$ by Fact 109; $\mathrm{disc}^{\square}(f) \to IC_\epsilon^{\square}(f)$ by Fact 98; $IC_\epsilon^{\square}(f) \to IC^{\square}(\pi)$ by Equation 3; $IC^{\square}(\pi) \to D_\epsilon^{\square}(f)$ by Fact 99.

From $\mathrm{disc}^{\square}(f)$, we have $\mathrm{disc}^{\square}(f) \to \mathrm{wreg}^{\square}(f)$ by Fact 95; $\mathrm{wreg}^{\square}(f) \to \mathrm{wprt}_\epsilon^{\square}(f)$ by Fact 96; $\mathrm{wprt}_\epsilon^{\square}(f) \to \mathrm{prt}_\epsilon^{+,\square}(f), \bar{\mathrm{prt}}_\epsilon^{\square}(f)$ by Fact 92; $\mathrm{prt}_\epsilon^{+,\square}(f), \bar{\mathrm{prt}}_\epsilon^{\square}(f) \to \mathrm{prt}_\epsilon^{\square}(f)$ by Fact 92; $\mathrm{prt}_\epsilon^{\square}(f) \to D_\epsilon^{\square}(f)$ by Fact 90.

In addition, $R_\epsilon^{A\to B,\square}(f) \leftrightarrow \mathrm{sub}_{\mathcal{Y},\epsilon}^{A\to B,\square}(f)$ by Fact 114; $D_\epsilon^{\square}(f) \leftrightarrow \mathrm{ardisc}_\epsilon^{\square}(f)$ by Fact 91; $\mathrm{prt}_\epsilon^{\square}(f) \leftrightarrow pIC_\epsilon^{\infty,\square}(f)$ by Fact 104; $\mathrm{prt}_\epsilon^{+,\square}(f) \leftrightarrow \mathrm{rdisc}_\epsilon^{\square}(f)$ by Fact 93; $IC^{\square}(\pi) \leftrightarrow IC^{\mathrm{ext},\square}(\pi)$ by Fact 117; $IC_\epsilon^{\square}(f) \leftrightarrow IC_\epsilon^{\mathrm{ext},\square}(f)$ by Fact 118; $IC_\epsilon^{\square}(f) \leftrightarrow AC_\epsilon^{\square}(f)$ by Fact 101. □

Now, we state facts used in the proof of Theorem 5. We begin by listing results from Section 3.3 that hold for product distribution.

Fact 89 ([82, Proposition 2.1], [37, Proposition 3.28]). *For every function $f : \mathcal{X}\times\mathcal{Y}\to\{0,1\}$ and every $\epsilon \geqslant 0$, $D_{1/2-\epsilon}^{\square}(f) \geqslant \log\left(\frac{2\epsilon}{\mathrm{disc}^{\square}(f)}\right)$.*

Fact 90 ([85, Corollary 1.4]). *For every $\epsilon > 0$ and every $f : \mathcal{X}\times\mathcal{Y}\to\{0,1\}$, $D_\epsilon^{\square}(f) \geqslant \log \bar{\mathrm{prt}}_\epsilon^{\square}(f)$.*

Fact 91 ([90, Corollary 2]). *For any $f : \mathrm{supp}(\mu)\to\{0,1\}$ and $\epsilon \in (0,1/8)$, we have $\log\left(\mathrm{ardisc}_\epsilon^{\square}(f)\right) \leqslant D_\epsilon^{\square}(f) \leqslant \left(\log \mathrm{ardisc}_{\epsilon/8}^{\square}(f) + 2\log\frac{1}{\epsilon} + 6\right)^2$.*

Fact 92 ([90, Proposition 1]). *For all f,ϵ, we have $\mathrm{wprt}_\epsilon^{\square}(f) \leqslant \mathrm{prt}_\epsilon^{+,\square}(f) \leqslant \mathrm{prt}_\epsilon^{\square}(f)$ and $\mathrm{wprt}_\epsilon^{\square}(f) \leqslant \bar{\mathrm{prt}}_\epsilon^{\square}(f) \leqslant \mathrm{prt}_\epsilon^{\square}(f)$.*

Fact 93 ([90, Theorem 5]). *Let $\square$ be a product distribution on $\mathcal{X}\times\mathcal{Y}$ and let f be a Boolean on the support of μ such that either $\mathrm{rdisc}_\epsilon^{\square}(f) \geqslant 1$ or $\mathrm{prt}_{4\epsilon}^{+,\square}(f) > 2$. Then, for any $\epsilon \in (0,1/4)$, we have $\frac{\epsilon}{2}\mathrm{prt}_{4\epsilon}^{+,\square}(f) \leqslant \mathrm{rdisc}_\epsilon^{\square}(f) \leqslant \mathrm{prt}_\epsilon^{+,\square}(f)$.*

Fact 94 ([90, Theorem 6]). *For any distribution μ, for any function $f : \mathrm{supp}(f)\to\{0,1\}$ and $\epsilon \in (0,1/4)$ such that either $\mathrm{ardisc}_\epsilon^{\square}(f) \geqslant 1$ or $\mathrm{pprt}_{4\epsilon}^{\square}(f) \geqslant 2$, we have $\frac{\epsilon}{2}\mathrm{pprt}_{4\epsilon}^{\square} \leqslant \mathrm{ardisc}_\epsilon^{\square}(f) \leqslant \mathrm{pprt}_\epsilon^{\square}(f)$.*

Fact 95 ([93, Proposition 2.16]). *Let $f : \mathcal{X}\times\mathcal{Y}\to\{0,1\}$. Then, $\mathrm{wereg}^{\square}(f) = \mathrm{disc}^{\square}(f)$.*

Fact 96 ([93, Proposition 2.18]). *Let $f : \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ be a total function. Then, $\mathrm{wprt}_\epsilon^{\square}(f) \geqslant \frac{1-\epsilon\cdot\frac{|\mathcal{Z}|}{|\mathcal{Z}|-1}}{\mathrm{wreg}^{\square}(f)}$.*

Definition 23 (Two-way product subdistribution bounds [110, Definition 3.3]). *For a distribution$\mu\mu$ over $\mathcal{X}\times\mathcal{Y}$, define $\mathrm{sub}_\epsilon^{\mu}(f) := \min_\lambda S^\infty(\lambda\|\mu)$, where λ is taken over all distributions which are both SM-like (simultaneous-message-like)[9] for μ and ϵ-monochromatic for f. When the minimization is restricted to distributions that are SM-like and (ϵ,z)-monochromatic for some fixed value $z\in\mathcal{Z}$, we denote the resulting measure as $\mathrm{sub}_\epsilon^{\mu,z}(f)$. The two-way product subdistribution bound, denoted as $\mathrm{sub}_\epsilon^{\square}(f) := \max_\mu \mathrm{sub}_\epsilon^{\mu}(f)$, where μ ranges over all product distributions on $\mathcal{X}\times\mathcal{Y}$.*

[9] We say that λ is SM-like for μ if there exists a distribution χ on $\mathcal{X}\times\mathcal{Y}$ such that χ is one-message-like (recall definition from Footnote 4)for μ with respect to $\mathcal{X}$ and λ is one-message-like for χ with respect to $\mathcal{Y}$.

