Research Statement

Alexander A. Sherstov

May 18, 2022

Can machines provably learn from past experience?

Are quantum computers substantially more powerful than classical ones?

What are the computational limits of a streaming algorithm? A neural network? A computer chip?

At first glance, this looks like a hodgepodge of unrelated questions. Surprisingly, research over the past forty years has shown that these questions are intimately related. They can all be distilled to a single mathematical concept, *communication*, and they can all be fruitfully studied within the framework of a single mathematical theory, *communication complexity theory*. The aim of my research is to understand communication as a pervasive computational phenomenon and use my findings to answer fundamental questions in circuit complexity, computational learning, quantum computing, and other models of computation. I have solved several longstanding open problems in these areas, which I describe below along with my vision for future research.

1 Background

Communication is a common bottleneck. It arises whenever a computational problem requires coordination among several parties, each with an incomplete view of the input. Communication complexity theory, initiated 43 years ago by Turing award winner Andrew Yao [88], is an area of theoretical computer science that studies communication as a computational resource, measured in bits. The simplest setting features two communicating parties, call them Alice and Bob, and a function F with two arguments. Alice has an input x, Bob has an input y, and their objective is to compute $F(x, y)$ by communicating back and forth according to an agreed-upon set of rules, called a *communication protocol*. Communication complexity theory studies the minimum amount of communication required to compute F, a quantity known as the *communication complexity* of F. A straightforward but prohibitively expensive communication protocol is for Alice to send her entire input to Bob. For some functions, this brute force approach is provably optimal, whereas for others one can accomplish the task with surprisingly little communication. From the simple two-party model just described to nondeterministic and quantum communication to sophisticated multiparty systems with overlapping inputs, communication complexity theory studies a wide variety of phenomena and does so in a clean, abstract, and unified way.

The communication bottleneck is often present even when no transmission of data occurs. Indeed, many computational problems involve an implicit flow of information among several components and are therefore subject to the limitations of multiparty communication. Over the years, communication complexity theory has proven to be a powerful tool in virtually all areas of theoretical computer science and beyond, including computational learning, pseudorandom generators, streaming algorithms, quantum computing, data structures, mechanism design, and chip layout.

2 Communication Complexity

My contributions to communication complexity theory include: a powerful and general method for proving lower bounds on communication complexity, the *pattern matrix method*; the determination of the multiparty communication complexity of the *set disjointness problem*, a central problem in the area; the determination of the communication complexity of DNF formulas and constant-depth circuits in the most challenging models (quantum, multiparty, and unbounded-error); a number of results establishing the relative power of different models of communication. My work solves a 22-year-old problem due to Babai, Frankl, and Simon, a 16-year-old problem due to Wigderson, an 11-year-old problem due to Kushilevitz and Nisan, and a 10-year-old problem due to Krause and Pudlák, among others.

In my research, communication is inextricably linked to the study of *multivariate polynomials*, which I consider to be the most fundamental objects in theoretical computer science. By studying communication protocols and polynomials jointly, I have been able to make surprising connections and resolve difficult questions far beyond communication complexity. I now review my contributions in detail.

2.1 The Pattern Matrix Method

Research in communication complexity theory centers around proving that a communication problem $F(x, y)$ of interest admits no efficient protocol. In other words, Alice and Bob cannot compute $F(x, y)$ reliably unless their protocol communicates many bits. Proving such *lower bounds* on communication is difficult—after all, one has to rule out every possible low-cost protocol. In [64, 66], I developed a novel method for proving communication lower bounds, the *pattern matrix method,* based on matrix analysis and linear programming duality. My method yields strong lower bounds on communication in a variety of models, including both classical and quantum communication channels, both high-accuracy communication protocols and protocols with accuracy vanishingly close to random guessing.

In more detail, the *ϵ-approximate degree* of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is denoted $\deg_\epsilon(f)$ and is defined as the minimum degree of a real polynomial p that approximates f within ϵ pointwise: $|f(x) - p(x)| \leqslant \epsilon$ for all $x \in \{0,1\}^n$. Given any function f of interest and any error parameter ϵ, the pattern matrix method considers the communication problem $F\colon \{0,1\}^{4n} \times \{0,1\}^{4n} \to \{0,1\}$ given by $F(x, y) = f(\bigvee_{i=1}^4 x_{1,i} \wedge y_{1,i}, \ldots, \bigvee_{i=1}^4 x_{n,i} \wedge y_{n,i})$ and gives a lower bound on the communication complexity of F in terms of the ϵ-approximate degree of f. The polynomial approximation of Boolean functions is an extensively studied topic, with powerful results available both in the bounded-error regime ($\epsilon = 1/3$) and small-bias regime ($\epsilon = \frac{1}{2} - o(1)$). As a result, the pattern matrix method makes it possible to prove a variety of communication lower bounds in different regimes by leveraging results from polynomial approximation.

As a first application of the pattern matrix method (we will see several others), I solved a problem in circuit complexity due to Krause and Pudlák [46] that had been open for 10 years. The problem concerns the two most studied models in circuit complexity: constant-depth circuits of $\wedge, \vee, \neg$ gates and constant-depth circuits of majority gates. Allender's classic result [2] states that every circuit of the first type can be efficiently converted into a depth-3 circuit of the second type. Krause and Pudlák asked if depth 2 suffices. I settled this question in the negative [64], proving that any conversion to depth 2 will involve an exponential blowup in circuit size. Thus, Allender's result from 1989 is best possible. More generally, I proved an exponentially small upper bound on the *discrepancy* of an $\mathbf{AC}^0$ circuit, an exponential improvement on previous work. The circuit class

$\mathbf{AC}^0$ is defined as the class of all polynomial-size constant-depth circuits of $\wedge, \vee, \neg$ gates. It plays a central role in circuit complexity and will resurface several times in the rest of this document.

I am excited to report that the pattern matrix method has since enabled considerable progress by other researchers and myself on longstanding open problems, e.g., [25, 49, 26, 27, 28, 11, 57, 33, 73, 72, 80]. Among them is the famous set disjointness problem, to which we turn our attention next.

