NOISY DECODING BY SHALLOW CIRCUITS WITH PARITIES: CLASSICAL AND QUANTUM

JOP BRIËT, HARRY BUHRMAN, DAVI CASTRO-SILVA,
AND NIELS M. P. NEUMANN

Abstract. We consider the problem of decoding corrupted error correcting codes with $\mathrm{NC}^0[\oplus]$ circuits in the classical and quantum settings. We show that any such classical circuit can correctly recover only a vanishingly small fraction of messages, if the codewords are sent over a noisy channel with positive error rate. Previously this was known only for linear codes with non-trivial dual distance, whereas our result applies to any code. By contrast, we give a simple quantum circuit that correctly decodes the Hadamard code with probability $\Omega(\varepsilon^2)$ even if a $(1/2-\varepsilon)$-fraction of a codeword is adversarially corrupted.

Our classical hardness result is based on an equidistribution phenomenon for multivariate polynomials over a finite field under biased input-distributions. This is proved using a structure-versus-randomness strategy based on a new notion of rank for high-dimensional polynomial maps that may be of independent interest.

Our quantum circuit is inspired by a non-local version of the Bernstein-Vazirani problem, a technique to generate "poor man's cat states" by Watts et al., and a constant-depth quantum circuit for the OR function by Takahashi and Tani.

1. Introduction

Error correcting codes (ECCs), formally introduced in Shannon's celebrated work [Sha48], protect digital signals from noise. An ECC is a map $E : \Sigma^k \to \Sigma^n$, for a finite alphabet Σ and positive integers $n \geq k$, with the property that any message $x \in \Sigma^k$ can be decoded from the codeword $E(x)$ even if the codeword is partially corrupted. If too many errors occur, however, recovering the original message may become impossible. In such cases one can instead resort to *list decoding*, an influential idea proposed in seminal works of Elias [Eli57] and Wozencraft [Woz58], which aims to give a small list of messages whose

This work was supported by the Dutch Research Council (NWO/OCW), as part of the Quantum Software Consortium programme (project number 024.003.037), the NETWORKS programme (grant no. 024.002.003), and the Quantum Delta NL program.

codewords are close to the received (corrupted) codeword. Complexity considerations appear naturally in this context, as encoding and decoding ideally allow for reliable communication with limited computational resources; they also appear because of the fundamental role played by ECCs in computational complexity itself (see e.g., [Tre04] for a survey).

1.1. **Error models.** In the error model considered by Shannon [Sha48], a codeword is corrupted according to some random process. A natural such process is given by the *symmetric channel*: for each coordinate of the codeword independently, the channel either transmits it unchanged with some probability ρ, or replaces it with a uniformly random element of Σ with probability $1-\rho$. We refer to ρ as the *bias* of the channel.[1] If $Z \in \Sigma^n$ is distributed according to the random outcome of the symmetric channel with bias ρ applied to a codeword $E(x)$, we write $Z \sim \mathcal{N}_\rho\big(E(x)\big)$. In this model the goal is to correctly decode a corrupted codeword with good probability over the noise.

The combinatorial worst-case error model of Hamming [Ham50] instead assumes that the codeword is corrupted arbitrarily on at most some $\delta \in [0,1)$ fraction of coordinates. We will refer to δ as the *error parameter*. In this setting, the number of errors that can be tolerated depends on the minimal Hamming distance between any pair of distinct codewords, or *minimal distance* of the code, denoted d_E. Since the Hamming ball of diameter $d_E - 1$ around any point $y \in \Sigma^n$ contains at most one codeword, a message can be retrieved if fewer than $d_E/2$ errors have occurred.

If more errors occur, faithful decoding is no longer possible and list decoding enters the picture. For $\delta \in [0,1)$ and positive integer L, a code is (δ, L)-*list decodable* if for any point $y \in \Sigma^n$, the Hamming ball of radius δn centered around y contains at most L codewords. It is well-known that any (δ, L)-list decodable code satisfies $L \geq \Omega(1/\varepsilon^2)$ when $\delta = (1-\varepsilon)(1-|\Sigma|^{-1})$ [GV10]. If fewer than a δ-fraction of codeword coordinates are corrupted, then a random element from this list will give the correct message with probability at least $1/L$.

1.2. **Circuits.** A well-studied problem is that of decoding corrupted ECCs by constant-depth circuits with n inputs, k outputs and size $\mathrm{poly}(n)$, for example in the context of black-box hardness amplification [STV99, TV07, Vio06]. Two classes of such circuits are AC^0, consisting of unbounded-fan-in AND, OR and NOT gates, and the class NC^0, consisting of arbitrary bounded-fan-in gates; without loss of generality, we may assume that the fan-in of any gate in NC^0 circuits is at most two.

[1] In this model, each coordinate is thus corrupted with probability $(1-\rho)(1-|\Sigma|^{-1})$, which is usually referred to as the *error rate*. For our purposes, however, the bias will be a more convenient parameterization.

The extensions of these classes where unbounded-fan-in parity gates are added to the gate sets are denoted by $\mathrm{AC}^0[\oplus]$ and $\mathrm{NC}^0[\oplus]$, respectively. These are proper extensions since parity cannot be computed by AC^0 circuits and NC^0 is a proper subset of AC^0 (see [AB09]). An important distinction is that the outputs of NC^0 circuits depend on only a constant number of coordinates of the input, whereas the outputs of $\mathrm{NC}^0[\oplus]$ circuits can depend on the whole input. The classes AC^0 and $\mathrm{NC}^0[\oplus]$ are incomparable since $\mathrm{NC}^0[\oplus]$ cannot compute the n-bit AND function; indeed, $\mathrm{NC}^0[\oplus]$ circuits can compute only constant-degree polynomials over $\mathbb{F}_2$ (see Section 3), whereas AND has degree n.

We also consider the quantum counterparts of the above circuit classes, denoted QX, where X is one of the classes discussed above; these classes were first introduced by Moore [Moo99] and Moore and Nilsson [MN01]. In contrast with their classical analogues, the classes $\mathrm{QNC}^0[\oplus]$ and QAC^0 are known to be equivalent [Moo99, GHMP02, HŠ05].[2]

1.3. **Quantum advantage.** The above-mentioned classes of quantum circuits have recently enjoyed renewed interest in the context of provable separations between quantum and classical computational complexity classes.

One of the principal challenges in quantum computing is to determine for which types of problems quantum computers offer a significant advantage over classical ones. Celebrated examples of practical importance, such a Shor's algorithm for integer factoring [Sho97], require quantum computers of a vastly larger scale than currently available. Moreover, formally proving classical hardness of factoring appears to be beyond the scope of currently-available techniques. Constant-depth circuits form an attractive computational model, as they will likely be easier to implement in practice and, from the perspective of complexity theory, provide one of the few settings currently amenable to provable lower bounds.

A recent series of works, starting with a breakthrough of Bravyi, Gosset and König [BGK18], considered the relative power of *shallow quantum circuits*, in particular QNC^0 [BWKST19, Gal19, BGKT20, GS20, CSV21]. A limitation of this class, similar to NC^0, is that the outputs of a QNC^0 circuit can only depend on a constant number of input-bits. For this reason, QNC^0 circuits are unsuited for the problem of (list) decoding a corrupted ECC, or more generally for any problem that admits a unique valid solution (i.e., a *function problem*). This

[2] The works of Moore furthermore showed that these classes are all equivalent to $\mathrm{QAC}^0[q]$, the class QAC^0 with additional modulo-q gates. For an integer $q > 1$, a modulo-q gate evaluates to 1 if the sum of its inputs equals 0 mod q and evaluates to 0 otherwise. Classically, the classes $\mathrm{AC}^0[p]$ and $\mathrm{AC}^0[q]$ are incomparable if p and q are powers of distinct primes [Raz87, Smo87].

is what motivates us to consider the extension $\mathrm{QNC}^0[\oplus]$ obtained by adding unbounded-fan-in parity gates, as well as its classical analogue $\mathrm{NC}^0[\oplus]$.

1.4. The Hadamard code.

A basic but important example of an ECC is the *Hadamard code*, which encodes k-bit messages into codewords of length $n = 2^k$ and is given by the $\mathbb{F}_2$-linear map $H(x) = (\langle x, y\rangle)_{y\in\mathbb{F}_2^k}$, where $\langle x, y\rangle = y^{\mathsf{T}}x$. This code has minimal distance $n/2$ and is $(1/2-\varepsilon, O(1/\varepsilon^2))$-list decodable for any $\varepsilon \in (0, 1/2]$, which is known to be optimal for any code [GV10].

Under the symmetric channel, the Chernoff bound implies that unique decoding of the Hadamard code is possible with high probability for any constant bias $\rho > 0$.[3] This is due to the fact that, with high probability, the Hamming ball of radius $(1/4-\rho/4)n$ around a corrupted version of a codeword C contains no other codewords than C itself.

For the worst-case Hamming model, Goldreich and Levin [GL89] famously gave an efficient list decoding algorithm for the Hadamard code that runs in time $\mathrm{poly}(k, 1/\varepsilon)$, for error parameter $\delta = 1/2 - \varepsilon$. For fixed $\varepsilon > 0$, their algorithm gives a probabilistic AC^0 circuit that, on input length n, correctly returns the original message with probability $\Omega(1)$.

2. Our results

Here we consider the following problem. Let $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ be a (binary) error correcting code. Given a map $\phi : \mathbb{F}_2^n \to \mathbb{F}_2^k$ representing some decoding procedure, we wish to bound the probability of correct message retrieval:

$$\Pr\big[\phi\big(E(x) + Z\big) = x\big], \tag{1}$$

where $x \in \mathbb{F}_2^k$ is some message and $Z \in \mathbb{F}_2^n$ is an error string. We consider two scenarios, one classical and one quantum.

2.1. Classical setting.

In the first scenario, ϕ represents an $\mathrm{NC}^0[\oplus]$ circuit, x is uniformly distributed and $Z \sim \mathcal{N}_\rho(0)$, so that $E(x)+Z$ is a random codeword corrupted according to the binary symmetric channel with bias ρ. Our main result in this setting says that (1) tends to zero, for any $\rho \in [0, 1)$ and any code:

Theorem 2.1 (Impossibility of decoding by $\mathrm{NC}^0[\oplus]$)**.** *For any $\rho \in [0, 1)$, $d \in \mathbb{N}$ and $\varepsilon \in (0, 1]$, there is a $k_0 = k_0(d, \rho, \varepsilon) \in \mathbb{N}$ such that the following holds. Let $k \geq k_0$ and n be positive integers, $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ be any map and $\phi : \mathbb{F}_2^n \to \mathbb{F}_2^k$ be a map computable by an $\mathrm{NC}^0[\oplus]$ circuit of depth at most d. Then, for a uniformly distributed $x \in \mathbb{F}_2^k$ and $Z \sim \mathcal{N}_\rho(0)$, we have that*

$$\Pr\big[\phi\big(E(x) + Z\big) = x\big] < \varepsilon.$$

[3] This even holds for any code over a large enough alphabet, as shown in [RU10].

In particular, this theorem shows that no $\mathrm{NC}^0[\oplus]$ circuit can correctly decode more than an ε-fraction of codewords with probability higher than ε over the noise distribution, if the messages are long enough depending on ε, the error rate $(1-\rho)/2 > 0$ and the depth of the circuit. As a consequence of Yao's minimax principle [Yao77] and the Chernoff bound, it follows that any probabilistic $\mathrm{NC}^0[\oplus]$ circuit will also fail (with high probability) to correctly decode any binary ECC in the worst-case Hamming model, for any constant error parameter $\delta \in (0, 1/2]$.