In the definition above, we say that a distribution λ is (ϵ, z)-monochromatic for f if $\Pr_{X,Y\sim\lambda}[(X,Y,z) \in f] \geqslant 1-\epsilon$. The distribution λ is ϵ-monochromatic for f if it is (ϵ, z)-monochromatic for f for some $z \in \mathcal{Z}$.

Fact 97 ([86, Lemma 3.1]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $\delta \in (0,1)$. Then, $\mathrm{rec}^{[]}_{\epsilon}(f) \geqslant \mathrm{sub}^{[]}_{\epsilon}(f) \geqslant \mathrm{rec}^{[]}_{\epsilon/\delta^2}(f) - \log \frac{1}{(1-\delta)^2}$.*

Fact 98 ([97, Theorem 1.1]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. Then, $IC^{[]}_{1/10}(f) \geqslant \log\left(\frac{1}{\mathrm{disc}^{[]}(f)}\right)$.*

Fact 99 ([49, Theorem 2.4]). *For every f, ϵ, there exists a protocol π that computes f on inputs drawn from a product distribution μ with probability of error at most ϵ and communication at most $D^{\mu^k}_{\rho}(f^k)$ such that $IC^{[]}(\pi) \leqslant \frac{2D^{[]^k}_{\epsilon}(f^k)}{k}$.*

Fact 100 ([85, Theorem 1.1]). *There exists a positive constant C such that for all functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. all $\epsilon, \delta \in (0, \frac{1}{2}]$, we have $IC^{[]}_{\epsilon}(f) \geqslant \frac{\delta^2}{C} \cdot \left(\log \bar{\mathrm{prt}}^{[]}_{\epsilon+3\delta}(f) - \log 2\right) - \delta$.*

Fact 101 ([100, Theorem 6.3]). *For any function f, distribution μ and error ϵ, $AC^{[]}_{\epsilon}(f) = IC^{[]}_{\epsilon}(f)$.*

Fact 102 ([102, Theorem 2]). *For any function f, distribution μ and error $\epsilon > 0$, we have $AQCC^{[]}_{\epsilon}(f) = QIC^{[]}_{\epsilon}(f)$.*

Fact 103 ([105]). *For any function f and error $\epsilon > 0$, we have $AQCC^{[]}_{\epsilon}(f) \leqslant AC^{[]}_{\epsilon}(f)$.*

Fact 104 ([93, Theorem 2.37]). *Let $f : \mathcal{X} \times \mathcal{Y} \to 2^{\mathcal{Z}}$ be any relation and let $\epsilon \in [0, 1/2)$. Then, $pIC^{\infty,[]}_{\epsilon}(f) = \log \mathrm{prt}^{[]}_{\epsilon}(f)$.*

The fact that quantum communication complexity is cheaper than randomized communication complexity extends to the one-way communication model under product distributions. Moreover, by definition of distributional communication complexity, the deterministic version of Fact 33 (see also [37, Exercise 4.2.1]) naturally extends to distributional communication complexity under product distribution. In addition, the one-way rectangle bound over product distributions serves as an upper bound on $R^{A\to B,[]}_{2\epsilon}(f)$ and a lower bound on $Q^{A\to B,[]}_{\epsilon}(f)$ under certain conditions.

Fact 105. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Then, $Q^{A\to B,[]}_{\epsilon}(f) \leqslant R^{A\to B,[]}_{\epsilon}(f)$.*

Proof. In order to simulate a private coin randomized protocol with a quantum protocol, Alice and Bob each prepare the state $\sum_i \sqrt{p(i)}\,|i\rangle$ and measure it locally. The rest of the protocol proceeds by replacing classical transformations with their quantum counterparts. □

Fact 106. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. We have $D^{A\to B,[]}_{\epsilon}(f) \geqslant \log D^{[]}_{\epsilon}(f)$.*

Fact 107 ([84, Theorem 4]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a total function and let $\epsilon \in (0, 1/2)$. Also, let $\mathrm{rec}^{A\to B,[]}_{\epsilon}(f) > 2\log(1/\epsilon)$. Then, $Q^{A\to B,[]}_{\epsilon^{3/8}}(f) \geqslant \Omega(\mathrm{rec}^{A\to B,[]}_{\epsilon}(f))$.*

Fact 108 ([84, 86]). *Let $\epsilon \in (0, 1/4)$, $R^{A\to B,[]}_{2\epsilon}(f) = O(\mathrm{rec}^{A\to B,[]}_{\epsilon}(f))$.*

Now, we introduce more complexity measure related to learning theory. In learning theory, a *concept class* $\mathtt{C}$ is any set of functions $f : \mathcal{X} \times \mathcal{Y} \to \{-1, 1\}$ for a finte set $\mathcal{X} \times \mathcal{Y}$. Ths *statistical query (SQ) dimension* under a given distribution μ, denoted as $\mathrm{sqdim}_{\mu}(\mathtt{C})$, is defined to be the largest $d \in \mathbb{Z}_+$ for which there are functions $f_1, f_2, \cdots, f_d \in \mathtt{C}$ and a probability

distribution μ on $\mathcal{X} \times \mathcal{Y}$ such that $\left|\mathbb{E}_{(x,y)\sim\mu}\left[f_i(x,y)f_j(x,y)\right]\right| \leqslant \frac{1}{d}$, for all $i \neq j$. We denote $\mathrm{sqdim}(\mathtt{C}) \coloneqq \max_\mu \mathrm{sqdim}_\mu(\mathtt{C})$ [111].