2.2 The Set Disjointness Problem

Consider k parties trying to pick a time to meet. There are a total of n time slots, and each party's availability is a subset of these n time slots. A given party knows his own availability and perhaps the availability of many others—but not all. The goal is to determine with high probability (say, 99%) whether there is a time slot that works for everyone. This is the most studied problem in communication complexity theory, known as the *set disjointness problem*. Abstractly, the set disjointness problem is for k parties to determine whether the given k subsets $S_1, S_2, \ldots, S_k \subseteq \{1, 2, \ldots, n\}$ have nonempty intersection, where no party knows all the k subsets. The standard, and most challenging, version of this problem is known as *number-on-the-forehead* set disjointness, where the i-th party knows all the subsets except S_i.

In a striking 1994 paper, Grolmusz [34] proved that this problem can be solved using roughly $O(n/2^k)$ bits of communication. This was known to be optimal for $k = 2$ parties [37, 55, 9], but starting at $k = 3$ the optimality of Grolmusz's communication protocol remained open for over a decade. This state of affairs changed drastically several weeks after the publication of my pattern matrix paper, with an exciting flurry of papers [49, 26, 11] using my method to prove a lower bound of roughly $\Omega(n/2^{k^3})^{1/(k+1)}$ bits on the communication required. This result is strong for small constant k but deteriorates rapidly as k grows. Yet to match Grolmusz's result, one would need to prove that the communication requirements remain high for as many as $k = \Theta(\log n)$ parties! Once again, progress on the set disjointness problem had come to a stall.

Four years later, I resolved the problem in its entirety [73, 72] by combining the pattern matrix method with a novel use of directional derivatives. Specifically, I proved a lower bound of $\Omega(\sqrt{n}/2^k k)$ on the communication requirements of k-party set disjointness. It had been an open problem for 16 years to prove this lower bound even in the special case $k = 4$; see Wigderson [6]. My result is optimal for quantum communication protocols, and by Grolmusz it is within a square of optimal for classical protocols.

2.3 Unbounded-Error Communication versus the Polynomial Hierarchy

The standard model of randomized communication [47] features parties Alice and Bob and a Boolean function $F\colon X \times Y \to \{0, 1\}$. On input $(x, y) \in X \times Y$, Alice and Bob receive the arguments x and y, respectively. In addition, each of them privately holds an unlimited supply of uniformly random bits which they can use in deciding what message to send at any given point in the protocol. As always, their objective is to compute F on any given input with minimal communication. The ϵ-error randomized communication complexity of F, denoted $R_\epsilon(F)$, is the least cost of a protocol that computes F with probability of error at most ϵ on every input.

Bounded-error communication complexity is the familiar quantity $R_{1/3}(F)$, corresponding to the error parameter $\epsilon = 1/3$. A deeper and more powerful notion is *unbounded-error* communication complexity, defined for a function F as the limit of $R_\epsilon(F)$ as $\epsilon \to 1/2$. The term "unbounded error" is a reference to the fact that the error probability can now be arbitrarily close to $1/2$, e.g., doubly (or triply, or quadruply) exponentially close. Remarkably, Paturi and Simon [54] discovered that unbounded-error communication complexity is synonymous with the matrix-analytic notion of *sign-*

rank. In more detail, the sign-rank of a matrix M is the minimum rank of a real matrix A that agrees with M in sign entrywise: $\operatorname{sgn} M_{i,j} = \operatorname{sgn} A_{i,j}$. The unbounded-error communication complexity of F equals, up to an additive constant, the logarithm of the sign-rank of $[(-1)^{F(x,y)}]_{x,y}$. Unbounded-error communication is the most powerful model of two-party communication for which explicit lower bounds have been proved to date. In particular, unbounded-error communication subsumes all other standard models, including deterministic, nondeterministic, bounded-error randomized, and quantum [57]. Its power makes it a very challenging model for which to prove lower bounds.

In a joint work [57], Razborov and I solved an open problem on unbounded-error communication that had been open for 22 years. This problem, due to Babai, Frankl, and Simon [5], asks whether unbounded-error communication is more powerful than the *polynomial hierarchy* ($\mathbf{PH}^{cc}$), another fundamental model which generalizes the familiar notion of nondeterminism. We gave a strong negative answer to this question by exhibiting a function in the polynomial hierarchy that has high unbounded-error communication complexity. Specifically, we showed that $F(x,y) = \bigwedge_{i=1}^{n^{1/3}} \bigvee_{j=1}^{n^{2/3}} (x_{i,j} \wedge y_{i,j})$ has unbounded-error communication complexity $\Omega(n^{1/3})$. Equivalently, we proved that F, when viewed in matrix form, has exponentially high sign-rank: $\exp(\Omega(n^{1/3}))$. This is the first exponential lower bound on the sign-rank of any function computed by a polynomial-size constant-depth circuit.

As I discuss below, our result explains why no efficient algorithm has been discovered for learning formulas in disjunctive normal form, a famous problem which in 1984 gave rise to computational learning theory and has since remained open.

2.4 Threshold Degree and Sign-Rank of Constant-Depth Circuits

Recall that the all-important class $\mathbf{AC}^0$ in circuit complexity consists of functions $f\colon \{0,1\}^n \to \{0,1\}$ computable by polynomial-size circuit of constant depth with gates $\wedge, \vee, \neg$. As discussed in the previous section, Razborov and I obtained strong lower bounds on the unbounded-error communication complexity and sign-rank of $\mathbf{AC}^0$. Our paper, however, left open the fundamental problem of obtaining *optimal* lower bounds on the unbounded-error communication complexity and sign-rank of this class. In fact, an even more basic problem had long remained open, that of proving an optimal lower bound on the *threshold degree* of $\mathbf{AC}^0$. I resolved both of these questions, in a research effort that spanned 11 years.