We note that the decay we obtain on the probability (1) of correct message retrieval as a function of the message length is extremely slow, making Theorem 2.1 a qualitative result rather than quantitative. Nevertheless, we conjecture that the true decay of this probability is exponential in the message length k; this would clearly be optimal, as can be seen by taking a constant map ϕ which always returns some fixed message. In Section 7 we will provide some evidence to support this conjecture.

2.2. **Quantum setting.** In the second scenario, we consider the worst-case Hamming model with constant-depth quantum circuits. Our main result in this setting is an explicit $\mathrm{QNC}^0[\oplus]$ circuit capable of decoding the Hadamard code.

Theorem 2.2 (Decoding Hadamard with $\mathrm{QNC}^0[\oplus]$)**.** *There is a family of* $\mathrm{QNC}^0[\oplus]$ *circuits* $(\mathcal{C}_n)_{n\in\mathbb{N}}$ *such that the following holds. Let* $k \in \mathbb{N}$*,* $n = 2^k$ *and* $\varepsilon \in (0, 1/2]$*. Then, for any* $y \in \mathbb{F}_2^n$ *and any* $x \in \mathbb{F}_2^k$ *satisfying* $d\big(y, H(x)\big) \leq (\frac{1}{2} - \varepsilon)n$*, on input* y *the circuit* $\mathcal{C}_n$ *returns* x *with probability* $\Omega(\varepsilon^2)$*.*

We note that the bound $\Omega(\varepsilon^2)$ obtained in the theorem is optimal, since in general there can be $\Theta(\varepsilon^{-2})$ messages $x \in \mathbb{F}_2^k$ satisfying $d\big(y, H(x)\big) \leq (\frac{1}{2} - \varepsilon)n$. This bound is non-trivial only when $\varepsilon = \Omega(1/\sqrt{n})$, as there are n possible messages.

As a simple corollary of Theorem 2.2, we obtain a similar result for the problem of list decoding the Hadamard code.

Corollary 2.3. *There is a family of* $\mathrm{QNC}^0[\oplus]$ *circuits* $(\mathcal{C}_n)_{n\in\mathbb{N}}$ *such that the following holds. Let* $k \in \mathbb{N}$*,* $n = 2^k$ *and* $\varepsilon \in [1/\sqrt{n}, 1/2]$*. Then, on any input* $y \in \mathbb{F}_2^n$*, with probability* $1 - \varepsilon$ *the circuit* $\mathcal{C}_n$ *returns a list* $L(y)$ *of size* $O\big(\varepsilon^{-2}\log(1/\varepsilon)\big)$ *which contains every* $x \in \mathbb{F}_2^k$ *with* $d\big(y, H(x)\big) \leq (\frac{1}{2} - \varepsilon)n$*.*

Proof: For a large enough constant $C > 0$, consider $C\varepsilon^{-2}\log(1/\varepsilon)$ parallel instances of the circuit from Theorem 2.2. This gives a list $L(y)$ of the claimed size such that any message $x \in \mathbb{F}_2^k$ satisfying $d\big(y, H(x)\big) \leq (\frac{1}{2} - \varepsilon)n$ appears in $L(y)$ with probability at least $1 - \varepsilon^3$. Since there are at most $O(1/\varepsilon^2)$ such messages, it follows from the union bound that with probability at least $1-O(\varepsilon)$ every such message appears in $L(y)$. □

Remark 2.4. Note that the circuits obtained in this corollary also output several messages whose codewords differ from the input y in more than $(\frac{1}{2}-\varepsilon)n$ coordinates; this differs from the usual notion of the list decoding problem, which aims to output a list of all messages $x \in \mathbb{F}_2^k$ with $d\big(y, H(x)\big) \leq (\frac{1}{2}-\varepsilon)n$ and none other. One can also solve the usual list decoding problem for the Hadamard code using $\mathrm{QNC}^0[\oplus]$ circuits, by making use of MAJORITY gates (and more general threshold gates) to prune the obtained list (see Section A.2). We omit the details, as they are not so relevant for us.

As a consequence of Theorem 2.1 and Theorem 2.2, we conclude that the problem of list decoding the Hadamard code separates the complexity classes $\mathrm{NC}^0[\oplus]$ and $\mathrm{QNC}^0[\oplus]$; this holds for any positive error parameter $\delta > 0$. The task of proving quantum advantage for a natural problem such as list decoding was the original motivation for the present work.

In the high-error regime where the parameter δ approaches the information-theoretic limit of $1/2$ (which is relevant for hardness amplification), a stronger separation follows by combining Theorem 2.2 with a result of Sudan showing hardness of noisy decoding by $\mathrm{AC}^0[\oplus]$ circuits (see Corollary 2.7 below).[4] To state this separation theorem precisely, we consider the following problem:

List-Hadamard problem: Let $\varepsilon : \mathbb{N} \to (0,1]$ be a function. For each dyadic number $n = 2^k$ we define the problem $\mathrm{LH}_n(\varepsilon)$ as follows: given $y \in \mathbb{F}_2^n$, output a list of at most $n/4$ elements in $\mathbb{F}_2^k$ containing every $x \in \mathbb{F}_2^k$ satisfying $d\big(y, H(x)\big) \leq \big(\frac{1}{2}-\varepsilon(n)\big)n$.

The most general form of our quantum advantage result is given by the following theorem:

Theorem 2.5 (Quantum-vs-classical separation). *For any constant $\delta \in (0, \frac{1}{2})$, list decoding the Hadamard code with error parameter δ separates* $\mathrm{QNC}^0[\oplus]$ *from* $\mathrm{NC}^0[\oplus]$. *Moreover, for any $(\log n)/\sqrt{n} \leq \varepsilon(n) \leq 1/(\log n)^{\omega(1)}$, the list-Hadamard problem* $\mathrm{LH}_n(\varepsilon)$ *separates* $\mathrm{QNC}^0[\oplus]$ *from* $\mathrm{AC}^0[\oplus]$.

2.3. **Related results and discussion.** Both the problem of decoding corrupted ECCs and the problem of proving quantum-versus-classical separations of complexity classes are well studied, and there are several results in the literature related to the results presented here.

[4] The same separation of complexity classes can also be obtained by combining other previously-known results; see Section 2.3 for a discussion.

The main strength of our Theorem 2.1 is that it holds for any code and for any positive error rate. Complementary results are known for restricted classes of codes, and also for when the error rate tends to 1/2. We will now expand on some of these results.

A code $E : \mathbb{F}_2^k \to \mathbb{F}_2^n$ is *t-wise independent* if, for any t-subset of coordinates $S \subseteq [n]$ and a uniformly random $X \in \mathbb{F}_2^k$, the restriction $E(X)_{|S}$ is uniformly distributed over $\mathbb{F}_2^S$. Many codes have this property; for instance, the dual code of a linear code of distance d is $(d-1)$-wise independent. Under the same noise model considered here, Lee and Viola [LV17], using earlier work of Viola [Vio09], showed that $\mathrm{NC}^0[\oplus]$ circuits cannot distinguish a corrupted uniformly random codeword of an $\omega(1)$-wise independent linear code from a uniformly random element of $\mathbb{F}_2^n$. Note that this problem is formally easier than (list) decoding.

Their result does not cover the Hadamard code, however, as it is not even 3-wise independent. Indeed, the Hadamard code is also easy to distinguish, as it contains the sub-code $(x_1, x_2, x_1 + x_2)$. Since the parity of these three bits is always zero, the parity under noise is biased towards zero and therefore easily distinguished from the parity of a random string.

In the very high-error regime where the error rate approaches the information-theoretic limit of 1/2 (which is relevant for hardness amplification), stronger results are also known. For instance, Sudan (see [Vio06, Section 6.2]) showed that list decoding with error parameter $1/2 - \varepsilon$ requires probabilistic $\mathrm{AC}^0[\oplus]$ circuits to have size $\exp(\mathrm{poly}(1/\varepsilon))$. Below we state his result when restricted to the Hadamard code, which is done for concreteness and better clarity; as can be easily seen from its proof, one could instead consider any other ECC.

Theorem 2.6 (MAJORITY from list-Hadamard)**.** *Let $\mathcal{C}$ be a probabilistic circuit that solves the list-Hadamard problem $\mathrm{LH}_n(\varepsilon)$ with probability at least $3/4$. There exists a (deterministic) oracle AC^0 circuit $\mathcal{D}$ of size $\mathrm{poly}(n, 1/\varepsilon)$ which, when given oracle access to $\mathcal{C}$ and the ability to fix its random bits, computes MAJORITY on $\Omega(1/\varepsilon)$ bits.*

This result can be readily deduced from Sudan's arguments exposed in [Vio06, Section 6.2]; since it is not given in this form elsewhere, we include its elegant proof in Appendix A. As a corollary, the circuit lower bound for MAJORITY due to Razborov [Raz87] and Smolensky [Smo87] gives the following (known) hardness result for list decoding the Hadamard code.

Corollary 2.7 (Hardness of list-Hadamard)**.** *If $\varepsilon(n) \leq 1/(\log n)^{\omega(1)}$, then the list-Hadamard problem $\mathrm{LH}_n(\varepsilon)$ cannot be solved by a probabilistic $\mathrm{AC}^0[\oplus]$ circuit with probability $\Omega(1)$.*

Combining this corollary with our $\mathrm{QNC}^0[\oplus]$ circuits for list-Hadamard given in Corollary 2.3, we obtain the second separation of complexity classes stated in Theorem 2.5.

The existence of the quantum circuits of Theorem 2.2 and Corollary 2.3 also follows from the Goldreich-Levin algorithm and the surprising fact that MAJORITY can be computed by a $\mathrm{QNC}^0[\oplus]$ circuit. This fact was proved by Høyer and Špalek [HŠ05] with one-sided error and by Takahashi and Tani [TT13] with zero error. The above-mentioned classical hardness of MAJORITY thus already implies a separation between $\mathrm{AC}^0[\oplus]$ and $\mathrm{QNC}^0[\oplus]$, showing that despite its simplicity, the latter class of quantum circuits is remarkably powerful.[5] In the opposite direction, one can use the ideas behind the proof of Theorem 2.6 to show that our quantum circuit from Corollary 2.3 also gives a $\mathrm{QNC}^0[\oplus]$ circuit for MAJORITY, albeit not exact (see Section A.2).

Finally, we note that a key enabling sub-routine in the Høyer-Špalek circuit is the powerful quantum fan-out gate (see below for further details). In our circuit for Corollary 2.3, we construct this gate explicitly using only classical parity gates and single- and two-qubit gates. These gates are native to many quantum architectures and as such, may give an easier way to implement quantum fan-out.

3. Techniques

To establish our main results, we use techniques from two different areas. Broadly speaking, Theorem 2.1 builds on ideas from higher-order Fourier analysis [Tao12, HHL19], while Theorem 2.2 (unsurprisingly) uses ideas from quantum computing [NC10].

3.1. Polynomial equidistribution.

The proof of Theorem 2.1 uses the basic observation that any function $\mathbb{F}_2^n \to \mathbb{F}_2^k$ that is computable by an $\mathrm{NC}^0[\oplus]$ circuit can be given by a collection of k constant-degree polynomials over $\mathbb{F}_2$ in n variables. Indeed, any gate with fan-in d implements a function $\mathbb{F}_2^d \to \mathbb{F}_2$ and any such function can be represented by a d-variable polynomial of total degree at most d. Degree is multiplicative under composition and composition occurs only between different layers of the circuit. Since the parities amount to addition in $\mathbb{F}_2$ and NC^0 circuits have constant depth, the total degree of the output is bounded.