We identify $\mathtt{C}$ with the sign matrix M, whose rows are indexed by the functions of $\mathcal{C}$ and columns indexed by the inputs $(x,y) \in \mathcal{X} \times \mathcal{Y}$, and entries given by $M_{f,(x,y)} = f(x,y)$. For a matrix $M \in \{-1,1\}^{n\times n}$, we define its SQ dimension, $\mathrm{sqdim}(M)$, to be the SQ dimension of the rows of M viewed as functions $\{1,2,\cdots,n\} \to \{-1,1\}$. The *statistical query complexity* $\mathrm{sq}(M)$ of M as the SQ dimension of its rows, i.e. $\mathrm{sq}(M) \coloneqq \mathrm{sqdim}(\{f_1,...,f_n\})$, where $f_1,\cdots,f_n$ are the rows of M.

Sherstov [111] showed that statical query complexity characterizes the discrepancy.

Fact 109 ([111, Theorem 7.1]). *Let $M \in \{-1,1\}^{m\times n}$. Then, $\sqrt{\frac{1}{2}\,\mathrm{sq}(M)} < \frac{1}{\mathrm{disc}^{[]}(M)} < 8\,\mathrm{sq}(M)^2$.*

A combinatorial quantity that captures the learning complexity of a concept class is the *Vapnik-Chervonenkis* (VC) dimension. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function whose communication complexity we are interested in. For any $x \in X$, let $f_x : \mathcal{Y} \to \mathcal{Z}$ be such that for every $y \in \mathcal{Y}$, $f_x(y) \coloneqq f(x,y)$. Define the set $f_\mathcal{X} = \{f_x | x \in \mathcal{X}\}$. Similarly, for $y \in \mathcal{Y}$, let $f_y(x) \coloneqq f(x,y)$ for all $x \in \mathcal{X}$ and $f_Y = \{f_y | y \in \mathcal{Y}\}$. In other words, the function class $f_\mathcal{X}$ (resp. $f_\mathcal{Y}$), is defined by the set of rows (resp. columns) of the communication matrix associated with f. Conversely, given a concept/function class $f_\mathcal{X}$, an element $f_x \in f_\mathcal{X}$ is a feature vector of dimension $|\mathcal{Y}|$. This corresponds to the geometric representation of boolean matrices as points and half spaces in the context of learning theory. The function $f(x,y)$ evaluates to 1 if and only if f_x belongs to the half-space $\{h : \langle h, y\rangle > 0\}$ and evaluates to 0 otherwise. This extends to the case of sign vectors and matrices [76]. The VC dimension of a matrix $M \in \{-1,1\}^{n\times n}$, denoted as $VC(M)$, is the largest $d \in \mathbb{Z}_+$ such that M features a $2^d \times d$ submatrix whose rows are the distinct elements of $\{-1,1\}^d$. Results from [82] and [112] show that statistical query complexity can be upper and lower bounded by VC dimension.

Fact 110 ([82, Lemma 4.1]). *Let $\mathtt{C}$ be a concept class. Then, $\mathrm{sq}(\mathtt{C}) \leqslant 2^{O(VC(\mathtt{C}))}$.*

Fact 111 ([112, Observation 20]). *For a concept class $\mathcal{C}$, $\mathrm{sq}(\mathtt{C}) = \Omega(VC(\mathtt{C}))$.*

Yao's principle [113] holds for one-way communication under product distribution. However, Sherstov [82] showed that the max relationship between $R_\epsilon^{[]}(f)$ and $D_\epsilon^{[]}(f)$ does not hold for the two-way model. In particular, they give a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ for which $\max_{[]} D_\epsilon^{[]}(f) = \Theta(1)$ but $R_\epsilon^{pub}(f) = \Theta(n)$.

Fact 112 ([4],[114, Theorem 2.2]). *For every function $f : \mathcal{X}\times\mathcal{Y} \to \{0,1\}$ and for every $0 < \epsilon < 1$, we have $R_\epsilon^{A\to B,[]}(f) = \max_{[]} D_\epsilon^{A\to B,[]}(f)$.*

The two results show the equivalence between $VC(f)$, $R_\epsilon^{A\to B,[]}(f)$, $\mathrm{sub}_{\mathcal{Y},\epsilon}^{A\to B,[]}(f)$ and $\mathrm{rec}_\epsilon^{A\to B,[]}(f)$.

Fact 113 ([114, Theorem 3.2]). *For every function $f : \mathcal{X}\times\mathcal{Y} \to \{0,1\}$ and for every $\epsilon \leqslant 1/8$,$R_{1/3}^{A\to B,[]}(f) = \Theta(VC(f))$.*

Fact 114 ([86, Theorem 4.5]). *Let $f \subseteq \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ be a relation and let $0 \leqslant \epsilon \leqslant 1/6$. There are universal constants c_1, c_2 such that $\mathrm{sub}_{\mathcal{Y},\epsilon}^{A\to B,[]}(f), -1 \leqslant c_1 \cdot R_\epsilon^{A\to B,[]}(f) \leqslant c_2 \cdot \left[\mathrm{sub}_{\mathcal{Y},\epsilon}^{A\to B,[]}(f) + \log\frac{1}{\epsilon} + 2\right]$.*

Moreover, the distributional complexity under product distributions is upper-bounded by VC dimension in both the one-way and two-way communication model.

Fact 115 ([114, Theorem 3.2], [82, Theorem 3.1]). *Let M be a sign matrix and let $0 < \epsilon \leqslant 1/3$. Then, $D_\epsilon^{A\to B,[]}(M) \leqslant O\left(\frac{1}{\epsilon} VC(M) \log\frac{1}{\epsilon}\right)$.*

The result below show that one-way quantum communication complexity under the product distribution can be lower bounded by the statistical query complexity and the rectangle bound.

Fact 116 ([115, Fact 2.8]). *Let $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$. For every $\gamma > 0$, we have $\Omega(\log(\gamma \cdot \sqrt{\mathrm{sq}(\mathtt{C})})) \leqslant Q^{A\to B,[]}_{1/2+\gamma}(\mathtt{C})$.*

Last but not least, we state results related to infromation complexity.

Fact 117 ([116, Proposition 3.12]). *For any protocol π, $IC^{[]}(\pi) = IC^{\mathrm{ext},[]}(\pi)$.*

Fact 118 ([116]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, then we have $IC^{\mathrm{ext},[]}_{\epsilon}(f) = IC^{[]}_{\epsilon}(f)$, $IC^{\mathrm{ext}}_{D,\epsilon}(f) = IC_{D,\epsilon}(f)$, $IC^{\mathrm{ext}}_{\epsilon}(f) = IC_{\epsilon}(f)$.*

4 Separation

In this subsection, we justify the separation between different classes. This is done by either mentioning known results, or finding a function f such that for a real function $g : \mathbb{R} \to \mathbb{R}$, complexity measures $\mathcal{C}$ from Class i and $\mathcal{C}'$ from Class $i+1$, we have $\mathcal{C}(f) \geqslant g(n)$ and $\mathcal{C}'(f) = O(1)$, where $i \in \{1,2,3,4\}$.