Formally, the threshold degree of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is defined to be the minimum degree of a polynomial p that represents f in sign: $\operatorname{sgn} p(x) = (-1)^{f(x)}$ for every $x \in \{0,1\}^n$. It is well-known and straightforward to verify that threshold degree is the limiting case of ϵ-approximate degree, namely, $\deg_{\pm}(f) = \lim_{\epsilon \nearrow 1/2} \deg_{\epsilon}(f)$. For decades, the notion of threshold degree has fueled algorithmic and complexity-theoretic breakthroughs in the study of $\mathbf{AC}^0$, and it has been an important open problem in the area to determine the maximum threshold degree of $\mathbf{AC}^0$. In their seminal 1969 monograph, Minsky and Papert [51] obtained a lower bound of $\Omega(n^{1/3})$ on the threshold degree of a polynomial-size DNF formula. Further progress proved difficult. For a long time, the only improvement was due to O'Donnell and Servedio [53], who obtained a threshold degree lower bound of $\Omega(n^{1/3} \log^{2(d-2)/3} n)$ for circuits of depth d. The authors of [53] formally posed the problem of breaking Minsky and Papert's "1/3" barrier and obtaining a lower bound with a larger exponent. I obtained such an improvement in [74], with a threshold degree lower bound of $\Omega(n^{(d-1)/(2d-1)})$ for circuits of depth d. In [76], I was able to strengthen my lower bound by a polynomial factor, to $\Omega(\sqrt{n})$. Progress then stalled once again, with a quadratic gap between the best lower bound $\Omega(\sqrt{n})$ and the trivial upper bound n.

Three years later, my Ph.D. student Pei Wu and I gave a near-complete solution [80] to the threshold degree problem for $\mathbf{AC}^0$. For any constant $\delta > 0$, we constructed a circuit in $\mathbf{AC}^0$ with threshold degree $\Omega(n^{1-\delta})$, which essentially matches the trivial upper bound of n. Using more advanced techniques, we also obtained in [80] a lower bound of $\Omega(n^{1-\delta})$ on the unbounded-error communication complexity of $\mathbf{AC}^0$ and a lower bound of $\exp(\Omega(n^{1-\delta}))$ on the sign-rank of $\mathbf{AC}^0$, essentially matching the trivial upper bounds of n and 2^n, respectively. Our results are some of the strongest communication lower bounds ever obtained for $\mathbf{AC}^0$.

2.5 Approximate Degree and Communication Complexity of DNF Formulas

Recall that the (bounded-error) approximate degree of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is the minimum degree of a real polynomial p that approximates f pointwise: $|f(x) - p(x)| \leqslant 1/3$. The pattern matrix method [64, 66, 73, 72] transforms approximate degree lower bounds into communication lower bounds in every bounded-error model, including the challenging quantum model and the number-on-the-forehead multiparty model. This connection has driven a major research effort to obtain an optimal lower bound on the approximate degree of $\mathbf{AC}^0$, which itself is a difficult problem. To appreciate this, consider the simple-looking function $f(x) = \bigwedge_{i=1}^{n} \bigvee_{j=1}^{m} x_{i,j}$, known as the AND-OR tree on nm variables. Remarkably, even for this function the approximate degree eluded complexity theorists and quantum researchers for 19 years [52, 82, 36, 3, 69], the conjectured answer being $\Theta(\sqrt{nm})$. I settled this conjecture in [68]. Independently, Bun and Thaler [23] obtained the same result using related techniques.

In a beautiful work four years later, Bun and Thaler [24] proved a lower bound of $\Omega(n^{1-\delta})$ on the approximate degree of $\mathbf{AC}^0$ circuits, where the constant $\delta > 0$ can be taken arbitrarily small at the expense of circuit depth. Bun and Thaler's lower bound shows that $\mathbf{AC}^0$ has essentially the maximum approximate degree—provided that one is willing to look at circuits of *arbitrarily large* constant depth. What happens at small depths has since been a wide open problem, with no techniques to address it. Bun and Thaler's circuit in [24] with approximate degree $\Omega(n^{1-\delta})$ can be "flattened" to produce a DNF formula of size $\exp(\log^{O(\log(1/\delta))} n)$, but this is superpolynomial and thus no longer in $\mathbf{AC}^0$. The only progress here has been an $\Omega(n^{3/4-\delta})$ lower bound, obtained for polynomial-size DNF formulas in [22, 50]. This has left us with a polynomial gap in the approximate degree lower bounds for small depth versus arbitrary constant depth.

In a recent paper [78], I resolved this question using a novel proof that departs completely from Bun and Thaler's approach [23, 24]. For every constant $\delta > 0$, I construct a polynomial-size DNF formula with approximate degree $\Omega(n^{1-\delta})$, which essentially matches the trivial upper bound of n. My result is optimal with respect to circuit depth since $\mathbf{AC}^0$ circuits of depth 1 (i.e., conjunctions and disjunctions) have approximate degree $O(\sqrt{n})$ [52]. Moreover, my constructed DNF formulas are the simplest possible in that they have width $O(1)$. Combined with the pattern matrix method, this result immediately gives polynomial-size DNF formulas with bounded-error quantum communication complexity $\Omega(n^{1-\delta})$ and randomized k-party number-on-the-forehead communication complexity $\Omega(n/4^k k^2)^{1-\delta}$. These are the strongest lower bounds on the bounded-error communication complexity of $\mathbf{AC}^0$, and my work achieves them for the simplest functions in $\mathbf{AC}^0$ (namely, polynomial-size DNF formulas of constant width). In a strong sense, the approximate degree and bounded-error communication complexity of $\mathbf{AC}^0$ can now be considered to be solved problems.

2.6 Product and Nonproduct Distributions

The basis for virtually all lower bounds for randomized computation is Yao's celebrated *minimax principle* [89], which he obtained by recasting randomized computation as a zero-sum game. In the

world of communication complexity, Yao's minimax principle states: to prove that a communication problem $F\colon X \times Y \to \{0,1\}$ is hard for randomized protocols, it is necessary and sufficient to find a probability distribution on $X \times Y$ with respect to which any low-cost deterministic protocol fails to reliably distinguish $F^{-1}(0)$ from $F^{-1}(1)$. Much of the art and magic of communication lower bounds is constructing such probability distributions. By far the simplest distributions to work with are the so-called *product distributions*, whereby Alice and Bob's inputs x and y are distributed independently. In their classic textbook on communication complexity, Kushilevitz and Nisan [47] asked whether the use of product distributions provably leads to (near-)optimal results. This question remained open for 11 years. I resolved it in [65], with the strongest possible answer in the negative. Specifically, I proved the existence of a Boolean function $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ whose communication complexity is the highest possible, $\Omega(n)$, but for which a product distribution can only certify a trivial $\Omega(1)$ lower bound.