We will therefore study the distribution of polynomial maps under biased input distributions. We will do so in a slightly more general setting over arbitrary

[5] This fact was missed on by the authors when working on the present paper. We thank an anonymous referee for bringing it to our attention.

finite fields of prime order.[6] For a prime p, let $\mathbb{F}_p$ denote the finite field with p elements. For $\rho \in [0,1]$, an $\mathbb{F}_p$-valued random variable Z is *ρ-biased* if with probability ρ it equals 0 and with probability $1-\rho$ it is uniformly distributed over $\mathbb{F}_p$. Note that this corresponds to the noise $\mathcal{N}_\rho(0)$ added by the symmetric channel when the alphabet is $\mathbb{F}_p$.

A mapping $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ is a *polynomial map* if there exist polynomials $f_1, \ldots, f_k \in \mathbb{F}_p[x_1, \ldots, x_n]$ such that $\phi = (f_1, \ldots, f_k)$. The degree of ϕ is the maximal degree among the f_i. To prove Theorem 2.1, it thus suffices to prove the following result.

Theorem 3.1 (Impossibility of decoding by polynomial maps)**.** *For any $d \in \mathbb{N}$ and $\rho, \varepsilon \in (0,1)$ there exists an integer $k_0 = k_0(p,d,\rho,\varepsilon)$ such that the following holds. Let $k \geq k_0$ and n be integers, $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ be a polynomial map of degree at most d and $E : \mathbb{F}_p^k \to \mathbb{F}_p^n$ be an arbitrary function. Then*

$$\mathrm{Pr}_{x \in \mathbb{F}_p^k, Z \sim \mathcal{N}_\rho(0)} \big[\phi\big(E(x) + Z\big) = x\big] \leq \varepsilon.$$

Studying the distribution of polynomial maps in many variables over a finite field falls within the purview of additive combinatorics. In the "unbiased" situation where Z is uniformly distributed there are powerful tools from higher-order Fourier analysis that can be used to study the distribution of $\phi(Z)$. In particular, Green and Tao [GT09] proved that if ϕ is "regular" (random-like), then $\phi(Z)$ is approximately uniformly distributed over $\mathbb{F}_p^k$. This implies that the probability of the event $\{\phi(E(x)+Z) = x\}$ considered is small for every x. A "regularity-type" lemma proved in [GT09] shows that one can "force" ϕ to be regular by restricting it to a partition defined by sufficiently many polynomial equations of degree less than the degree of ϕ. However, these techniques cause the size of the polynomial map ϕ considered to blow up considerably, and are only effective if k is an extremely slowly growing function of n.

In order to deal with this issue, and to adapt these results to the case where Z is no longer uniform but biased, we employ a dichotomy often used in additive combinatorics that studies the "pseudorandom" case of regular maps separately from the "structured" case of maps that carry a certain algebraic structure. This is done by defining and studying a new notion of rank for (high-dimensional) polynomial maps, which we call the *analytic rank*,[7] and

[6] The restriction to prime order is done for notational reasons and for ease of exposition. Our arguments can be readily adapted to the case of non-prime finite fields.

[7] A very similar notion of rank was defined for multilinear forms by Gowers and Wolf [GW11], who coined the term analytic rank. We use the same name to highlight the similarity between our two notions, which are relevant for distinct types of mathematical objects.

which measures how well-equidistributed the values taken by the considered map are.

In the pseudorandom case, a key tool we use is a new random restriction result for high-rank polynomial maps proved in a companion paper [BCS22]. We use this to show that the distribution of values taken by a high-rank polynomial map will be close to uniform even under a biased input distribution. This implies that the event considered in the theorem has very low probability for any fixed x, in which case we can conclude by averaging.

In the structured case we deal instead with polynomial maps of low rank, whose values are in a sense poorly distributed. Results from higher-order Fourier analysis then imply that they can be determined by "few" lower-degree polynomial maps (plus a few extra polynomials); by a simple Fourier-analytic argument we can reduce the analysis of a low-rank polynomial map to those lower-degree maps which specify it, making it amenable to an inductive argument.

3.2. **Building the quantum circuit.**

The quantum circuit of Theorem 2.2 is inspired by a distributed version of the Bernstein-Vazirani algorithm [BV97]. Given a corrupted Hadamard codeword $H(x)$, this single-query quantum algorithm returns x with probability $\Omega(\varepsilon^2)$. The distributed version describes an entangled strategy for a particular non-local game [CHTW04] consisting of n players who, when given unique coordinates of $H(x)$, must each return an element of $\mathbb{F}_2^k$. They win if and only if the sum of their answers equals x. It turns out that by sharing an n-partite GHZ state of local dimension 2^k, they can simulate the Bernstein-Vazirani algorithm and achieve the same success probability. We then turn this entangled strategy into a quantum circuit that only uses single and two-qubit gates and classical parity gates. For this we use two constant-depth sub-routines, one for preparing GHZ states and another for the quantum *fan-out gate* [PS13], which implements the map $|x\rangle\,|y_1\rangle\dots|y_n\rangle \mapsto |x\rangle\,|y_1 \oplus x\rangle\dots|y_n \oplus x\rangle$.

To generate the GHZ state, we use a poor man's cat state [BWKST19], which is a GHZ state with some of its qubits flipped. We correct this poor man's cat state to a GHZ state by flipping qubits based on parity computations. The input for these parity computations follows from the procedure that generates the poor man's cat state.

To implement the quantum fan-out gate, we use ideas from distributed quantum computing. These ideas use GHZ states and classical parity gates together with single and two-qubit gates. With the quantum fan-out gate, we also obtain the quantum parity gate, by conjugating the quantum fan-out gate with Hadamard gates.

Part of the circuit is applying phase-flips, conditional on the bits of the corrupted codeword. To do this, we use quantum fan-out gates in a circuit, exponential in size in k to correctly apply the phase-flips [TT13].

The depth of the list-decoding circuit is constant, whereas the circuit size is $O(n^2 \log n)$. We also show how to reduce this complexity to $O(n \log n \log \log n)$, while increasing the depth by only a small constant number. We do this by preparing a state on $\lceil \log(k+1) \rceil$ qubits. Evaluating an OR on this newly prepared state yields the same result as evaluating an OR on the original k qubits [HŠ05]. Applying the same exponential size circuit as before on this newly prepared state indeed gives the reduced circuit size.

4. Warm-up: The linear case and a non-local game

This section is meant to give some intuition for the proofs of Theorem 3.1 and Theorem 2.2, as well as provide the first steps in those proofs.

4.1. **Impossibility of decoding for linear maps.** To motivate our later arguments, here we present a proof of the first nontrivial case of Theorem 3.1, namely that of maps $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ of degree 1. In this case, there is a matrix $U \in \mathbb{F}_p^{k\times n}$ and a vector $v \in \mathbb{F}_p^k$ such that

$$\phi(y) = Uy + v \quad \text{for all } y \in \mathbb{F}_p^n.$$

Let x be a uniformly distributed random variable over $\mathbb{F}_p^k$ and Z be an $\mathcal{N}_\rho(0)$-distributed random variable over $\mathbb{F}_p^n$. Our goal is then to bound the probability of the event

$$U(E(x)+Z)+v = x. \tag{2}$$

We distinguish two cases based on the rank of U. Let $r \in [k]$ be an integer to be set later. If U has rank at most r, then its image $\mathrm{im}(U)$ is a subspace of size at most p^r. If (2) holds, then x is contained in the coset $v + \mathrm{im}(U)$ of this subspace, which (for x uniform over $\mathbb{F}_p^k$) happens with probability at most p^r/p^k. Hence, (2) holds with probability at most $p^{-(k-r)}$ in this case.

For the "pseudorandom case" of high-rank matrices, we make the following simple but important observation: one can sample $Z \sim \mathcal{N}_\rho(0)$ by first sampling the set $I \subseteq [n]$ of "corrupted coordinates", then sampling the "noise" y uniformly at random from $\mathbb{F}_p^I$ and setting[8] $Z_{|I} = y$, $Z_{|[n]\setminus I} = 0$. Each index $i \in [n]$ has probability $1-\rho$ of belonging to the random set I, with these events being mutually independent; we denote this sampling scheme by $I \sim [n]_{1-\rho}$.

[8] Given $x \in \mathbb{F}_p^n$ and $I \subseteq [n]$, we denote by $x_{|I} \in \mathbb{F}^I$ the restriction of x to the coordinates indexed by I.

If we denote by $U_I \in \mathbb{F}_p^{k\times I}$ the restriction of U to the columns indexed by $I \subseteq [n]$, it follows that the random variable UZ has the same distribution as the random variable $U_I y$, where $I \sim [n]_{1-\rho}$ and y is uniformly distributed over $\mathbb{F}_p^I$. Thus, for any given $x \in \mathbb{F}_p^k$, we have

$$\mathrm{Pr}_{Z\sim\mathcal{N}_\rho(0)}\big[U(E(x)+Z)+v=x\big] = \mathbb{E}_{I\sim[n]_{1-\rho}}\mathrm{Pr}_{y\in F^I}\big[U_I y = x - UE(x) - v\big].$$

Now, if $I \subseteq [n]$ is fixed and y is uniformly distributed over $\mathbb{F}_p^I$, then the random variable $U_I y$ is uniformly distributed over $\mathrm{im}(U_I)$; hence

$$\max_{w\in\mathbb{F}_p^k} \mathrm{Pr}_{y\in F^I}\big[U_I y = w\big] = \frac{1}{|\,\mathrm{im}(U_I)|} = \frac{1}{p^{\mathrm{rk}(U_I)}}.$$

Taking the expectation over $I \sim [n]_{1-\rho}$ and $x \in \mathbb{F}_p^k$, we conclude that event (2) holds with probability at most $\mathbb{E}_{I\sim[n]_{1-\rho}} p^{-\mathrm{rk}(U_I)}$.

Suppose now that U has rank at least r, and let $J \subseteq [n]$ be a set of r linearly independent columns of U. By the Chernoff bound (see e.g. [HR90]), we have that

$$\mathrm{Pr}_{I\sim[n]_{1-\rho}}\left[|I\cap J| \le \frac{(1-\rho)r}{2}\right] \le e^{-(1-\rho)r/8}.$$

Thus U_I will contain more than $(1-\rho)r/2$ linearly independent columns with probability at least $1 - e^{-(1-\rho)r/8}$; whenever this happens we have $\mathrm{rk}(U_I) \ge (1-\rho)r/2$. It follows that

$$\begin{aligned}\mathrm{Pr}_{x\in\mathbb{F}_p^k, Z\sim\mathcal{N}_\rho(0)}\big[U(E(x)+Z)+v=x\big] &\le \mathbb{E}_{I\sim[n]_{1-\rho}} p^{-\mathrm{rk}(U_I)}\\ &\le e^{-(1-\rho)r/8} + p^{-(1-\rho)r/2}\end{aligned}$$

in this “high-rank” case.

Setting $r = k/2$ (say) implies that in both cases the probability that event (2) holds decays exponentially in k, which concludes the analysis.

4.2. **Quantum decoding in a non-local game.** Our quantum algorithm is inspired by the analysis of a particular non-local game. In a non-local game, a referee randomly sends questions to a set of players, according to a probability distribution known to the players in advance. Then, without communicating with each other, the players individually answer the referee. Finally, the referee determines whether the players win or lose based solely on the questions and answers. The rule used by the referee is known to the players in advance as well. With a (deterministic) classical strategy, the players decide before the game starts what to answer to each possible question. With an entangled strategy, the players base their answers on the outcomes of local measurements of their respective parts of a shared entangled state. We refer to [CHTW04] for further background on non-local games.