Separation between Class 1 and Class 2 Consider the Equality problem defined as $EQ : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ such that for all $(x,y) \in \{0,1\}^n \times \{0,1\}^n$,

$$EQ(x,y) = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{otherwise.} \end{cases}$$

The Equality problem provides a polynomial separation between Class 1 and Class 2. More specifically, $D(EQ) = n+1$ by [37, Example 1.21]. On the other hand, observe that $\gamma_2(EQ) = 1$ since the communication matrix M_{EQ} of EQ is the identity matrix I and $\gamma_2(I)$=1.

Remark 1. *We note that separation in the other direction is not possible as deterministic communication complexity is always at least randomized communication complexity, even in the public coin and exact error case.*

Separation between Class 2 and Class 3 An exponential separation between Class 2 and Class 3 was prove by Cheung et al.. [33, Corollary 6]. However, to the best of our knowledge, an unbounded separation is not known to exist.

Fact 119 (Corollary 6 in[33]). *There is a Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with communication matrix M_f, such that $\gamma_2(M_f) \geqslant 2^{n/43}$ and $\gamma_2^{\alpha}(M_f) \leqslant O(\mathrm{poly}(n))$.*

Define the Hamming Distance problem as $HD_1 : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ such that for all $(x,y) \in \{0,1\}^n \times \{0,1\}^n$,

$$HD_1(x,y) = \begin{cases} 1, & \text{if } d(x,y) = 1 \\ 0, & \text{otherwise,} \end{cases}$$

where $d(x,y)$ denotes the Hamming distance between x and y.

Lemma 6. $R^{\mathrm{pub}}_{\epsilon}(HD_1) = O(1)$.

Proof. Assume that the following Equality Testing protocol costs $O(1)$ and has negligible error $\epsilon > 0$:

$$\mathrm{EQTEST}_{\epsilon}(x,y) = \begin{cases} 1, \text{ with probability } 1-\epsilon, \text{ if } x = y \\ 0, \text{ with probability } 1-\epsilon, \text{ if} x \neq y. \end{cases}$$

The randomized public coin protocol with constant error for the HD_1 protocol goes as follows:

1. Perform EQTEST_ϵ on (x, y).

2. If $\text{EQTEST}_\epsilon(x, y) = 1$, output 0 for HD_1 and terminate the protocol. Otherwise, remove half of the entries of x, y using a public random coin. Denote the shortened strings as x', y' respectively. Perform EQTEST_ϵ on (x', y') and accept as per the Equality Test.

3. Repeat Step 2 for 100 times and accept if EQTEST outputs 0 for at least 33 times.

Now, we analyze the correctness of the protocol. There are two cases:
There are two cases:

(i) If $d(x, y) = 0$, then EQTEST_ϵ returns 1 with probability at least $1 - \epsilon$. The protocol is terminated and we get $HD_1(x, y) = 0$ with probability at least $1 - \epsilon$.

(ii) If $d(x, y) \neq 0$,,

(a) If $d(x, y) = 1$, then $\text{EQTEST}(x, y) = 1$ with probability ϵ. Moreover, $d(x', y')$ is either 0 or 1 with probability 1/2 each. In the former case, $\text{EQTEST}_\epsilon(x', y') = 1$ with probability $1 - \epsilon$, while in the latter case, with probability ϵ. So the total probability that EQTEST outputs 1 on x, y is at least

$$\epsilon + \frac{1}{2} \cdot (1 - \epsilon) + \frac{1}{2} \cdot \epsilon = \frac{1}{2} + \epsilon.$$

(b) If $d(x, y) = 2$, $\text{EQTEST}(x, y) = 1$ with probability ϵ. Furthermore, $d(x', y') = 0$ with probability 1/4 and $1 \leqslant d(x', y') \leqslant 2$ with probability 3/4. In this case, the total probability that EQTEST outputs 1 on x, y is at most

$$\epsilon + \frac{1}{4} \cdot (1 - \epsilon) + \frac{3}{4} \cdot \epsilon \approx \frac{1}{4} + \epsilon.$$

By a similar argument, if $d(x, y) = k$ for some $k > 2$, $\text{EQTEST}(x, y) = 1$ with probability ϵ. Also, $x' = y'$ with probability $1/2^k$ and hence the total probability that EQTEST outputs 1 on x, y is approximately $\frac{1}{2^k} + \epsilon$. Hence, the error of the above protocols depends on the Hamming distance of x, y.

This gives a standard randomized public coin protocol with constant communication.

□

We are now interested in the lower bound on $\gamma_2(HD_1)$. Notice that HD_1 corresponds to a special case of the threshold function defined in the work of Hambardzymyan et al. [22]. Namely, the threshold function is given by $\bar{\text{thr}}_k : \{0,1\}^n \to \{0,1\}$ such that

$$\bar{\text{thr}}_k(x) = 1 \quad \Leftrightarrow \quad \sum_{i=1}^{n} x_i < k.$$

HD_1 corresponds to the case when $k = 2$. The authors showed the following result.

Fact 120 ([22, Lemma 2.15])**.** $\gamma_2(HD_1) = \Theta(\sqrt{n})$.

Nevertheless, we give a simpler proof for their result, avoids the heavy machinery of Fourier analysis. We consider the Boolean variant of the communication matrix and show the following lemma.

Lemma 7. $\gamma_2(HD_1) \geqslant \Omega(\sqrt{n})$.

Proof. Note that we have by an alternative definition of γ_2 that and the above observation that

$$\gamma_2(M_n) = \max_{u,v:\|u\|_2=\|v\|_2=1} \|uv^T \circ M_n\|_{tr}.$$

By picking $u = v = (1/\sqrt{2^n}, \cdots, 1/\sqrt{2^n})^T \in \mathbb{R}^{2^n}$, we get by the structure of M_n, that

$$\gamma_2(M_n) \geqslant \left\| \frac{1}{2^n} J \circ M_n \right\|_{tr} \geqslant \left\| \frac{1}{2^n} J \circ M_n \right\|_F = \sqrt{2^n \cdot n \cdot \frac{1}{2^n}} = \sqrt{n},$$

where $\|\cdot\|_{tr}$ and $\|\cdot\|_F$ denotes the nuclear and Frobenius norm respectively, and J is the all ones matrix. This completes the proof. □

For completeness, we give an upper bound on $\gamma_2(HD_1)$.