In an earlier work [63], I studied *discrepancy*, which I consider to be the single most important complexity measure of a communication problem. I constructed communication problems with an exponential gap between the discrepancy under product versus nonproduct distributions. My work additionally proved that unbounded-error communication protocols can be exponentially more efficient than a related communication model, $\mathbf{PP}^{cc}$—more than 20 years after the introduction of the two models. I received the "Best Student Paper" award for these contributions at the 22nd IEEE Conference on Computational Complexity (CCC 2007). Independently of my work, Buhrman, Vereshchagin, and de Wolf [21] obtained the same separation for $\mathbf{PP}^{cc}$ with much different techniques.

2.7 Robust Approximation by Polynomials

Approximation theory has played a central role in my contributions to communication complexity theory, including the pattern matrix method and the set disjointness problem. Conversely, the insights that I gained along the way allowed me to resolve several questions in polynomial approximation. One such, due to Buhrman et al. [20], is whether every bounded polynomial on the Boolean hypercube can be made robust to noise in the inputs with only a constant-factor increase in degree. Such robustness to noise becomes crucial if one needs to do anything nontrivial with the polynomials, e.g., compose them.

This problem had been unsolved for 9 years when I resolved it in full in [70]. For any $\delta > 0$ and any polynomial $p\colon \{0,1\}^n \to [-1,1]$, I showed how to construct a corresponding polynomial $p_{\text{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree only $O(\deg p + \log \frac{1}{\delta})$ that is robust to noise in the inputs: $|p(x) - p_{\text{robust}}(x+\epsilon)| < \delta$ for all Boolean inputs $x \in \{0,1\}^n$ and their perturbations $\epsilon \in [-1/3, 1/3]^n$. This result is optimal with respect to all parameters and reveals that polynomials possess an extraordinary capacity to tolerate noise. Moreover, the transformation from p to p_{robust} is fully explicit, with a clean formulaic description.

2.8 Interactive Coding for Insertions, Deletions, and Substitutions

Shannon [61, 62] famously considered the problem of transmitting a one-way message across an unreliable channel. In seminal work, Schulman [58, 59, 60] generalized Shannon's problem to the interactive setting. Here, two parties Alice and Bob communicate back and forth according to a communication protocol agreed upon in advance. As in Shannon's case, the communication channel is controlled by an adversary who can change a small constant fraction of the symbols, chosen at will, as they transit through the channel. The goal is to overcome these corruptions by cleverly simulating the original protocol with some redundant communication. Interactive coding is a highly

active research area, with a vast literature on virtually every aspect of the problem. In particular, Braverman and Rao [16] proved that any communication protocol can be simulated in Schulman's model with corruption rate up to $\frac{1}{4} - \epsilon$ for any $\epsilon > 0$, and established a matching impossibility result for corruption rate $\frac{1}{4}$.

Braverman, Gelles, Mao, and Ostrovsky [15] proposed a far-reaching generalization of Schulman's canonical model, whereby the adversary can additionally manipulate the channel by *inserting* and *deleting* symbols. Insertions and deletions are drastically more difficult to handle than substitutions, even in the one-way setting of coding theory. As their main result, Braverman et al. [15] proved that any communication protocol can be simulated in the generalized model with substitutions, insertions, and deletions as along as the corruption rate does not exceed $\frac{1}{18} - \epsilon$, for an arbitrarily small constant $\epsilon > 0$. The authors of [15] posed the problem of determining the highest possible corruption rate that can be tolerated.

My Ph.D. student Pei Wu and I gave a complete solution [81] to this problem. We showed that any protocol can be simulated in the generalized model with substitutions, insertions, and deletions with corruption rate up to $\frac{1}{4} - \epsilon$ for any $\epsilon > 0$. Recall that this corruption tolerance is optimal even in the setting of substitutions alone. Analogous to all previous work [60, 16, 15], our solution uses a constant-size alphabet and incurs only a constant-factor overhead in communication.

2.9 Compressing Interactive Communication

Classic work by Shannon [61, 62] shows how to optimally compress one-way communication to its information content, achieving in the limit a transmission cost equal to the entropy of the message. The corresponding compression problem for *interactive* communication has attracted increasing attention over the past two decades. Consider two parties Alice and Bob with inputs $X \in \mathscr{X}$ and $Y \in \mathscr{Y}$, respectively, where $\mathscr{X}$ and $\mathscr{Y}$ are finite sets and the pair (X, Y) is distributed according to some known probability distribution on $\mathscr{X} \times \mathscr{Y}$. Alice and Bob exchange messages back and forth according to an agreed-upon randomized protocol in order to implement some functionality that depends on both inputs. *Information complexity theory* studies a protocol's *information cost*, defined as the amount of information that Alice and Bob learn on average about each other's inputs from the history of messages exchanged between them.

A protocol's communication cost is always at least as large as its information cost, and the gap between these two quantities can be arbitrary. In this light, it is natural to ask whether the communication in every protocol can be compressed to its information content while approximately preserving the protocol's functionality. For every protocol of information cost I, Braverman [14] and Braverman and Weinstein [17] showed how to compress the communication to $2^{O(I)}$ bits. Ganor, Kol, and Raz [30, 32, 31] proved that this upper bound is optimal in general. Therefore, it is in general impossible to compress interactive communication to its information content. Put another way, efficient protocol compression requires additional assumptions—structural, information-theoretic, or otherwise—on the original communication protocol.

A fundamental special case of the protocol compression problem occurs when Alice and Bob's inputs X and Y are distributed independently. In this setting, the best results prior to my work achieved compression to $I \operatorname{polylog}(C)$ bits [10] and $I^2 \operatorname{polylog}(I)$ bits [45], where I and C are the original protocol's information cost and communication cost, respectively. These bounds are incomparable, with the former bound becoming meaningless for C large enough or infinite. The best possible compression cost remained unknown, with $\tilde{O}(I)$ being the most natural conjecture. In [75], I resolved this problem in full by giving a compression scheme that achieves communication cost $O(I \log^2 I)$ bits, independent of the communication cost C of the original protocol. My compression essentially matches the well-known lower bound of $\Omega(I)$.

3 Quantum Computing

Today's fastest algorithms for factoring integers run in exponential time. Indeed, the security of the widely used RSA public-key cryptosystem hinges on our inability (as yet?) to factor integers efficiently. In 1994, Peter Shor made a spectacular discovery [83]: he developed a polynomial-time algorithm for factoring integers on a *quantum* computer. Shor's breakthrough spurred an intense effort to understand the power and limitations of quantum computing. In this line of research, communication remains a natural and important computational resource to study. Since the introduction of quantum communication complexity by Yao [90] thirty years ago, the subject has become an integral part of communication complexity theory and is an essential part of my research program.