Let $H : \mathbb{F}_2^k \to \mathbb{F}_2^n$ be the Hadamard code, where $n = 2^k$ and let $\varepsilon \in (0, 1/2)$ be a constant (independent of n). We identify the codewords $H(x)$ with functions $\mathbb{F}_2^k \to \mathbb{F}_2$ given by $H(x)(y) = \langle x, y\rangle$. We consider the following non-local game, which we shall refer to as the *Hadamard game*. There are n players, each labeled uniquely with an element in $\mathbb{F}_2^k$. The referee picks a uniformly chosen message $x \in \mathbb{F}_2^k$ and randomly corrupts the codeword $H(x)$ using the binary symmetric channel with error rate $1/2 - \varepsilon$, resulting in a function $c : \mathbb{F}_2^k \to \mathbb{F}_2$. He then sends player y the value $c(y)$. The players each return a string in $\mathbb{F}_2^k$ and they win the game if the sum of their answers equals x.

Here we show that entangled players can win the Hadamard game with probability $\Omega(1)$. The corresponding strategy is inspired by the famous Bernstein-Vazirani algorithm [BV97]. The strategy is based on an n-partite GHZ state of local dimension 2^k, shared by the n players:

$$\frac{1}{\sqrt{n}} \sum_{y \in \mathbb{F}_2^k} |y\rangle \otimes \ldots \otimes |y\rangle \, .$$

Upon receiving their input $c(y)$, player y applies a conditional phase flip on their part of the shared state:

$$|z\rangle \mapsto \begin{cases} (-1)^{c(y)} |z\rangle & \text{if } z = y \\ |z\rangle & \text{else} \end{cases} . \tag{3}$$

Once all players have done this, they share the state

$$\frac{1}{\sqrt{n}} \sum_{y} (-1)^{c(y)} |y\rangle \otimes \ldots \otimes |y\rangle \, .$$

Each player then applies a k-qubit Hadamard gate to their local register and measures in the computational basis. The measurement results are then returned by each player. The state before the measurements is:

$$n^{-(n+1)/2} \sum_{y \in \mathbb{F}_2^k} \sum_{b_1,\ldots,b_n \in \mathbb{F}_2^k} (-1)^{c(y)} (-1)^{\langle y, b_1 + \cdots + b_n\rangle} |b_1\rangle \otimes \ldots \otimes |b_n\rangle \, .$$

The probability that the measurement results sum to a string $z \in \mathbb{F}_2^k$ is therefore given by

$$\begin{aligned} \Pr\Big[\sum_{i=1}^{n} b_i = z\Big] &= \frac{1}{n^{(n+1)}} \sum_{b_1 + \cdots + b_n = z} \Big| \sum_{y \in \mathbb{F}_2^k} (-1)^{c(y) + \langle y, z\rangle} \Big|^2 \\ &= \Big| 1 - 2\frac{d\big(c, H(z)\big)}{n} \Big|^2 . \end{aligned}$$

It follows from the Chernoff bound [HR90] that for fixed $x \in \mathbb{F}_2^k$ and the random c obtained by corrupting the codeword $H(x)$,

$$\Pr\left[\frac{d\big(c, H(x)\big)}{n} \geq \frac{1-\varepsilon}{2}\right] \leq \exp(-C\varepsilon^2 n).$$

Hence, by the union bound, for fixed $\varepsilon \in (0, 1/2)$, the players win with probability at least $C\varepsilon^2$, where the probability is taken over the message x, the noise corrupting the codeword $H(x)$ to c and the measurements done by the players.

Note that this strategy in fact succeeds with probability $C\varepsilon^2$ for *every* x and whenever at most any $(1/2-\varepsilon)$-fraction of the coordinates of $H(x)$ are flipped.

5. Classical hardness of list decoding

In this section we will prove Theorem 3.1, which – as explained before – implies that $\mathrm{NC}^0[\oplus]$ circuits are unable to perform list decoding, no matter which specific code is considered (see Theorem 2.1 for a formal statement). We will do so by following a similar strategy as we used for maps of degree 1 in Section 4.1, dividing the analysis into the "pseudorandom" case of high-rank polynomial maps and the "structured" case of low-rank polynomial maps.

There is a well-studied notion of rank for polynomials $P \in \mathbb{F}_p[x_1, \ldots, x_n]$, first introduced by Green and Tao [GT09], which is defined (roughly speaking) as the smallest number of lower-degree polynomials needed to compute P. A different notion of rank, called the analytic rank, was later introduced by Gowers and Wolf [GW11] when studying linear systems of equations over $\mathbb{F}_p^n$. It is related to the *bias* of the polynomial P, or more specifically to the bias of the symmetric $\deg(P)$-multilinear form associated to P. The bias of a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ is an analytic measure of how well-equidistributed the values of f are when evaluated on a uniformly random input; formally,

$$\mathrm{bias}(f) = |\mathbb{E}_{x\in\mathbb{F}_p^n}\omega^{f(x)}| \tag{4}$$

where we write $\omega = e^{2i\pi/p}$ for a primitive p-th root of unity.

When dealing with a polynomial P of some bounded degree d, having non-negligible bias implies that it has a significant amount of internal structure. Such a result was first proven by Green and Tao [GT09] in the case of polynomials whose degree d is smaller than the characteristic p of the field considered, and motivated the introduction of both their notion of rank and Gowers and Wolf's notion of analytic rank. We will need a similar result, proven by Kaufman and Lovett [KL08], which generalizes this theorem to characteristics $p \leq d$ and also gives more precise information on the structure of the polynomial.

For a vector $h \in \mathbb{F}_p^n$ and a polynomial $P \in \mathbb{F}_p[x_1, \ldots, x_n]$, the derivative of P in direction h is defined by

$$\Delta_h P(x) = P(x+h) - P(x).$$

Note that $\Delta_h P$ is also a polynomial on $\mathbb{F}_p^n$, and (as with usual derivatives in real analysis) its degree is strictly smaller than $\deg(P)$. The derivatives of a polynomial map $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ are defined analogously, and also satisfy $\deg(\Delta_h \phi) < \deg(\phi)$.

The following result of Kaufman and Lovett shows that polynomials with large bias must be highly structured:

Theorem 5.1 (Bias implies low rank)**.** *For every $d \in \mathbb{N}$ and $\varepsilon > 0$, there is an $r = r(p, d, \varepsilon) \in \mathbb{N}$ such that the following holds. If $P \in \mathbb{F}_p[x_1, \ldots, x_n]$ is a polynomial of degree at most d with* $\operatorname{bias}(P) \geq \varepsilon$*, then there exist $h_1, \ldots, h_r \in \mathbb{F}_p^n$ and a map $\Gamma : \mathbb{F}_p^r \to \mathbb{F}_p$ such that*

$$P(x) \equiv \Gamma\big(\Delta_{h_1} P(x), \ldots, \Delta_{h_r} P(x)\big).$$

5.1. **The analytic rank of polynomial maps.** Inspired by Gowers and Wolf's notion of analytic rank for multilinear forms and polynomials [GW11], we introduce a new notion of rank for higher-dimensional polynomial maps which we also call the *analytic rank*. This notion will be crucial in our proof of Theorem 3.1; intuitively, it measures how well a given polynomial map ϕ can be approximated by lower-degree maps.

For integers $d, n, k \geq 1$, we denote by $\mathrm{Pol}_{\leq d}(\mathbb{F}_p^n, \mathbb{F}_p^k)$ the space of all polynomial maps $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ of degree at most d.

Definition 5.2 (Analytic rank)**.** Given a polynomial map $\phi \in \mathrm{Pol}_{\leq d}(\mathbb{F}_p^n, \mathbb{F}_p^k)$, we define its analytic rank $\mathrm{arank}_d(\phi)$ by

$$\mathrm{arank}_d(\phi) = -\log_p \left(\max_{\psi : \mathbb{F}_p^n \to \mathbb{F}_p^k,\, \deg(\psi) < d} \Pr_{x \in \mathbb{F}_p^n} \big[\phi(x) = \psi(x)\big] \right).$$

Note that, for affine-linear maps $\phi \in \mathrm{Pol}_{\leq 1}(\mathbb{F}_p^n, \mathbb{F}_p^k)$, this definition coincides with the usual notion of rank for the matrix $U \in \mathbb{F}_p^{k \times n}$ encoding its linear part. Indeed, write $\phi(x) = Ux + v$ for some $v \in \mathbb{F}_p^k$. Since Ux is uniformly distributed over $\mathrm{im}(U) \simeq \mathbb{F}_p^{\mathrm{rk}(U)}$ when x is uniformly distributed over $\mathbb{F}_p^k$, we have that

$$\Pr_{x \in \mathbb{F}_p^n} \big[Ux + v = w\big] = \begin{cases} p^{-\mathrm{rk}(U)} & \text{if } w - v \in \mathrm{im}(U), \\ 0 & \text{if } w - v \notin \mathrm{im}(U). \end{cases}$$

This might help explain the reason for the $-\log_p$ in the definition of analytic rank, as well as the need to maximize the probability of agreement over all lower-degree maps.

Another useful way of viewing the analytic rank of a polynomial map ϕ is as a measure of how well-equidistributed its values are in $\mathbb{F}_p^k$, up to lower-degree perturbations. Indeed, we can equivalently write

$$\operatorname{arank}_d(\phi) = \min_{\psi:\mathbb{F}_p^n\to\mathbb{F}_p^k,\, \deg(\psi)<d} -\log_p \left(\mathbb{E}_{v\in\mathbb{F}_p^k, x\in\mathbb{F}_p^n} \omega^{\langle v,\, \phi(x)-\psi(x)\rangle}\right).$$

The expectation inside the logarithm above is analogous to the notion of bias (4) given before, and can be seen as an analytic measure of how close to uniformly distributed over $\mathbb{F}_p^k$ the values taken by $\phi - \psi$ are.

It is clear from the definition that the function arank_d is non-negative (since probabilities are bounded by 1), and that $\operatorname{arank}_d(\phi) = 0$ if and only if $\deg(\phi) \leq d - 1$. It also satisfies several useful properties in common with the rank of matrices; in order to state them we will need some notation for considering coordinate restrictions:

Definition 5.3 (Restriction)**.** For a polynomial map $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ and subset $I \subseteq [n]$, we define the restriction $\phi_{|I} : \mathbb{F}_p^I \to \mathbb{F}_p^k$ to be the map given by $\phi_{|I}(y) = \phi(\bar{y})$, where $\bar{y} \in \mathbb{F}_p^n$ agrees with y on the coordinates in I and is zero elsewhere.

The properties of analytic rank which will be important to us are summarized in the next lemma. Those rank functions for polynomial maps which satisfy all these properties are called *natural rank functions* in [BCS22].

Lemma 5.4 (Properties of analytic rank)**.** *For all integers $d, n, k \geq 1$, the analytic rank function* arank_d *satisfies:*

(1) *Symmetry:*
$\operatorname{arank}_d(\phi) = \operatorname{arank}_d(-\phi)$ *for all* $\phi \in \operatorname{Pol}_{\leq d}(\mathbb{F}^n, \mathbb{F}^k)$.

(2) *Sub-additivity:*
$\operatorname{arank}_d(\phi + \gamma) \leq \operatorname{arank}_d(\phi) + \operatorname{arank}_d(\gamma)$ *for all* $\phi, \gamma \in \operatorname{Pol}_{\leq d}(\mathbb{F}^n, \mathbb{F}^k)$.

(3) *Monotonicity under restrictions:*
$\operatorname{arank}_d(\phi_{|I}) \leq \operatorname{arank}_d(\phi)$ *for all* $\phi \in \operatorname{Pol}_{\leq d}(\mathbb{F}^n, \mathbb{F}^k)$ *and all sets* $I \subseteq [n]$.