Lemma 8. $\gamma_2(HD_1) \leqslant O(\sqrt{n})$.

Proof. The communication matrix of HD_1 for strings length n is a $2^n \times 2^n$ Boolean matrix M_n. When $n = 1$,

$$M_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

For the case of $n = 2$,

$$M_2 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}.$$

Hence, M_n can be described recursively as follows:

$$M_n = \left[\begin{array}{c|c} M_{n-1} & I \\ \hline I & M_{n-1} \end{array}\right]$$

By its trivial decomposition, we can write $M_n = IM_n$. Since every row/column of M_n contains n ones, we have

$$\gamma_2(M_n) = \gamma_2(IM_n) \leqslant \|I\|_{2\to\infty}\|M_n\|_{1\to 2} = 1 \cdot \sqrt{n},$$

where $\|M_n\|_{p\to q} := \sup_{v\neq 0} \frac{\|M_n v\|_q}{\|v\|_p}$. □

Combining Lemmas 7 and 8 give $\gamma_2(HD_1) = \Theta(\sqrt{n})$.

Remark 2. *Separation in the other direction is not possible since $\gamma_2(f) \geqslant \gamma_2^\alpha(f)$ for all f and $\alpha > 1$ by Fact 35.*

Separation between Class 3 and Class 4 Let $C \approx 2^{n/3}$. Alice gets the vector $x \in [-C, C]^3$ and Bob gets the vector $y \in [-C, C]^3$. Define

$$f(x, y) = \text{sign}\langle x, y\rangle,$$

where sign : $\mathbb{R} \to \{-1, 1\}$ is the sign function, mapping positive inputs to 1 and negative inputs to -1; and $\langle\cdot,\cdot\rangle$, denotes the inner product between two vectors. Hatami Hosseini and Lovett [29] showed that the above function provides separation between Class 3 and Class 4.

Fact 121 ([29, Theorem 4])**.** *The function f defined above satisfies* $\text{signrank}(f) = 3$ *and* $\text{disc}(f) = 2^{-\Omega(n)}$*. In particular,* $R_\epsilon^{\text{pub}}(f) = \Omega(n)$.

Remark 3. *Separation in the opposite direction is not possible since the unbounded-error communication complexity of a function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ is never much more than its complexity in the other standard models. For instance, $UPP(f) \leqslant O(R_{1/3}^{\text{pub}}(f) + \log\log(|\mathcal{X}| + |\mathcal{Y}|))$ and $UPP(f) \leqslant O(Q_{1/3}^*(f) + \log\log(|\mathcal{X}| + |\mathcal{Y}|))$ [117, Section 2.1].*

Separation between Class 4 and Class 5 We first define the notion of a maximum class. Let $C \subseteq \{-1,1\}^N$ be a class with VC dimension d. Then, C is called a *maximum class* if it meets the Sauer-Shelah's bound [118] with equality, i.e. $|C| = \sum_{i=0}^{d} \binom{N}{i}$.

Alon et al. [108] showed a polynomial separation between $\text{signrank}(f)$ and $VC(f)$ [108] via a maximum class of low VC dimension and high signrank. Specifically, let P be the set of points in a projective plane of order n and let L be the set of lines in it. Let $N = N_{n,2} = |P| = |L|$, where $N_{n,2} := n^2 + n + 1 = \frac{n^3-1}{n-1}$. For every $\ell \in L$, fix some linear order on the points in ℓ. A set $T \subset P$ is called an *interval* if $T \subseteq \ell$ for some line $\ell \in L$, and T forms an interval with respect to the order fixed on ℓ.

Fact 122 (Theorem 8 in [108])**.** *The class Q of all intervals is a maximum class of VC dimension 2. Moreover, there exists a choice of linear orders for the lines in L such that the resulting Q has signrank $\Omega(n^{1/2}/\log n)$.*

Remark 4. *Separation between Class 3 and 4 in the opposite direction is not possible since* $\text{signrank}(f) \geqslant VC(f)$ *[108].*

5 Conclusion and open problems

We outline a hierarchy for constant communication complexity. by categorizing communication complexity measures into five classes. In each class, one complexity measure is related by a function of another complexity measure, independent of the input size. We show separation between consecutive classes. Such separations are only meaningful in one direction.

Our classification is mainly based on known results in the literature. Although we have made every attempt to comprehensively include all relevant complexity measures, some measures may have been inadvertently omitted. There is also a small number of measures that we are unable to classify. For example, it is natural for one to include $QMA(f)$ in Class 3 since both $MA(f)$ and $Q^*_\epsilon(f)$ are also in that class. In particular, having $MA(f) \leqslant R^{\text{pub}}_\epsilon(f)$, $QMA(f) \leqslant Q^*_\epsilon(f)$ should hold analogously. In his paper, Klauck proved that $QMA(f)$ is lower bounded by the one-sided smooth discrepancy, $\text{sdisc}^1_\delta(f)$ [73, Theorem 2]. To the best of our knowledge, no lower bound on $\text{sdisc}^1_\delta(f)$ is known. If one could prove a lower bound on $\text{sdisc}^1_\delta(f)$ in terms of any complexity measure in Class 3 and independent of the input size, then both $QMA(f)$ and $\text{sdisc}^1_\delta(f)$ can be added to Class 3.

Open Problem 1. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. Which of the complexity measures in Class 3 lower bounds* $\text{sdisc}^1_\delta(f)$*, without dependency on the input size?*

Conjecture 1. *$QMA(f)$ and* $\text{sdisc}^1_\delta(f)$ *belong to Class 3.*

On the other hand, it is surprising to see that AM and MA complexities belong to different classes despite both models being almost similar. This is due to the fact that $AM_\epsilon(f) \leqslant O(SBP(f) + \log n)$ [89].

Other unclassified measures include the (two-way) Las Vegas communication complexity, $LV(f)$. Jain and Klauck introduced the Las Vegas partition bound and proved that it is at most exponential in $LV(f)$ [83, Lemma 7]. However, it is not clear if $LV(f)$ is lower bounded by $R^{\text{pub}}_\epsilon(f)$ in Class 3 or upper bounded by $D(f)$ in Class 1.

With regards to information complexity, Braverman showed that $IC_0(f) = IC_{D,0}(f)$ [50, Theorem 3.6] and $IC^{\text{ext}}_0(f) = IC^{\text{ext}}_{D,0}(f)$ [50, Theorem, 3.17]. Since internal information complexity is upper bounded by external information complexity, which is always upper bounded by communication complexity, this implies that $IC_0(f)$ and $IC_{D,0}(f)$ are respectively upper bounded by $IC^{\text{ext}}_0(f)$ and $IC^{\text{ext}}_{D,0}(f)$, which are both upper bounded by $R^{\text{pub}}_0(f)$. As there are limited

results on exact information complexity, it is not clear whether $IC_0(f)$ can be lower bounded by any of the elements in Class 1, without dependency on the input size.