3.1 Lower Bounds for Quantum Communication

Analogous to the classical case, quantum communication features two parties who exchange quantum messages according to an agreed-upon protocol in order to solve a two-party communication problem $F\colon X \times Y \to \{0,1\}$. As always, an input $(x,y) \in X \times Y$ is split between the parties, with one party knowing only x and the other party knowing only y. We allow arbitrary *prior entanglement* at the start of the communication. A measurement at the end of the protocol produces a single-bit answer, which is interpreted as the protocol output. Proving lower bounds for bounded-error quantum communication is significantly more challenging than for randomized communication. An illustrative example is the two-party set disjointness problem on n bits. Babai, Frankl, and Simon [5] obtained an $\Omega(\sqrt{n})$ randomized communication lower bound using a short and elementary proof, which was later improved to a tight $\Omega(n)$ in [37, 55, 9]. This is in stark contrast with the quantum model, where the best lower bound for two-party set disjointness was for a long time a trivial $\Omega(\log n)$ until a tight $\Omega(\sqrt{n})$ lower bound was proved in a breakthrough by Razborov [56].

My technique for communication lower bounds, the pattern matrix method [64, 66], applies not only to classical communication but quantum as well. This makes it possible to prove quantum communication lower bounds in a blackbox manner by appealing to lower bounds on approximate degree. As of this writing, the pattern matrix method is one of the strongest and most general tools ever developed for quantum communication complexity. In particular, the pattern matrix method gives a completely new and much simpler proof of Razborov's celebrated quantum lower bounds [56], including his $\Omega(\sqrt{n})$ lower bound for the set disjointness problem. Quantum lower bounds obtained via the pattern matrix method generalize immediately to multiparty communication (the number-on-the-forehead quantum communication model).

In my latest contribution to the area, I obtained for every $\delta > 0$ an $\Omega(n^{1-\delta})$ lower bound on the quantum communication complexity of polynomial-size DNF formulas. This result definitively resolves the quantum communication complexity of $\mathbf{AC}^0$, showing that even the simplest circuits in this class require near-maximum communication. Prior to my work, an $\Omega(n^{1-\delta})$ quantum lower bound was only known [24] for $\mathbf{AC}^0$ circuits of depth that grows with $1/\delta$.

3.2 An Optimal Separation of Quantum and Classical Query Complexity

Understanding the relative power of quantum and classical computing is of basic importance in theoretical computer science. This question has been studied most actively in the *query model*, which is tractable enough to allow unconditional lower bounds yet rich enough to capture most of the known quantum algorithms. Illustrative examples include the quantum algorithms of Deutsch and Jozsa [29], Bernstein and Vazirani [12], Grover [35], and Shor's period finding [83]. In the query

model, the task is to evaluate a fixed function f on an unknown n-bit input x. In the classical setting, query algorithms are commonly referred to as *decision trees.* A decision tree accesses the input one bit at a time, choosing the bits to query in an adaptive fashion. The objective is to determine $f(x)$ by querying as few bits as possible. The quantum model is a far-reaching generalization of the classical decision tree whereby all bits can be queried in superposition with a single query. The catch is that the outcomes of those queries are then also in superposition, and it is not clear a priori whether quantum query algorithms are more efficient than decision trees.

The comparative power of randomized and bounded-error quantum query algorithms has been studied for more than two decades. Simon [84] exhibited a problem with bounded-error quantum query complexity $O(\log^2 n)$ and randomized query complexity $\Omega(\sqrt{n})$, a striking example of the computational advantages afforded by the quantum model. The raises a fundamental question: what is the largest possible separation between bounded-error quantum and randomized query complexity, for a problem with n-bit input? This question was formally posed by Buhrman et al. [19] and, a decade later, by Aaronson and Ambainis [1], who presented it as being essential to understanding the phenomenon of quantum speedups. Toward this goal, the authors of [1] exhibited a problem that can be solved to bounded error with a single quantum query but has randomized query complexity $\tilde{\Omega}(\sqrt{n})$. Tal [85] gave an improved separation, of $O(1)$ versus $\Omega(n^{2/3-\epsilon})$, for bounded-error quantum and randomized query complexities for any constant $\epsilon > 0$. The challenge of obtaining an optimal separation remained wide open.

In joint work [79], my Ph.D. students Andrey Storozhenko and Pei Wu and I resolved this problem in full. For every integer $k \geqslant 1$, we obtained a separation of k versus $\tilde{\Omega}(n^{1-\frac{1}{2k}})$ between bounded-error quantum query complexity and randomized query complexity. Our separation is best possible for every k, by the results of Aaronson and Ambainis [1] and Bravyi et al. [18]. Furthermore, we obtained an essentially optimal separation of $O(\log n)$ versus $\Omega(n^{1-\epsilon})$ for bounded-error quantum versus randomized *communication* complexity, for any $\epsilon > 0$. The best previous separation [85] was polynomially weaker: $O(\log n)$ versus $\Omega(n^{2/3-\epsilon})$. Independently and concurrently with our work, analogous quantum-classical separations were obtained by Bansal and Sinha [8] with completely different techniques.

3.3 Strong Direct Product Theorems

A natural question to ask of any computational model is how the resources needed to solve n instances of a problem scale with n. More concretely, suppose that solving a single instance of a given decision problem, with probability of correctness $2/3$, requires R units of a computational resource (such as time, memory, communication, or queries). How many units of the resource are needed to solve n independent instances of the problem? Common sense suggests that the answer should be on the order of nR. After all, with significantly fewer than R units per instance, any algorithm seems doomed to have to guess random answers for many of the instances, resulting in overall success probability $2^{-\Omega(n)}$. Such a statement is called a *strong direct product theorem.* A related notion is an *XOR lemma*, which asserts that computing the XOR of the answers to the n problem instances requires $\Omega(nR)$ resources, even if one is willing to settle for a success probability of $\frac{1}{2} + 2^{-\epsilon n}$ (where $\epsilon > 0$ is a small enough absolute constant). While highly plausible, XOR lemmas and strong direct product theorems are notoriously hard (and sometimes impossible) to prove. To some extent, the difficulty stems from the claimed exponential decay in the probability of successful computation. Dropping this part of the claim from strong direct product theorems results in *direct sum theorems*, which nevertheless are also very challenging.