(4) *Restriction Lipschitz property:*
$\operatorname{arank}_d(\phi_{|I\cup J}) \leq \operatorname{arank}_d(\phi_{|I}) + |J|$ *for all* $\phi \in \operatorname{Pol}_{\leq d}(\mathbb{F}^n, \mathbb{F}^k)$ *and all sets* $I, J \subseteq [n]$.

Proof: The first property is trivial. To prove property (2), let $\psi, \chi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ be polynomial maps of degree at most $d - 1$ such that

$$\begin{aligned}
\operatorname{arank}_d(\phi) &= -\log_p \Pr_{x\in\mathbb{F}_p^n}\big[\phi(x) = \psi(x)\big],\\
\operatorname{arank}_d(\gamma) &= -\log_p \Pr_{x\in\mathbb{F}_p^n}\big[\gamma(x) = \chi(x)\big].
\end{aligned}$$

Then $p^{-\operatorname{arank}_d(\phi)-\operatorname{arank}_d(\gamma)}$ can be expressed as

$$\begin{aligned}&\Pr_{x,y\in\mathbb{F}_p^n}\big[\phi(x)=\psi(x)\ \ \&\ \ \gamma(y)=\chi(y)\big]\\ &\qquad\qquad=\Pr_{x,y}\big[\phi(x)=\psi(x)\ \ \&\ \ \phi(x)+\gamma(x+y)=\psi(x)+\chi(x+y)\big],\end{aligned}$$

where we performed the change of variables $(x,y)\mapsto(x,x+y)$. Since

$$\gamma(x+y)=\gamma(x)+\Delta_y\gamma(x),$$

this equals

$$\begin{aligned}&\Pr_{x,y}\big[\phi(x)=\psi(x)\ \ \&\ \ \phi(x)+\gamma(x)=\psi(x)+\chi(x+y)-\Delta_y\gamma(x)\big]\\ &\qquad\qquad\le\Pr_{x,y}\big[\phi(x)+\gamma(x)=\psi(x)+\chi(x+y)-\Delta_y\gamma(x)\big].\end{aligned}$$

Note that, for any fixed $y\in\mathbb{F}_p^n$, the function

$$x\mapsto\psi(x)+\chi(x+y)-\Delta_y\gamma(x)$$

is a polynomial map of degree at most $d-1$. The last probability above is then bounded by

$$\begin{aligned}&\max_y\Pr_x\big[\phi(x)+\gamma(x)=\psi(x)+\chi(x+y)-\Delta_y\gamma(x)\big]\\ &\qquad\qquad\le\max_{\zeta:\mathbb{F}_p^n\to\mathbb{F}_p^k,\,\deg(\zeta)<d}\Pr_x\big[\phi(x)+\gamma(x)=\zeta(x)\big]\\ &\qquad\qquad=p^{-\operatorname{arank}_d(\phi+\gamma)}.\end{aligned}$$

Sub-additivity now follows by taking logarithms.

To prove property (3) it suffices to show that $\operatorname{arank}_d(\phi_{|[n]\setminus\{i\}})\le\operatorname{arank}_d(\phi)$ for any $i\in[n]$, which can then be applied iteratively. Assume for notational convenience that $i=n$, and let $\psi:\mathbb{F}_p^n\to\mathbb{F}_p^k$ be a polynomial map of degree at most $d-1$ which satisfies

$$p^{-\operatorname{arank}_d(\phi)}=\Pr_{x\in\mathbb{F}_p^n}[\phi(x)=\psi(x)].$$

Factoring out the variable x_n allows us to write the probability on the right-hand side as

$$\mathbb{E}_{x_n\in\mathbb{F}_p}\Pr_{y\in\mathbb{F}_p^{n-1}}\big[\phi_{|[n-1]}(y)+\phi'(y,x_n)x_n=\psi_{|[n-1]}(y)+\psi'(y,x_n)x_n\big],$$

where ϕ' and ψ' are some polynomial maps of degree at most $d-1$. By the averaging principle, this is at most

$$\begin{aligned}&\max_{x_n\in\mathbb{F}_p}\Pr_{y\in\mathbb{F}_p^{n-1}}\big[\phi_{|[n-1]}(y)=\psi_{|[n-1]}(y)+\psi'(y,x_n)x_n-\phi'(y,x_n)x_n\big]\\ &\qquad\le\max_{\zeta:\mathbb{F}_p^{n-1}\to\mathbb{F}_p^k,\,\deg(\zeta)<d}\Pr_{y\in\mathbb{F}_p^{n-1}}\big[\phi_{|[n-1]}(y)=\psi_{|[n-1]}(y)+\zeta(y)\big]\\ &\qquad=p^{-\operatorname{arank}_d(\phi_{|[n-1]})},\end{aligned}$$

showing that $\operatorname{arank}_d(\phi_{|[n-1]})\le\operatorname{arank}_d(\phi)$ as wished.

Finally, for the Lipschitz property (4), let $\psi : \mathbb{F}_p^I \to \mathbb{F}_p^k$ be a map with $\deg(\psi) < d$ maximizing the agreement probability $\Pr_{x\in\mathbb{F}_p^I}\big[\phi_I(x) = \psi(x)\big]$, and suppose without loss of generality that $J \cap I = \emptyset$. Then

$$\begin{aligned}
p^{-\operatorname{arank}_d(\phi_{|I\cup J})} &\geq \Pr_{x\in\mathbb{F}_p^I,\, y\in\mathbb{F}_p^J}\big[\phi_{|I\cup J}(x,y) = \psi(x)\big] \\
&\geq \Pr_{x\in\mathbb{F}_p^I,\, y\in\mathbb{F}_p^J}\big[\phi_{|I\cup J}(x,0) = \psi(x) \ \ \& \ \ y = 0\big] \\
&= p^{-|J|}\Pr_{x\in\mathbb{F}_p^I}\big[\phi_I(x) = \psi(x)\big] \\
&= p^{-\operatorname{arank}_d(\phi_I)-|J|},
\end{aligned}$$

and the restriction Lipschitz property follows. □

5.2. **Biased equidistribution of high-rank maps.** As in the degree-1 case considered in Section 4.1, we will need to study the distribution of values $\phi(Z)$ taken by a polynomial map ϕ when the input is a ρ-biased random variable $Z \sim \mathcal{N}_\rho(y)$. This can be done by considering restrictions of ϕ to random subsets of variables, which model the coordinates "corrupted" by the studied random process.

Motivated by this problem, the behavior of rank functions under random coordinate restrictions was studied in detail by the first and third authors [BCS22]. In the nomenclature of that paper, Lemma 5.4 shows that the analytic rank is a natural rank function. Applying [BCS22, Theorem 1.8] we then immediately obtain the following result, which shows that random restrictions of a high-rank polynomial map will also have high rank with high probability. (Recall that $I \sim [n]_\sigma$ denotes the random process of sampling a subset $I \subseteq [n]$ where each $i \in [n]$ belongs to I with probability σ, all events being mutually independent.)

Theorem 5.5 (Random restriction theorem)**.** *For every $d \in \mathbb{N}$ and $\sigma, \varepsilon \in (0,1]$, there exist $\kappa = \kappa(d,\sigma) > 0$ and $R = R(d,\sigma,\varepsilon) \in \mathbb{N}$ such that the following holds. For every map $\phi \in \operatorname{Pol}_{\leq d}(\mathbb{F}^n, \mathbb{F}^k)$ with $\operatorname{arank}_d(\phi) \geq R$, we have that*

$$\Pr_{I\sim[n]_\sigma}\big[\operatorname{arank}_d(\phi_{|I}) \geq \kappa \cdot \operatorname{arank}_d(\phi)\big] \geq 1-\varepsilon.$$

With the help of this theorem, it is easy to show that high-rank polynomial maps are approximately equidistributed even under biased inputs:

Lemma 5.6 (Biased equidistribution lemma)**.** *For every $d \in \mathbb{N}$ and $\rho, \varepsilon \in (0,1)$ there exists a constant $R_0 = R_0(d,\rho,\varepsilon) > 0$ such that the following holds. If $\phi \in \operatorname{Pol}_{\leq d}(\mathbb{F}^n, \mathbb{F}^k)$ satisfies $\operatorname{arank}_d(\phi) \geq R_0$, then*

$$\Pr_{Z\sim\mathcal{N}_\rho(0)}\big[\phi(y+Z) = w\big] \leq \varepsilon \quad \textit{for all } y \in \mathbb{F}_p^n, w \in \mathbb{F}_p^k.$$

Proof: It suffices to prove the special case where both y and w are zero, that is

$$\Pr_{Z\sim\mathcal{N}_\rho(0)}\big[\phi(Z)=0\big]\leq\varepsilon.$$

Indeed, for fixed $y\in\mathbb{F}_p^n$ and $w\in\mathbb{F}_p^k$, the map $\tilde{\phi}: x\mapsto\phi(y+x)-w$ has the same degree and same analytic rank as ϕ, and satisfies $\tilde{\phi}(x)=0$ if and only if $\phi(y+x)=w$.

We can sample $Z\sim\mathcal{N}_\rho(0)$ by first sampling $I\sim[n]_{1-\rho}$ (the "corrupted coordinates"), then sampling z uniformly from $\mathbb{F}_p^I$ (the "noise") and setting $Z_{|I}=z$, $Z_{|[n]\setminus I}=0$; thus

$$\begin{aligned}\Pr_{Z\sim\mathcal{N}_\rho(0)}\big[\phi(Z)=0\big]&=\mathbb{E}_{I\sim[n]_{1-\rho}}\Pr_{z\in\mathbb{F}_p^I}\big[\phi_{|I}(z)=0\big]\\&\leq\mathbb{E}_{I\sim[n]_{1-\rho}}p^{-\operatorname{arank}_d(\phi_{|I})}.\end{aligned}$$

Let $R=R(d,1-\rho,\varepsilon/2)$ and $\kappa=\kappa(d,1-\rho)$ be the constants guaranteed by Theorem 5.5. If $\operatorname{arank}_d(\phi)\geq R$, from that result we obtain

$$\mathbb{E}_{I\sim[n]_{1-\rho}}p^{-\operatorname{arank}_d(\phi_{|I})}\leq\varepsilon/2+p^{-\kappa\cdot\operatorname{arank}_d(\phi)}.$$

Taking[9] $R_0=\max\big\{R,\,\log_p(2/\varepsilon)/\kappa\big\}$ we conclude that

$$\Pr_{Z\sim\mathcal{N}_\rho(0)}\big[\phi(Z)=0\big]\leq\mathbb{E}_{I\sim[n]_{1-\rho}}p^{-\operatorname{arank}_d(\phi_{|I})}\leq\varepsilon$$

whenever $\operatorname{arank}_d(\phi)\geq R_0$, as wished. □

5.3. **The proof of Theorem 3.1.** We are now ready to present the proof of Theorem 3.1, which proceeds by induction on the degree d. For degree-1 maps the result was already proven in the warm-up section,[10] so let $d\geq 2$ and assume the result holds for maps of degree at most $d-1$.

As was done in the base case, we will divide the argument into two parts, corresponding to whether the analytic rank of ϕ is "high" (the pseudorandom case) or "low" (the structured case). The pseudorandom case immediately follows from Lemma 5.6, the biased equidistribution lemma: let $R_0=R_0(d,\rho,\varepsilon)$ be the constant guaranteed by that lemma, and suppose that $\operatorname{arank}_d(\phi)>R_0$. Then for every $x\in\mathbb{F}_p^k$ we have that

$$\Pr_{Z\sim\mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big)=x\big]\leq\varepsilon,$$

and we conclude by averaging over all such x.