Open Problem 2. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. Which of the complexity measures in Class 1 lower bounds $IC_0(f)$, without dependency on the input size?*

Conjecture 2. *$IC_0(f), IC_{D,0}(f), IC_0^{\text{ext}}(f)$ and $IC_{D,0}^{\text{ext}}(f)$ belong to Class 1.*

Furthermore, the equivalence between the internal zero-error information complexity of a function f and its amortized communication complexity, does not hold in the case of amortized zero-error communication complexity. In fact, it was conjectured that the (average case) amortized zero-error communication complexity of a function is exactly characterized by its external information complexity [119, 120]. If this conjecture is proven to be true, then the average case amortized zero-error communication complexity belongs to the same class as the complexity measures mentioned in Conjecture 2.

Touchette showed that $AQCC_\epsilon^\mu(f) = QIC_\epsilon^\mu(f)$ for any function f and error $\epsilon > 0$ [102, Theorem 2]. While one can draw the relation $Q_\epsilon^{*,\mu}(f) \to QIC_\epsilon^\mu(f) \to AQCC_\epsilon^\mu(f) \to AC_\epsilon^\mu(f) \to IC_\epsilon^\mu(f)$ in Class 3, the lower bound on $Q_\epsilon^{*,\mu}(f)$ is not known. In fact, this was posed as an open problem in the work of Jain and Zhang [84].

Open Problem 3. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. Which of the complexity measures in Class 3 lower bounds $Q_\epsilon^{*,\mu}(f)$, without dependency on the input size?*

Conjecture 3. *$Q_\epsilon^{*,\mu}(f), QIC_\epsilon^\mu(f)$ and $AQCC_\epsilon^\mu(f)$ belong to Class 3.*

Under product distribution over inputs, we have a similar open problem and conjecture.

Conjecture 4. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a Boolean function. Which of the complexity measures in Class 5 lower bounds $Q_\epsilon^{*,[]}(f)$, without dependency on the input size?*

Conjecture 5. *$Q_\epsilon^{*,[]}(f), QIC_\epsilon^{[]}(f)$ and $AQCC_\epsilon^{[]}(f)$ belong to Class 5.*

6 Acknowledgments

DL specially thanks Alexander Belov, Srijita Kundu, João F. Doriguello, Dmitry Gavinsky and Uma Girish for helpful discussions. AA and DL acknowledge funding from the Latvian Quantum Initiative under EU Recovery and Resilience Facility under project no. 2.3.1.1.i.0/1/22/I/CFLA/001. DL and HK thanks funding from the Singapore Ministry of Education (partly through the Academic Research Fund Tier 3 MOE2012-T3-1-009) and by the Singapore National Research Foundation. DL and HK is also supported by Majulab UMI 3654.

References

[1] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213, 1979.

[2] László Babai, Noam Nisant, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, 1992.

[3] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science (FOCS 1986)*, pages 337–347. IEEE, 1986.

[4] Andrew Chi-Chih Yao. Lower bounds by probabilistic arguments. In *24th Annual Symposium on Foundations of Computer Science (FOCS 1983)*, pages 420–428. IEEE, 1983.

[5] Alon Orlitsky and Aly El Gamal. Communication complexity. In *Complexity in information theory*, pages 16–61. Springer, 1988.

[6] László Lovász. Communication complexity: A survey. In *Paths, Flows, and VLSI Layout*, page 235–265. Springer-Verlag, 1990.

[7] Thomas Lengauer. VLSI theory. In *Algorithms and Complexity*, pages 835–868. Elsevier, 1990.

[8] Magnús M Halldórsson, Xiaoming Sun, Mario Szegedy, and Chengu Wang. Streaming and communication complexity of clique approximation. In *International Colloquium on Automata, Languages, and Programming*, pages 449–460. Springer, 2012.

[9] László Babai, Anna Gál, and Avi Wigderson. Superpolynomial lower bounds for monotone span programs. *Combinatorica*, 19(3):301–319, 1999.

[10] Farid Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoretical Computer Science*, 157(2):139–159, 1996.

[11] Noam Nisan. The communication complexity of threshold gates. *Combinatorics, Paul Erdos is Eighty*, 1(301-315):6, 1993.

[12] Hans Dietmar Gröger and György Turán. On linear decision trees computing boolean functions. In *International Colloquium on Automata, Languages, and Programming*, pages 707–718. Springer, 1991.

[13] Ashok K Chandra, Merrick L Furst, and Richard J Lipton. Multi-party protocols. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 94–99, 1983.

[14] Noga Alon and Wolfgang Maasst. Meanders and their applications in lower bounds arguments. *Journal of Computer and System Sciences*, 37(2):118–129, 1988.

[15] László Babai, Pavel Pudlák, Vojtech Rödl, and Endre Szemerédi. Lower bounds to the complexity of symmetric boolean functions. *Theoretical Computer Science*, 74(3):313–323, 1990.

[16] Vwani P Roychowdhury, Alon Orlitsky, and Kai-Yeung Siu. Lower bounds on threshold and related circuits via communication complexity. *IEEE Transactions on Information Theory*, 40(2):467–474, 1994.

[17] Mikael Goldmann. Communication complexity and lower bounds for threshold circuits. In *Theoretical advances in neural computation and learning*, pages 85–125. Springer, 1994.

[18] Kai-Yeung Siu, Vwani Roychowdhury, and Thomas Kailath. *Discrete neural computation: A theoretical foundation*. Prentice-Hall, Inc., 1995.

[19] Alexander A. Razborov. Applications of matrix methods to the theory of lower bounds in computational complexity. *Combinatorica*, 10(1):81–93, 1990.

[20] Peter Bro Miltersen. Lower bounds for union-split-find related problems on random access machines. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 625–634, 1994.

[21] Noam Chomsky. Three models for the description of language. *IRE Transactions on information theory*, 2(3):113–124, 1956.

[22] Lianna Hambardzumyan, Hamed Hatami, and Pooya Hatami. Dimension-free bounds and structural results in communication complexity. *Israel Journal of Mathematics*, 253(2):555–616, 2023.

[23] Nathaniel Harms, Sebastian Wild, and Viktor Zamaraev. Randomized communication and implicit graph representations. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1220–1233, 2022.