In particular, it is a longstanding problem to prove a strong direct product theorem for quantum communication ("solving n instances of a communication problem requires $\Omega(n)$ times the commu-

nication cost of a single instance, even for probability of correctness exponentially close to random guessing"). Prior to my work, progress on this question was quite limited. My paper [67] comes close to fully resolving the question. More precisely, I prove a strong direct product theorem for *generalized discrepancy*, a complexity measure which from a practical standpoint is equivalent to quantum communication complexity: the former always gives a lower bound on the latter, and they are equivalent for every communication problem studied to date. In particular, my work gives a strong direct product theorem for every quantum communication problem studied in the literature.

4 Computational Learning

Humans have a striking ability to learn from past experience. Is it possible to replicate this ability with a computer algorithm, at least for well-defined and highly structured tasks? *Computational learning theory* studies precise mathematical models of learning and aims to design efficient learning algorithms or to understand why some learning tasks are inherently hard. In the canonical model due to Valiant [86], known as PAC learning, the learner receives values of an unknown Boolean function f at independently sampled points from the domain. The goal is to efficiently find an accurate approximation to f, based on these training examples and the fact that f belongs to some known class $\mathscr{C}$ of functions.

Efficient algorithms have been discovered for several classes $\mathscr{C}$. Other important classes, such as DNF formulas, $\mathbf{AC}^0$ circuits, and intersections of halfspaces, have remained off-limits despite decades of extensive research. I have established strong intractability results for these key learning problems, which I now discuss.

4.1 Intersections of Halfspaces

A *halfspace* in $\mathbb{R}^n$, also known as a *linear threshold function*, is a function of the form $f(x) = \operatorname{sgn}(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n - \theta)$ for some fixed reals $\alpha_1, \alpha_2, \ldots, \alpha_n, \theta$. While there are many efficient algorithms for learning a single halfspace, it is a longstanding challenge in the area to efficiently learn intersections of two or more halfspaces [13, 48, 87, 4, 41, 42, 40]. Klivans and I explained this lack of progress by establishing the first exponential lower bounds [43] on the sample complexity of learning the intersection of n halfspaces in the *statistical-query model* [38]. This work received the "Best Student Paper" award at the 19th Annual Conference on Learning Theory (COLT 2006). In subsequent work [44], Klivans and I proved that under widely accepted cryptographic assumptions, there is no efficient learning algorithm for intersections of n halfspaces even in the standard PAC model. These intractability results closed an important line of research in computational learning.

The only possibility left open by our work was that of efficiently learning the intersection of k halfspaces for small k, such as $k = 2$. I addressed this question in [69, 71]. I focused on learning *by sign-representing polynomials*, a generalization of the perceptron algorithm and one of the few methods known for learning under arbitrary distributions. I proved that the intersection of even two halfspaces has threshold degree $\Theta(n)$, i.e., requires a sign-representing polynomial of the largest possible degree. This means that the perceptron approach cannot yield anything better than a trivial, exponential-time algorithm. My work solved an open problem due to Klivans [39] and received the "Best Student Paper" award at the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009). In [77], I further gave an *explicit* construction of the intersection of two halfspaces with threshold degree $\Theta(n)$.

4.2 DNF Formulas and Constant-Depth Circuits

Progress in computational learning theory has been achieved in great part via *analytic methods*, such as representations of unknown functions by halfspaces or their higher-degree analogues. I have contributed several articles on analytic complexity measures of learning problems, clarifying the relations among them and exposing the limitations of analytic methods. In [63], I proved an exponential separation between the *sign-rank* and *margin complexity* of a natural learning problem, showing that high-margin classification can be off-limits even on learning problems with a simple geometry. In follow-up work [57], Razborov and I studied DNF formulas. Learning this concept class efficiently is a central unresolved problem in computational learning theory. Razborov and I proved that DNF formulas of size n have sign-rank $\exp(\Theta(n^{1/3}))$, which is the first unconditional, exponential lower bound for learning DNF formulas by a wide range of algorithms (including perceptrons of arbitrary degree and any other finite-dimensional kernel methods).

Recall that a polynomial-size DNF formula is a special case of an $\mathbf{AC}^0$ circuit, with depth bounded by 2. A fundamental question that long puzzled researchers in learning theory and communication complexity theory alike is whether every $\mathbf{AC}^0$ circuit has a perceptron of relatively low degree, i.e., can be represented by the sign of a low-degree real polynomial. An answer in the affirmative would give an efficient learning algorithm for this concept class; a strong answer in the negative would be an important impossibility result in learning and communication complexity theory. Recall from Section 2.4 that I fully resolved this question, in a series of works [74, 76, 80] that spanned 11 years. In the latest work, my Ph.D. student Pei Wu and I proved that $\mathbf{AC}^0$ circuits have threshold degree $\Omega(n^{1-\delta})$ and sign-rank $\exp(\Omega(n^{1-\delta}))$ for every $\delta > 0$. This result breaks the longstanding Minsky–Papert barrier from 1969 and shows that $\mathbf{AC}^0$ has essentially the highest possible threshold degree and sign-rank. In particular, $\mathbf{AC}^0$ circuits cannot be learned in time $2^{o(n)}$ by sign-representing polynomials or any other finite-dimensional kernel methods, which are among the most powerful and successful algorithmic tools in the area.

5 Future Directions

I will focus here on three of the most important open problems in communication complexity theory, which are a top priority in my research program. All three are basic scientific questions about the power and limitations of communication in the physical universe that we occupy. Their resolution will additionally have a revolutionary impact on other areas of theoretical computer science, including circuit complexity, computational learning theory, and quantum computing.