[9] Note that this bound is non-increasing on the value of p, so we can obtain a field-independent bound by considering the smallest case $p=2$.

[10] It would also be possible to start the induction from the trivial base case $d=0$ of constant maps, but we thought it more instructive to first present the argument for degree-1 maps in order to gain some intuition.

Now suppose that $\operatorname{arank}_d(\phi) \leq R_0$, and let $\psi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ be a map of degree at most $d-1$ such that $\Pr_{x\in\mathbb{F}_p^n}\big[\phi(x) = \psi(x)\big] \geq p^{-R_0}$. Denote $\tilde{\phi} = \phi - \psi$ for convenience, and let $P \in \mathbb{F}_p[y_1, \dots, y_n, v_1, \dots, v_k]$ be the polynomial given by

$$P(y, v) = \langle v, \tilde{\phi}(y)\rangle.$$

This polynomial has non-negligible bias:

$$\operatorname{bias}(P) = \mathbb{E}_{y\in\mathbb{F}_p^n}\mathbb{E}_{v\in\mathbb{F}_p^k}\omega^{\langle v,\tilde{\phi}(y)\rangle} = \mathbb{E}_{y\in\mathbb{F}_p^n}\mathbf{1}\big[\tilde{\phi}(y) = 0\big] \geq p^{-R_0},$$

where $\omega = e^{2\pi i/p}$. By Theorem 5.1, there exist $s = s(p, d, R_0) \in \mathbb{N}$, pairs $(h_1, w_1), \dots, (h_s, w_s) \in \mathbb{F}_p^n \times \mathbb{F}_p^k$ and a map $\Gamma : \mathbb{F}_p^s \to \mathbb{F}_p$ such that

$$P(y, v) = \Gamma\big(\Delta_{(h_1,w_1)}P(y, v), \dots, \Delta_{(h_s,w_s)}P(y, v)\big).$$

Let $f : \mathbb{F}_p^s \to \mathbb{C}$ be the map given by $f(t) = \omega^{\Gamma(t)}$ and let $\widehat{f} : \mathbb{F}_p^s \to \mathbb{C}$ be its Fourier transform,

$$\widehat{f}(\alpha) = \mathbb{E}_{t\in\mathbb{F}_p^s} f(t)\omega^{-\langle\alpha,t\rangle}.$$

Since P is linear in the last k coordinates, it follows that

$$\begin{aligned}\Delta_{(h,w)}P(y, v) &= P(y + h, v + w) - P(y + h, v) + P(y + h, v) - P(y, v)\\ &= \langle w, \tilde{\phi}(y + h)\rangle + \langle v, \Delta_h\tilde{\phi}(y)\rangle.\end{aligned}$$

By the Fourier inversion formula, we conclude that

$$\begin{aligned}\omega^{P(y,v)} &= f\big(\Delta_{(h_1,w_1)}P(y, v), \dots, \Delta_{(h_s,w_s)}P(y, v)\big)\\ &= \sum_{\alpha\in\mathbb{F}_p^s}\widehat{f}(\alpha)\omega^{Q_\alpha(y)+\langle v,\gamma_\alpha(y)\rangle},\end{aligned}$$

where for $\alpha \in \mathbb{F}_p^s$ we denote

$$\begin{aligned}Q_\alpha(y) &= \sum_{i=1}^{s}\langle\alpha_i w_i, \tilde{\phi}(y + h_i)\rangle,\\ \gamma_\alpha(y) &= \sum_{i=1}^{s}\alpha_i\Delta_{h_i}\tilde{\phi}(y).\end{aligned}$$

Note crucially that $\deg(\gamma_\alpha) \leq d-1$ for all $\alpha \in \mathbb{F}_p^s$, which is what will eventually allow us to apply the induction hypothesis.

It follows from our expression for $\omega^{P(y,v)}$ that

$$\begin{aligned}\mathbf{1}[\phi(y)=x] &= \mathbb{E}_{v\in\mathbb{F}_p^k}\omega^{\langle v,\phi(y)-x\rangle}\\ &= \mathbb{E}_{v\in\mathbb{F}_p^k}\omega^{P(y,v)+\langle v,\psi(y)-x\rangle}\\ &= \sum_{\alpha\in\mathbb{F}_p^s}\widehat{f}(\alpha)\omega^{Q_\alpha(y)}\,\mathbb{E}_{v\in\mathbb{F}_p^k}\omega^{\langle v,\,(\gamma_\alpha+\psi)(y)-x\rangle}.\end{aligned}$$

Taking $y=E(x)+Z$, we then obtain

$$\begin{aligned}\Pr\big[\phi\big(E(x)+Z\big)=x\big] &= \mathbb{E}_{x,Z}\mathbf{1}\big[\phi\big(E(x)+Z\big)=x\big]\\ &\le \sum_{\alpha\in\mathbb{F}_p^s}|\widehat{f}(\alpha)|\,\mathbb{E}_{x,Z}\Big|\mathbb{E}_{v\in\mathbb{F}_p^k}\omega^{\langle v,\,(\gamma_\alpha+\psi)(E(x)+Z)-x\rangle}\Big|\\ &\le \Bigg(\sum_{\alpha\in\mathbb{F}_p^s}|\widehat{f}(\alpha)|\Bigg)\max_{\alpha\in\mathbb{F}_p^s}\mathbb{E}_{x,Z}\mathbf{1}\big[(\gamma_\alpha+\psi)\big(E(x)+Z\big)=x\big]\\ &\le p^{s/2}\max_{\alpha\in\mathbb{F}_p^s}\Pr\big[(\gamma_\alpha+\psi)\big(E(x)+Z\big)=x\big],\end{aligned}$$

where we have used the Cauchy-Schwarz inequality and Parseval's identity in the last line. Since $\deg(\gamma_\alpha+\psi)\le d-1$ and s ultimately depends only on p, d, ρ and ε, by taking

$$k\ge k_0(p,d,\rho,\varepsilon):=k_0(p,d-1,\rho,\varepsilon\,p^{-s/2})$$

we conclude from the induction hypothesis that

$$\Pr\big[\phi\big(E(x)+Z\big)=x\big]\le\varepsilon$$

in this case as well. The theorem follows.

6. Quantum circuit for decoding the Hadamard code

In this section we give the constant-depth quantum circuit to decode the Hadamard code. We first give the operations on a high level, then we present more details on the implementations and finally we show how to further reduce the total complexity.

6.1. High-level quantum algorithm.

Interestingly, we can implement each of the operations performed by the entangled players described in Section 4.2 with constant-depth quantum circuits using only single and two-qubit gates and classical parity gates, thus giving a $\mathrm{QNC}^0[\oplus]$ circuit.

To generate GHZ states (which we will need on multiple occasions), we use a technique of Watts et al. [BWKST19] that starts by generating a so-called *poor man's cat state*: $\frac{1}{\sqrt{2}}(|z\rangle+|\bar{z}\rangle)$ for some binary vector z. To generate an n-qubit poor man's cat state, we apply Hadamard gates to an n-qubit register initialized in the all-zeros state and then compute the parity between adjacent qubits in $n-1$ ancilla qubits. After measuring the ancilla qubits, we are left with a poor man's cat state. With the parity measurements d_i we can correct the poor man's cat state to a GHZ state by flipping qubits conditioned on a prefix-sum computation of the measurement outcomes.

To apply the conditional phase-flip (3), we use an ancilla qubit and compute the AND function

$$|z\rangle\,|b\rangle \quad \mapsto \quad |z\rangle\,|\mathrm{AND}(z_1,\dots,z_k)\oplus b\rangle\,.$$

We then apply a phase-flip on the last qubit conditioned on $c(y)=1$. Using X-gates, we can ensure that AND evaluates to 1 if and only if $z=y$. For ease of implementation, we use the identity $\mathrm{AND}(z_1,\dots,z_k)=\neg\mathrm{OR}(\neg z_1,\dots,\neg z_k)$.

We cannot use standard decomposition techniques, such as using Toffoli gates, to implement the OR, as these do not give constant-depth circuits. Instead, we use the constant-depth Exact OR-implementation of [TT13]. Their method uses single and two-qubit gates and additionally makes use of quantum fan-out gates that implement the general map

$$|b\rangle\,|y_1\rangle\dots|y_n\rangle \mapsto |b\rangle\,|y_1\oplus b\rangle\dots|y_n\oplus b\rangle\,.$$

Below, we show how to implement these quantum fan-out gates with $\mathrm{QNC}^0[\oplus]$ circuits. These quantum fan-out gates compute parity when conjugated with Hadamard gates on each input. We use the quantum fan-out gates to compute and sum the parity of each subset of the inputs to compute the OR of the inputs in the first qubit [TT13, Lemma 1].

Implementing this constant-depth approach does come at the cost of a circuit size exponential in k: $O(k2^k)=O(n\log n)$. We can implement this for each input in parallel which gives a quantum circuit of size $O(n^2\log n)$. As we have a constant number of constant-depth operations, we have constant-depth circuit for list-decoding the Hadamard code.

6.2. **Details of quantum algorithm.** Now, we present details of the operations given in the previous section.

Figure 1 shows the quantum circuit to generate a 3-qubit GHZ state. If the measurement results were $d_1=1$ and $d_2=1$, then we have the state $\frac{1}{\sqrt{2}}(|010\rangle+|101\rangle)$. We flip the second qubit, because the first measurement result, d_1, equals 1. We do not flip the third qubit, because the parity of the measurement results, $d_1\oplus d_2$, equals 0. This indeed gives the 3-qubit GHZ state $\frac{1}{\sqrt{2}}(|000\rangle+|111\rangle)$ as desired.

This method extends naturally to larger n. A prefix-sum computation is then used to determine which qubits have to be flipped. The depth of the circuit does not increase with larger n. In our GHZ state construction, we

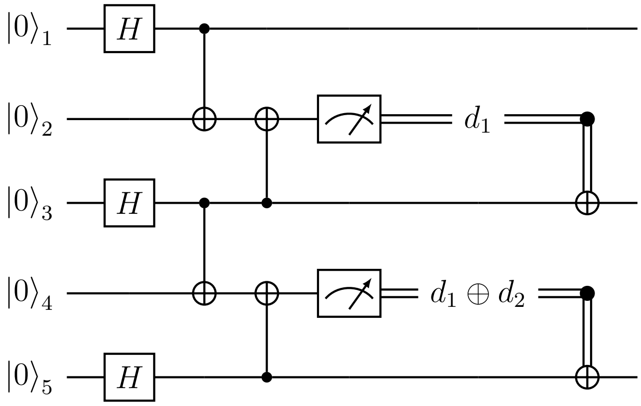

FIGURE 1. The quantum circuit to generate a 3-qubit GHZ state. First we obtain a poor man's cat state $\frac{1}{\sqrt{2}}(|z\rangle + |\bar{z}\rangle)$ with each $z \in \mathbb{F}_2^3$ equally likely to be found. The parity gates compute a prefix sum on the measurement results d_1 and d_2 and determine if a qubit has to be flipped to obtain the GHZ state.

implicitly assumed the qubits to be arranged in a linear architecture. Other arrangements work equally well, as shown in [BWKST19].

A powerful tool in the implementation above is the quantum fan-out gate. We now show how to implement this gate using only single and two-qubit gates and classical parity gates. For this, we combine the GHZ state construction introduced above with ideas from distributed quantum computing, specifically, the non-local CNOT-gate [EJPP00, YL04]. The term non-local CNOT originates from the fact that we can imagine the control and target qubits being hosted on different quantum devices which share a GHZ state. We apply this gate in a local setting to construct the quantum fan-out gate in constant-depth.