[24] Nathaniel Harms and Viktor Zamaraev. Randomized communication and implicit representations for matrices and graphs of small sign-rank. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1810–1833. SIAM, 2024.

[25] Yuting Fang, Lianna Hambardzumyan, Nathaniel Harms, and Pooya Hatami. No complete problem for constant-cost randomized communication. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1287–1298, 2024.

[26] Nati Linial, Shahar Mendelson, Gideon Schechtman, and Adi Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27:439–463, 2007.

[27] Nati Linial and Adi Shraibman. Learning complexity vs communication complexity. *Combinatorics, Probability and Computing*, 18(1-2):227–245, 2009.

[28] Abhishek Bhrushundi, Sourav Chakraborty, and Raghav Kulkarni. Property testing bounds for linear and quadratic functions via parity decision trees. In *Computer Science-Theory and Applications: 9th International Computer Science Symposium in Russia, CSR 2014, Moscow, Russia, June 7-11, 2014. Proceedings 9*, pages 97–110. Springer, 2014.

[29] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Sign rank vs discrepancy. In *35th Computational Complexity Conference (CCC 2020)*, volume 169 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.

[30] Nathaniel Harms. Universal communication, universal graphs, and graph labeling. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:27. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.

[31] Hamed Hatami, Pooya Hatami, William Pires, Ran Tao, and Rosie Zhao. Lower Bound Methods for Sign-Rank and Their Limitations. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*, volume 245 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

[32] Louis Esperet, Nathaniel Harms, and Andrey Kupavskii. Sketching Distances in Monotone Graph Classes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*, volume 245 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

[33] Tsun-Ming Cheung, Hamed Hatami, Kaave Hosseini, and Morgan Shirley. Separation of the factorization norm and randomized communication complexity. In *38th Computational Complexity Conference (CCC 2023)*, volume 264 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023.

[34] Manasseh Ahmed, Tsun-Ming Cheung, Hamed Hatami, and Kusha Sareen. Communication complexity of half-plane membership. In *Electron. Colloquium Comput. Complex*, 2023.

[35] Hamed Hatami and Pooya Hatami. Structure in communication complexity and constant-cost complexity classes. *SIGACT News*, 55(1):67–93, 2024.

[36] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *Computational Complexity*, 27(2):245–304, 2018.

[37] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996.

[38] Anna Gál and Ridwan Syed. Upper Bounds on Communication in Terms of Approximate Rank. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12730 LNCS(6):116–130, 2021.

[39] Matthias Krause. Geometric arguments yield better bounds for threshold circuits and distributed computing. *Theoretical Computer Science*, 156(1-2):99–117, 1996.

[40] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 699–708, 2007.

[41] Jiří Matoušek, Aleksandar Nikolov, and Kunal Talwar. Factorization norms and hereditary discrepancy. *International Mathematics Research Notices*, 2020(3):751–780, 2020.

[42] Shanmugavelayutham Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1285–1292, 2012.

[43] Alexander Edmonds, Aleksandar Nikolov, and Jonathan Ullman. The power of factorization mechanisms in local and central differential privacy. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 425–438, 2020.

[44] Monika Henzinger and Jalaj Upadhyay. Constant matters: Fine-grained complexity of differentially private continual observation using completely bounded norms. *Cryptology ePrint Archive*, 2022.

[45] Ashwin V Nayak. *Lower bounds for quantum computation and communication*. PhD thesis, University of California, Berkeley, 1999.

[46] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 124–133, 2001.

[47] Alon Orlitsky. Worst-case interactive communication. ii. two messages are not optimal. *IEEE Transactions on Information Theory*, 37(4):995–1005, 2002.

[48] Nan Ma and Prakash Ishwar. Some results on distributed source coding for interactive function computation. *IEEE Transactions on Information Theory*, 57(9):6180–6195, 2011.

[49] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC '10: Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 67–76, New York, USA, 2010. ACM.

[50] Mark Braverman. Interactive information complexity. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 505–524, 2012.

[51] Lovett Shachar. Communication is bounded by root of rank. *Journal of the ACM*, 63(1):1–9, 2016.

[52] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. *Proceedings of the Annual IEEE Conference on Computational Complexity*, pages 120–130, 2001.

[53] Charles H Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Physical review letters*, 70(13):1895, 1993.

[54] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 63–68, 1998.

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

[56] Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 352–361. IEEE, 1993.

[57] Ilan Kremer. *Quantum communication*. PhD thesis, Hebrew University of Jerusalem, 1995.

[58] Hartmut Klauck. One-way communication complexity and the nečiporuk lower bound on formula size. *SIAM Journal on Computing*, 37(2):552–583, 2007.

[59] Hartmut Klauck. On quantum and probabilistic communication: Las vegas and one-way protocols. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 644–651, 2000.

[60] Nikhil S Mande, Swagato Sanyal, and Suhail Sherif. One-way communication complexity and non-adaptive decision trees. In *39th International Symposium on Theoretical Aspects of Computer Science (STACS 2022)*, volume 219 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 49:1–49:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

[61] Ben Davis, Hamed Hatami, William Pires, Ran Tao, and Hamza Usmani. On public-coin zero-error randomized communication complexity. *Information Processing Letters*, 178:106293, 2022.

[62] Juraj Hromkovič and Georg Schnitger. On the power of las vegas for one-way communication complexity, obdds, and finite automata. *Information and Computation*, 169(2):284–296, 2001.

[63] Louis Deaett and Venkatesh Srinivasan. Linear algebraic methods in communication complexity. *Linear Algebra and Its Applications*, 436(12):4459–4472, 2012.

[64] LeRoy B Beasley and Thomas J Laffey. Real rank versus nonnegative rank. *Linear Algebra and its applications*, 431(12):2330–2335, 2009.

[65] Gillat Kol, Shay Moran, Amir Shpilka, and Amir Yehudayoff. Approximate nonnegative rank is equivalent to the smooth rectangle bound. In *Automata, Languages, and Programming: 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I 41*, pages 701–712. Springer, 2014.

[66] Troy Lee, Adi Shraibman, et al. Lower bounds in communication complexity. *Foundations and Trends® in Theoretical Computer Science*, 3(4):263–399, 2009.

[67] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 223–228, 1988.

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

[69] Christos H Papadimitriou and Stathis K Zachos. Two remarks on the power of counting. In *Theoretical Computer Science: 6th Gl-Conference Dortmund, January 5–7, 1983*, pages 269–275. Springer, 1982.

[70] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *Journal of Computer and System Sciences*, 65(4):612–625, 2002.