Quantifiers in communication. A deep and pervasive idea in computer science is the use of existential and universal quantifiers as a vehicle in computation. The familiar class $\mathbf{P}$ involves no quantifiers, whereas $\mathbf{NP}$ features a single quantifier (corresponding to the existence of a membership certificate). Allowing two or more quantifiers results in a class known as the *polynomial hierarchy*, denoted $\mathbf{PH}$. Understanding the computational power of quantifiers is one of the most important open problems in science. Analogous to these Turing complexity classes, two-party communication has its own classes $\mathbf{P}^{cc}, \mathbf{NP}^{cc}, \mathbf{PH}^{cc}$. For example, the polynomial hierarchy $\mathbf{PH}^{cc}$ consists of all communication problems $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ of the form $F(x,y) = \bigwedge_{i_1=1}^{m} \bigvee_{i_2=1}^{m} \bigwedge_{i_3=1}^{m} \cdots \bigvee_{i_k=1}^{m} A_{i_1,i_2,\ldots,i_k}(x) \wedge B_{i_1,i_2,\ldots,i_k}(y)$, where k is an arbitrary constant, the fan-ins satisfy $m = \exp(\log^{O(1)} n)$, and the functions $A_{i_1,i_2,\ldots,i_k}, B_{i_1,i_2,\ldots,i_k}\colon \{0,1\}^n \to \{0,1\}$ are arbitrary. It is a longstanding challenge in communication complexity theory to exhibit a communication problem not solvable in the polynomial hierarchy. I am developing a novel, spectral line of attack on the problem. If successful, my approach will also give a nonuniform circuit fam-

ily of quasipolynomial size for PAC learning DNF formulas and $\mathbf{AC}^0$ circuits under an arbitrary distribution, solving a major open problem in computational learning theory.

Multiparty communication. It is a significant challenge to analyze the communication requirements of functions whose arguments are distributed among many parties. This challenge gets harder as the number of parties grows. As usual, we are interested in the standard model of multiparty communication (the number-on-the-forehead model), which allows an arbitrary overlap among the arguments available to different parties. Known approaches in this setting break down completely as soon as the number of parties exceeds the *logarithm* of the bit length of the function's arguments, a famous obstacle in theoretical computer science known as the *logarithmic barrier* [7]. Breaking this barrier is of basic scientific importance, not to mention its vast consequences for circuit complexity theory and other areas. I am actively working on resolving this challenge.

Quantum versus classical communication. Whether quantum communication can be substantially more powerful than classical is a major open problem in both quantum computing and communication complexity theory. While quantum communication protocols are known to afford exponential savings in sampling and computing partial functions, it remains a central open problem whether quantum communication can be exponentially more efficient at any *total* communication problem $F\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. This question is closely related to what is perhaps the most famous open problem in communication, the *log-rank conjecture*: roughly speaking, both problems ask whether a *combinatorial view* of communication is equivalent to an *analytic view*.

References

[1] S. Aaronson and A. Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM J. Comput.*, 47(3):982–1038, 2018.

[2] E. Allender. A note on the power of threshold circuits. In *Proceedings of the Thirtieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 580–584, 1989.

[3] A. Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[4] R. I. Arriaga and S. Vempala. An algorithmic theory of learning: Robust concepts and random projection. *Mach. Learn.*, 63(2):161–182, 2006.

[5] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the Twenty-Seventh Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 337–347, 1986.

[6] L. Babai, T. P. Hayes, and P. G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001.

[7] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. Syst. Sci.*, 45(2):204–232, 1992.

[8] N. Bansal and M. Sinha. k-forrelation optimally separates quantum and classical query complexity. In *Proceedings of the Fifty-Third Annual ACM Symposium on Theory of Computing* (STOC), pages 1303–1316, 2021.

[9] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[10] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. *SIAM J. Comput.*, 42(3):1327–1363, 2013.

[11] P. Beame and T. Huynh. Multiparty communication complexity and threshold circuit size of AC^0. *SIAM J. Comput.*, 41(3):484–518, 2012.

[12] E. Bernstein and U. V. Vazirani. Quantum complexity theory. *SIAM J. Comput.*, 26(5):1411–1473, 1997.

[13] A. Blum and R. Kannan. Learning an intersection of a constant number of halfspaces over a uniform distribution. *J. Comput. Syst. Sci.*, 54(2):371–380, 1997.

[14] M. Braverman. Interactive information complexity. *SIAM J. Comput.*, 44(6):1698–1739, 2015.

[15] M. Braverman, R. Gelles, J. Mao, and R. Ostrovsky. Coding for interactive communication correcting insertions and deletions. *IEEE Trans. Information Theory*, 63(10):6256–6270, 2017.

[16] M. Braverman and A. Rao. Toward coding for maximum errors in interactive communication. *IEEE Trans. Information Theory*, 60(11):7248–7255, 2014.

[17] M. Braverman and O. Weinstein. A discrepancy lower bound for information complexity. *Algorithmica*, 76(3):846–864, 2016.

[18] S. Bravyi, D. Gosset, D. Grier, and L. Schaeffer. Classical algorithms for Forrelation. Available at `https://arxiv.org/abs/2102.06963`, 2021. October 2021.

[19] H. Buhrman, L. Fortnow, I. Newman, and H. Röhrig. Quantum property testing. *SIAM J. Comput.*, 37(5):1387–1400, 2008.

[20] H. Buhrman, I. Newman, H. Röhrig, and R. de Wolf. Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007.

[21] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), pages 24–32, 2007.

[22] M. Bun, R. Kothari, and J. Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *Theory Comput.*, 16:1–71, 2020.

[23] M. Bun and J. Thaler. Dual lower bounds for approximate degree and Markov–Bernstein inequalities. *Inf. Comput.*, 243:2–25, 2015.

[24] M. Bun and J. Thaler. A nearly optimal lower bound on the approximate degree of $\mathbf{AC}^0$. *SIAM J. Comput.*, 49(4), 2020.

[25] A. Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proceedings of the Forty-Eighth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 449–458, 2007.

[26] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. In *Electronic Colloquium on Computational Complexity (ECCC)*, January 2008. Report TR08-002.

[27] M. David and T. Pitassi. Separating NOF communication complexity classes RP and NP. In *Electronic Colloquium on Computational Complexity (ECCC)*, February 2008. Report TR08-014.

[28] M. David, T. Pitassi, and E. Viola. Improved separations between nondeterministic and randomized multiparty communication. *ACM Transactions on Computation Theory (TOCT)*, 1(2), 2009.

[29] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A*, 439:553–558, 1992.

[30] A. Ganor, G. Kol, and R. Raz. Exponential separation of information and communication. In *Proceedings of the Fifty-Fifth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 176–185, 2014.