Figure 2 shows the quantum fan-out gate implementation for one control and two targets. This method extends to an arbitrary number of targets by reapplying the same operations to all target qubits and the corresponding qubits in the GHZ state in parallel. The last Z-gate is only applied if the parity over all measurement results equals 1.

Lemma 6.1. *The circuit of Figure 2 implements a quantum fan-out gate.*

Proof: Let $|x\rangle$ be an n-qubit computational basis state and $|\phi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ be any single qubit quantum state. We will prove that the circuit implements the quantum fan-out gate on the state $|\phi\rangle\,|x\rangle$. The lemma then follows by linearity of the operations.

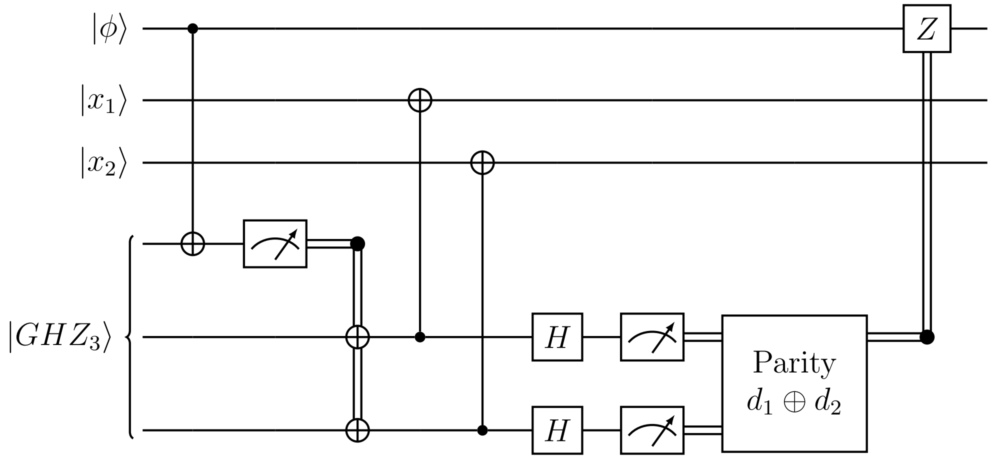

FIGURE 2. Implementation of a quantum fan-out gate with one control qubit $|\phi\rangle$ and two target qubits $|x_1\rangle$ and $|x_2\rangle$. Only single and two-qubit gates and classical parity gates are used. The bottom three qubits are in the GHZ$_3$ state.

The action of the quantum fan-out gate on the quantum state is given by

$$|\phi\rangle\,|x\rangle \overset{\text{fan-out}}{\mapsto} \alpha\,|0\rangle\,|x\rangle + \beta\,|1\rangle\,X^{\otimes n}\,|x\rangle = \alpha\,|0\rangle\,|x\rangle + \beta\,|1\rangle\,|\bar{x}\rangle\,, \tag{5}$$

where $|\bar{x}\rangle$ is the computational basis state $|x\rangle$ with all qubits flipped.

To see why this works, assume we have a GHZ$_{n+1}$ state and we apply the operations as shown in Figure 2 generalized to arbitrary n. Up to a normalization factor of $1/\sqrt{2}$ coming from the GHZ state we then have:

$$\begin{aligned}
&\big[\alpha\,|0\rangle + \beta\,|1\rangle\big]\,|x\rangle \otimes \big[\,|00\cdots 0\rangle + |11\cdots 1\rangle\,\big] \\
&\quad\xmapsto{(1)} \alpha\,|0\rangle\,|x\rangle \otimes \big[\,|00\cdots 0\rangle + |11\cdots 1\rangle\,\big] + \beta\,|1\rangle\,|x\rangle \otimes \big[\,|10\cdots 0\rangle + |01\cdots 1\rangle\,\big] \\
&\quad\xmapsto{(2)} \alpha\,|0\rangle\,|x\rangle\,|d_0 0\cdots 0\rangle + \beta\,|1\rangle\,|x\rangle\,|d_0 1\cdots 1\rangle \\
&\quad\xmapsto{(3)} \alpha\,|0\rangle\,|x\rangle\,|d_0 0\cdots 0\rangle + \beta\,|1\rangle\,X^{\otimes n}\,|x\rangle\,|d_0 1\cdots 1\rangle \\
&\quad\xmapsto{(4)} \frac{1}{2^{n-1}}\sum_{d\in\mathbb{F}_2^n}\big[\alpha\,|0\rangle\,|x\rangle + \beta(-1)^{d_1+\ldots+d_n}\,|1\rangle\,X^{\otimes n}\,|x\rangle\,\big]\,|d_0 d_1\ldots d_n\rangle \\
&\quad\xmapsto{(5)} \alpha\,|0\rangle\,|x\rangle\,|d_0 d_1\ldots d_n\rangle + (-1)^{d_1+\ldots+d_n}\beta\,|1\rangle\,X^{\otimes n}\,|x\rangle\,|d_0 d_1\ldots d_n\rangle \\
&\quad\xmapsto{(6)} \big[\alpha\,|0\rangle\,|x\rangle + \beta\,|1\rangle\,|\bar{x}\rangle\,\big]\,|d_0 d_1\ldots d_n\rangle\,.
\end{aligned}$$

In Step (1), we perform a CNOT operation from the control qubit to the first qubit of the GHZ state. In Step (2), we measure that qubit, with outcome d_0, and apply an X-gate to the remaining n qubits of the GHZ state if $d_0 = 1$. Next we perform CNOT gates between the $i+1$-th qubit of the GHZ state and the i-th target qubit. In Steps (4) and (5), we first apply Hadamard gates to each unmeasured qubit of the GHZ state and subsequently measure it. Finally, we compute the parity $d_1 \oplus \ldots \oplus d_n$ and apply a Z-gate to the control qubit if this parity equals one. This indeed gave the desired final state and hence the shown quantum circuit implements the quantum fan-out gate. □

6.3. **Reducing the algorithm's complexity.** We now show how to reduce the complexity of the quantum algorithm from $O(n^2 \log n)$ to $O(n \log n \log\log n)$.

With the current implementation, the complexity of applying a single conditional phase-flip is polynomial in the codeword length n. We can reduce this to polynomial in the message length k. For this we apply the OR-reduction [HŠ05]. Instead of evaluating an OR on k inputs, we use a $O(k \log k)$ size constant depth circuit to prepare a quantum state on $\lceil \log(k+1) \rceil$ qubits, such that the OR on these $\lceil \log(k+1) \rceil$ qubits evaluates to the same value as the OR on the original k qubits. This OR-reduction uses the quantum fan-out gate.

The OR-reduction increases the depth by an additive constant, however, it reduces the complexity of applying a single conditional phase-flip to $O(k \log k)$. The complexity of the quantum circuit therefore reduces to $O(n \log n \log\log n)$, using $k = \log n$.

Note that the phase flip is only applied if the input is one, hence, we have to apply the OR-reduction and the Exact OR implementation only if this input is one.

7. The high-characteristic setting

In this section we give some evidence to support our conjecture (made in Section 2.1) that the probability of correct message retrieval by $\mathrm{NC}^0[\oplus]$ circuits decays exponentially with the message length. This is done by proving the following theorem, which a "high-characteristic" analogue of Theorem 3.1 with much better bounds; as we see no reason to believe a result of this kind has a strong dependence on the characteristic of the finite field considered,[11] we believe that a similar bound also holds for low-characteristic fields such as $\mathbb{F}_2$.

Theorem 7.1 (Exponential decay in high characteristic)**.** *For every $d \in \mathbb{N}$ and $\rho \in [0, 1)$ there exist constants $C = C(\rho, d)$ and $c = c(\rho, d) > 0$ such that the following holds. Let $p > d$ be a prime, and let n, k be integers with $k \geq p$.*

[11] While the result is stated in the setting of prime fields $\mathbb{F}_p$, it easily generalizes to the case of non-prime finite fields $\mathbb{F}_q$, with only minor modifications in the proof.

Then for every polynomial map $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ of degree at most d and every function $E : \mathbb{F}_p^k \to \mathbb{F}_p^n$ we have

$$\Pr_{x\in\mathbb{F}_p^k, Z\sim\mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big)=x\big] \leq Ce^{-ck/(\log k)^{d^2}}.$$

Remark 7.2. The presence of the poly-logarithmic term in the exponential above is due to a poly-logarithmic loss when passing between two distinct notions of tensor rank in our proof of Theorem 7.1. It is a widely-believed conjecture in additive combinatorics that these two notions of rank (see Section 7.1 below) are within a constant multiplicative factor of one another, in which case our proof would give an upper bound of the form Ce^{-ck} for the probability of correct message retrieval (and this would be the best possible).

As with the proof of Theorem 3.1, we will prove Theorem 7.1 by induction on the degree d, using the degree-1 case shown in Section 4.1 as the base case of the induction. The inductive argument will also share many similarities with the one presented in Section 5, in particular relying on a structure-versus-randomness dichotomy based on a notion of rank associated with the polynomial map ϕ. The reason for the better bounds we obtain now stems from the fact that, in the high-characteristic case, one can work with tensors (i.e. multilinear forms) rather than with general polynomial maps. In the quasirandom case of our argument, we can then use a stronger version of the random restriction theorem for the analytic rank of tensors (also proved in [BCS22]), while in the structured case we use a recently-proved close connection between analytic rank and partition rank of tensors [MZ22].

7.1. **Tensors associated to polynomial maps.** Given a polynomial map $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ of degree at most d, we can define a $(d+1)$-tensor $T : (\mathbb{F}_p^n)^d \times \mathbb{F}_p^k \to \mathbb{F}_p$ associated to it by

$$T(y_1,\dots,y_d,v) = \big\langle v,\, \Delta_{y_1}\cdots\Delta_{y_d}\phi(0)\big\rangle,$$

where we recall that $\Delta_y\phi(x) = \phi(x+y)-\phi(x)$. While not immediately obvious, the formula above indeed defines a tensor (i.e., it is linear in each variable separately). This follows from the fact that $\Delta_{y_1}\cdots\Delta_{y_d}\phi$ does not depend on the order of the derivatives, and that the polynomial map $\Delta_{y_1}\cdots\Delta_{y_{d-1}}\phi$ has degree at most 1 (since ϕ has degree at most d); note that, if ψ is a linear map, then $h \mapsto \Delta_h\psi$ is linear in h.

If the characteristic p of the field in strictly higher than the degree d, then we also have the *integration formula*

$$\phi(y) = \frac{1}{d!}T(y,\dots,y,\cdot) + \psi(y) \quad \text{for all } y \in \mathbb{F}_p^n,$$

where y is repeated d times inside T and ψ is a polynomial map of degree at most $d-1$. This follows from the (discrete) Taylor expansion theorem, and allows us to pass back and forth between tensors and polynomial maps.

We will use the following two notions of rank for tensors, originally introduced by Gowers and Wolf [GW11] and by Naslund [Nas20], respectively.

Definition 7.3 (Tensor analytic rank)**.** Let $X_1, \dots, X_r$ be finite sets and $T : \mathbb{F}_p^{X_1} \times \cdots \times \mathbb{F}_p^{X_r} \to \mathbb{F}_p$ be an r-tensor. The bias of T is defined as

$$\operatorname{bias}(T) = \mathbb{E}_{x_1 \in \mathbb{F}_p^{X_1}, \dots, x_r \in \mathbb{F}_p^{X_r}} \omega^{T(x_1, \dots, x_r)},$$

where $\omega = e^{2\pi i/p}$. The bias is always real and positive,[12] and the analytic rank of T is defined by

$$\operatorname{arank}(T) = -\log_p \operatorname{bias}(T).$$

Definition 7.4 (Partition rank)**.** A nonzero r-tensor $T : \mathbb{F}_p^{X_1} \times \cdots \times \mathbb{F}_p^{X_r} \to \mathbb{F}_p$ is said to have partition rank 1 if there is a nonempty strict subset $I \subset [r]$ and tensors $S : \prod_{i \in I} \mathbb{F}^{X_i} \to \mathbb{F}$ and $R : \prod_{i \in [r] \setminus I} \mathbb{F}^{X_i} \to \mathbb{F}$ such that T can be factored as $T = SR$. The partition rank of T, denoted prank(T), is defined as the least $m \in \mathbb{N}$ such that there is a decomposition $T = T_1 + \cdots + T_m$ where each T_i has partition rank 1.