[71] Mika Göös, Toniann Pitassi, and Thomas Watson. Zero-information protocols and unambiguity in arthur-merlin communication. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 113–122, 2015.

[72] Hartmut Klauck. Rectangle size bounds and threshold covers in communication complexity. In *18th IEEE Annual Conference on Computational Complexity, 2003. Proceedings.*, pages 118–134. IEEE, 2003.

[73] Hartmut Klauck. On arthur merlin games in communication complexity. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 189–199. IEEE, 2011.

[74] Satyanarayana V Lokam. Spectral methods for matrix rigidity with applications to size–depth trade-offs and communication complexity. *Journal of Computer and System Sciences*, 63(3):449–473, 2001.

[75] Amit Chakrabarti, Graham Cormode, Andrew McGregor, Justin Thaler, and Suresh Venkatasubramanian. Verifiable stream computation and arthur-merlin communication. *Leibniz international proceedings in informatics (LIPIcs)*, pages 217–243, 2015.

[76] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8:21–49, 1999.

[77] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–399, 2007.

[78] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[79] Anurag Anshu, Shalev Ben-David, Ankit Garg, Rahul Jain, Robin Kothari, and Troy Lee. Separating quantum communication and approximate rank. *Leibniz International Proceedings in Informatics, LIPIcs*, 79, 2017.

[80] Nikhil Shekhar Mande. *Communication complexity of xor functions*. PhD thesis, PhD thesis, Tata Institute of Fundamental Research Mumbai, 2018.

[81] Hartmut Klauck. Lower bounds for quantum communication complexity. *SIAM Journal on Computing*, 37(1):20–46, 2007.

[82] Alexander A. Sherstov. Communication complexity under product and nonproduct distributions. *Computational Complexity*, 19(1):135–150, 2010.

[83] Rahul Jain and Hartmut Klauck. The Partition Bound for Classical Communication Complexity and Query Complexity. In *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity*, CCC '10, pages 247–258, 2010.

[84] Rahul Jain and Shengyu Zhang. New bounds on classical and quantum one-way communication complexity. *Theoretical Computer Science*, 410(26):2463–2477, 2009.

[85] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *SIAM Journal on Computing*, 44(5):1550–1572, 2015.

[86] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct Product Theorems for Classical Communication Complexity via Subdistribution Bounds. *Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 599–608, 2008.

[87] Mark Braverman, Ankit Garg, Young Kun Ko, Jieming Mao, and Dave Touchette. Near-Optimal Bounds on Bounded-Round Quantum Communication Complexity of Disjointness. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 773–791, 10 2015.

[88] Rahul Jain. Strong direct product conjecture holds for all relations in public coin randomized one-way communication complexity. *arXiv preprint arXiv:1010.0522*, 2010.

[89] Mika Göös and Thomas Watson. Communication complexity of set-disjointness for all probabilities. *Theory of Computing*, 12(1):1–23, 2016.

[90] Lila Fontes, Rahul Jain, Iordanis Kerenidis, Sophie Laplante, Mathieu Laurière, and Jérémie Roland. Relative discrepancy does not separate information and communication complexity. *ACM Transactions on Computation Theory (TOCT)*, 9(1):1–15, 2016.

[91] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. *Journal of the ACM (JACM)*, 63(5):1–31, 2016.

[92] Rahul Jain, Troy Lee, and Nisheeth K Vishnoi. A quadratically tight partition bound for classical communication complexity and query complexity. *arXiv preprint arXiv:1401.4512*, 2014.

[93] Alexandre Nolin. *Communication complexity: large output functions, partition bounds, and quantum nonlocality*. PhD thesis, Université Paris Cité, 2020.

[94] Arkadev Chattopadhyay, Pavel Dvořák, Michal Kouckỳ, Bruno Loff, and Sagnik Mukhopadhyay. Lower bounds for elimination via weak regularity. In *Leibniz International Proceedings in Informatics, LIPIcs*, volume 66, page 21. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.

[95] Sophie Laplante, Virginie Lerays, and Jérémie Roland. Classical and quantum partition bound and detector inefficiency. In *International Colloquium on Automata, Languages, and Programming*, pages 617–628. Springer, 2012.

[96] Rahul Jain and Srijita Kundu. A direct product theorem for quantum communication complexity with applications to device-independent QKD. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1285–1295, 2022.

[97] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. *Algorithmica*, 76:846–864, 2016.

[98] Ziv Bar-Yossef, Thathachar S Jayram, Ravi Kumar, and D Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[99] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 270–278. IEEE, 2001.

[100] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, 2014.

[101] Manoj M. Prabhakaran and Vinod M. Prabhakaran. Rényi information complexity and an information theoretic characterization of the partition bound. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP*, volume 55 of *LIPIcs*, pages 88:1–88:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

[102] Dave Touchette. Quantum information complexity. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 317–326, 2015.

[103] Igor Devetak and Jon Yard. Exact cost of redistributing multipartite quantum states. *Physical Review Letters*, 100(23):230501, 2008.

[104] Jon T Yard and Igor Devetak. Optimal quantum source coding with quantum side information at the encoder and decoder. *IEEE Transactions on Information Theory*, 55(11):5339–5351, 2009.

[105] Anurag Anshu, Dave Touchette, Penghui Yao, and Nengkun Yu. Exponential separation of quantum communication and classical information. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 277–288, 2017.

[106] Richard Cleve and Harry Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201, 1997.

[107] Mark Braverman, Ankit Garg, Young Kun Ko, Jieming Mao, and Dave Touchette. Near-optimal bounds on the bounded-round quantum communication complexity of disjointness. *SIAM Journal on Computing*, 47(6):2277–2314, 2018.

[108] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus vc dimension. In *Conference on Learning Theory*, pages 47–80. PMLR, 2016.

[109] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, 1984-Octob:118–126, 1984.

[110] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 599–608, 2008.

[111] Alexander A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[112] Lev Reyzin. Statistical queries and statistical algorithms: Foundations and applications. *arXiv preprint arXiv:2004.00557*, 2020.

[113] Andrew Chi Chih Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of the 18th IEEE Symposium on Foundations of Computer Science (FOCS 1977)*, pages 222–227, 1977.

[114] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

[115] Hartmut Klauck. Lower bounds for quantum communication complexity. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 288–297. IEEE, 2001.

[116] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th annual ACM Symposium on Theory of Computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM.

[117] Alexander A Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011.

[118] Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972.

[119] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 151–160, 2013.

[120] Mark Braverman. Coding for interactive computation: progress and challenges. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1914–1921. IEEE, 2012.