[31] A. Ganor, G. Kol, and R. Raz. Exponential separation of communication and external information. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (STOC), pages 977–986, 2016.

[32] A. Ganor, G. Kol, and R. Raz. Exponential separation of information and communication for Boolean functions. *J. ACM*, 63(5):46:1–46:31, 2016.

[33] D. Gavinsky and A. A. Sherstov. A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(10):227–245, 2010.

[34] V. Grolmusz. The BNS lower bound for multi-party protocols in nearly optimal. *Inf. Comput.*, 112(1):51–54, 1994.

[35] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (STOC), pages 212–219, 1996.

[36] P. Høyer, M. Mosca, and R. de Wolf. Quantum search on bounded-error inputs. In *Proceedings of the Thirtieth International Colloquium on Automata, Languages and Programming* (ICALP), pages 291–299, 2003.

[37] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[38] M. J. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998.

[39] A. R. Klivans. *A Complexity-Theoretic Approach to Learning*. PhD thesis, MIT, 2002.

[40] A. R. Klivans, P. M. Long, and A. K. Tang. Baum's algorithm learns intersections of halfspaces with respect to log-concave distributions. In *Proceedings of the Thirteenth International Workshop on Randomization and Computation* (RANDOM), pages 588–600, 2009.

[41] A. R. Klivans, R. O'Donnell, and R. A. Servedio. Learning intersections and thresholds of halfspaces. *J. Comput. Syst. Sci.*, 68(4):808–840, 2004.

[42] A. R. Klivans and R. A. Servedio. Learning intersections of halfspaces with a margin. *J. Comput. Syst. Sci.*, 74(1):35–48, 2008.

[43] A. R. Klivans and A. A. Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2–3):97–114, 2007.

[44] A. R. Klivans and A. A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. *J. Comput. Syst. Sci.*, 75(1):2–12, 2009.

[45] G. Kol. Interactive compression for product distributions. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* (STOC), pages 987–998, 2016.

[46] M. Krause and P. Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theor. Comput. Sci.*, 174(1–2):137–156, 1997.

[47] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

[48] S. Kwek and L. Pitt. PAC learning intersections of halfspaces with membership queries. *Algorithmica*, 22(1/2):53–75, 1998.

[49] T. Lee and A. Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[50] N. S. Mande, J. Thaler, and S. Zhu. Improved approximate degree bounds for k-distinctness. In *Proceedings of the 15th Conference on the Theory of Quantum Computation, Communication and Cryptography* (TQC), volume 158, pages 2:1–2:22, 2020.

[51] M. L. Minsky and S. A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass., 1969.

[52] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[53] R. O'Donnell and R. A. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010.

[54] R. Paturi and J. Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[55] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[56] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Sciences, Mathematics*, 67:145–159, 2002.

[57] A. A. Razborov and A. A. Sherstov. The sign-rank of $\mathbf{AC}^0$. *SIAM J. Comput.*, 39(5):1833–1855, 2010.

[58] L. J. Schulman. Communication on noisy channels: A coding theorem for computation. In *Proceedings of the Thirty-Third Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 724–733, 1992.

[59] L. J. Schulman. Deterministic coding for interactive communication. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), pages 747–756, 1993.

[60] L. J. Schulman. Coding for interactive communication. *IEEE Trans. Information Theory*, 42(6):1745–1756, 1996.

[61] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[62] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(4):623–656, October 1948.

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

[64] A. A. Sherstov. Separating $\mathbf{AC}^0$ from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009.

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

[66] A. A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[67] A. A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012.

[68] A. A. Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9(20):653–663, 2013.

[69] A. A. Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013.

[70] A. A. Sherstov. Making polynomials robust to noise. *Theory of Computing*, 9:593–615, 2013.

[71] A. A. Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013.

[72] A. A. Sherstov. Communication lower bounds using directional derivatives. *J. ACM*, 61(6):1–71, 2014.

[73] A. A. Sherstov. The multiparty communication complexity of set disjointness. *SIAM J. Comput.*, 45(4):1450–1489, 2016.

[74] A. A. Sherstov. Breaking the Minsky–Papert barrier for constant-depth circuits. *SIAM J. Comput.*, 47(5):1809–1857, 2018.

[75] A. A. Sherstov. Compressing interactive communication under product distributions. *SIAM J. Comput.*, 47(2):367–419, 2018.

[76] A. A. Sherstov. The power of asymmetry in constant-depth circuits. *SIAM J. Comput.*, 47(6):2362–2434, 2018.

[77] A. A. Sherstov. The hardest halfspace. *Comput. Complex.*, 30(11):1–85, 2021.

[78] A. A. Sherstov. The approximate degree of DNF and CNF formulas. In *Proceedings of the Fifty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2022. To appear.

[79] A. A. Sherstov, A. A. Storozhenko, and P. Wu. An optimal separation of randomized and quantum query complexity. In *Proceedings of the Fifty-Third Annual ACM Symposium on Theory of Computing* (STOC), pages 1289–1302, 2021.

[80] A. A. Sherstov and P. Wu. Near-optimal lower bounds on the threshold degree and sign-rank of $\mathbf{AC}^0$. In *Proceedings of the Fifty-First Annual ACM Symposium on Theory of Computing* (STOC), pages 401–412, 2019.

[81] A. A. Sherstov and P. Wu. Optimal interactive coding for insertions, deletions, and substitutions. *IEEE Trans. Inf. Theory*, 65(10):5971–6000, 2019.

[82] Y. Shi. Approximating linear restrictions of Boolean functions. Manuscript, 2002.

[83] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997.

[84] D. R. Simon. On the power of quantum computation. *SIAM J. Comput.*, 26(5):1474–1483, 1997.

[85] A. Tal. Towards optimal separations between quantum and randomized query complexities. In *Proceedings of the Sixty-First Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 228–239, 2020.

[86] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.

[87] S. Vempala. A random-sampling-based algorithm for learning intersections of halfspaces. *J. ACM*, 57(6):32, 2010.

[88] A. C.-C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing* (STOC), pages 209–213, 1979.

[89] A. C.-C. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the Twenty-Fourth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 420–428, 1983.

[90] A. C.-C. Yao. Quantum circuit complexity. In *Proceedings of the Thirty-Fourth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 352–361, 1993.