While these two notion of rank are defined in very different ways, it turns out that they are intimately related to one another. Lovett has shown that $\operatorname{arank}(T) \le \operatorname{prank}(T)$ holds for all tensors [Lov19], and it is a well-known open problem to determine whether a similar inequality holds in the converse direction, up to an absolute multiplicative factor. Very recently, Moshkovitz and Zhu [MZ22] proved that the relation between these two rank functions is at worst quasilinear.

Theorem 7.5 (Moshkovitz–Zhu)**.** *For every $r \ge 2$ there exists $L_r > 0$ such that for every r-tensor T over any finite field, we have*

$$\operatorname{arank}(T) \le \operatorname{prank}(T) \le L_r \operatorname{arank}(T) \log^{r-1}\big(1 + \operatorname{arank}(T)\big). \tag{6}$$

This result will be an important ingredient in our proof of Theorem 7.1; we note that the decay obtained could be improved to Ce^{-ck} if Theorem 7.5 were proven without the poly-logarithmic factor on the right-hand side of (6). Another important ingredient is the following random restriction theorem for tensors [BCS22], stated here for the special case of the analytic rank.

[12] It is not hard to show that $\operatorname{bias}(T) = \Pr_{x_1 \in \mathbb{F}_p^{X_1}, \dots, x_{r-1} \in \mathbb{F}_p^{X_{r-1}}} \big[T(x_1, \dots, x_{r-1}, \cdot) \equiv 0\big]$.

Theorem 7.6 (Tensor random restriction theorem)**.** *For every $d \in \mathbb{N}$ and $\sigma \in (0,1]$, there exist constants $C, \kappa > 0$ such that for any order-d tensor T over any field, we have that*

$$\mathrm{Pr}_{I\sim[n]_\sigma}\big[\mathrm{arank}(T_{|I}) \geq \kappa \cdot \mathrm{arank}(T)\big] \geq 1 - Ce^{-\kappa\,\mathrm{arank}(T)}.$$

7.2. **The proof of Theorem 7.1.** The proof will proceed by induction on the degree of the polynomial map. Recall that in the base case of degree-1 maps the result has already been proven in Section 4.1.

Let now $\phi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ be a polynomial map of degree at most d, with $2 \leq d < p$, and suppose the theorem holds for polynomial maps of degree at most $d-1$. Define the $(d+1)$-tensor $T : (\mathbb{F}_p^n)^d \times \mathbb{F}_p^k \to \mathbb{F}_p$ by

$$T(y_1, \dots, y_d, v) = \big\langle v,\, \Delta_{y_1} \dots \Delta_{y_d}\phi(0)\big\rangle.$$

We split the analysis into two cases, depending on whether the analytic rank of T is above or below some cut-off value $r = \Theta\big(k/(\log k)^{d^2}\big)$.

Pseudorandom case. Assume that $\mathrm{arank}(T) \geq r$. We will show that, for any given $x \in \mathbb{F}_p^n$, the probability

$$\mathrm{Pr}_{Z\sim\mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big) = x\big] \tag{7}$$

decays exponentially on $\mathrm{arank}(T)$; we then conclude the pseudorandom case by averaging over all x.

Fix some $x \in \mathbb{F}_p^n$. As before (in Section 5), we write

$$\begin{aligned}\mathrm{Pr}_{Z\sim\mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big) = x\big] &= \mathbb{E}_{I\sim[n]_{1-\rho}}\mathrm{Pr}_{y\in\mathbb{F}_p^I}\big[\phi\big(E(x)+y\big) = x\big] \\ &= \mathbb{E}_{I\sim[n]_{1-\rho}}\mathbb{E}_{y\in\mathbb{F}_p^I, v\in\mathbb{F}_p^k}\omega^{\langle v,\, \phi(E(x)+y)-x\rangle}.\end{aligned}$$

Note that we can write $\phi(E(x)+y) - x = \phi(y) + \psi(y)$, where

$$\psi(y) := \Delta_{E(x)}\phi(y) - x$$

has degree at most $d-1$. Using this identity and the triangle inequality, it follows that

$$\begin{aligned}\mathrm{Pr}_{Z\sim\mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big) = x\big] &= \mathbb{E}_{I\sim[n]_{1-\rho}}\mathbb{E}_{y\in\mathbb{F}_p^I, v\in\mathbb{F}_p^k}\omega^{\langle v,\, \phi(y)+\psi(y)\rangle} \\ &\leq \mathbb{E}_{I\sim[n]_{1-\rho}}\mathbb{E}_{v\in\mathbb{F}_p^k}\big|\mathbb{E}_{y\in\mathbb{F}_p^I}\omega^{\langle v,\, (\phi+\psi)(y)\rangle}\big|.\end{aligned}$$

Repeated applications of the Cauchy-Schwarz inequality (or equivalently, the monotonicity property of the Gowers uniformity norms [TV06, pp. 420]) shows that, for any fixed $v \in \mathbb{F}_p^k$, $I \subseteq [n]$, we have

$$\big|\mathbb{E}_{y\in\mathbb{F}_p^I}\omega^{\langle v,\, (\phi+\psi)(y)\rangle}\big| \leq \big(\mathbb{E}_{y_0,y_1,\dots,y_d\in\mathbb{F}_p^I}\omega^{\langle v,\, \Delta_{y_1}\dots\Delta_{y_d}(\phi+\psi)(y_0)\rangle}\big)^{1/2^d}.$$

We then conclude that

$$\begin{aligned}
&\mathrm{Pr}_{Z\sim\mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big)=x\big]\\
&\qquad\le \mathbb{E}_{I\sim[n]_{1-\rho}}\mathbb{E}_{v\in\mathbb{F}_p^k}\big(\mathbb{E}_{y_0,y_1,\ldots,y_d\in\mathbb{F}_p^I}\omega^{\langle v,\,\Delta_{y_1}\ldots\Delta_{y_d}(\phi+\psi)(y_0)\rangle}\big)^{1/2^d}\\
&\qquad\le \mathbb{E}_{I\sim[n]_{1-\rho}}\big(\mathbb{E}_{v\in\mathbb{F}_p^k}\mathbb{E}_{y_0,y_1,\ldots,y_d\in\mathbb{F}_p^I}\omega^{\langle v,\,\Delta_{y_1}\ldots\Delta_{y_d}(\phi+\psi)(y_0)\rangle}\big)^{1/2^d},
\end{aligned}$$

where we have applied Hölder's inequality once (or, alternatively, Cauchy-Schwarz d further times).

Now we need to relate this last expression to the analytic rank of T. Deriving d times a polynomial map of degree at most $d-1$ gives the zero map, and so $\Delta_{y_1}\ldots\Delta_{y_d}\psi(y_0)\equiv 0$. Moreover, since $\deg(\phi)\le d$, the d-th derivative $\Delta_{y_1}\ldots\Delta_{y_d}\phi$ is a constant map. We conclude that

$$\mathbb{E}_{v\in\mathbb{F}_p^k}\mathbb{E}_{y_0,y_1,\ldots,y_d\in\mathbb{F}_p^I}\omega^{\langle v,\,\Delta_{y_1}\ldots\Delta_{y_d}(\phi+\psi)(y_0)\rangle}=\mathbb{E}_{v\in\mathbb{F}_p^k}\mathbb{E}_{y_1,\ldots,y_d\in\mathbb{F}_p^I}\omega^{\langle v,\,\Delta_{y_1}\ldots\Delta_{y_d}\phi(0)\rangle}.$$

For each $v\in\mathbb{F}_p^k$, let $S(v)$ be the d-tensor given by $T(\cdot,\ldots,\cdot,v)$. Then the above is precisely the bias of the restricted tensor $S(v)_{|I^d}$, averaged over v, which (by definition) equals the average of $p^{-\operatorname{arank}(S(v)_{|I^d})}$. The probability (7) is then bounded from above by

$$\mathbb{E}_{v\in\mathbb{F}_p^k}\mathbb{E}_{I\sim[n]_{1-\rho}}p^{-\operatorname{arank}(S(v)_{|I^d})/2^d}.$$

Theorem 7.6 now implies that for some absolute constant $C=C(d,\rho)>0$, the last quantity is bounded from above by $Cp^{-\operatorname{arank}(T)/C}$. This settles the pseudorandom case.

Structured case. Now we assume that $\operatorname{arank}(T)<r$.

Denote the partition rank of T by $s:=\operatorname{prank}(T)$. Theorem 7.5 shows that $s\le L_{d+1}r(\log r)^d$, where L_{d+1} is a universal constant. We can then write

$$T(y_{[d]},v)=\sum_{i=1}^{s}R_i(y_{I_i})S_i(y_{I_i^c},v)$$

for some non-empty sets $I_i\subseteq[d]$, $|I_i|$-tensors R_i and $(d-|I_i|+1)$-tensors S_i. Since $d<p$, by Taylor's expansion theorem we have that

$$\phi(y)=\frac{1}{d!}\Delta_y\ldots\Delta_y\phi(0)+\psi_0(y),\quad \deg(\psi_0)<d.$$

Define $q_i:\mathbb{F}_p^n\to\mathbb{F}$, $\psi_i:\mathbb{F}_p^n\to\mathbb{F}_p^k$ $(i\in[s])$ by

$$q_i(y)=\frac{1}{d!}R_i(y^{I_i}),\quad \langle v,\psi_i(y)\rangle=S_i(y^{I_i^c},v),$$

and note that $\deg(\psi_i) < d$ for all $i \in [s]$. By the definition of T we conclude that

$$\phi(y) = \psi_0(y) + \sum_{i=1}^{s} q_i(y)\psi_i(y), \quad \text{with } \deg(\psi_i) < d \text{ for } 0 \le i \le s.$$

Let $\mathcal{A} = \{A_1, \dots, A_m\}$ be the partition of $\mathbb{F}_p^n$ given by the level sets of the polynomial map $(q_1, \dots, q_s) : \mathbb{F}_p^n \to \mathbb{F}_p^s$; note that $m \le p^s \le p^{L_{d+1} r (\log r)^d}$. For each $j \in [m]$, ϕ will coincide on A_j with a polynomial map $\psi_{A_j} : \mathbb{F}_p^n \to \mathbb{F}_p^k$ of degree at most $d-1$ (just substitute the $q_i(y)$ on the formula above by their value on $A_j \in \mathcal{A}$). Define the random events

$$\mathcal{E}_j = \big\{E(x) + Z \in A_j : x \sim \mathcal{U}(\mathbb{F}_p^k),\, Z \sim \mathcal{N}_\rho(0)\big\}, \quad j \in [m].$$

Since these events partition the probability space, it follows that

$$\begin{aligned}
&\Pr_{x \in \mathbb{F}_p^k, Z \sim \mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big) = x\big] \\
&\qquad = \sum_{i=1}^{m} \Pr_{x,Z}\big[\phi\big(E(x)+Z\big) = x \;\;\&\;\; \mathcal{E}_j\big] \\
&\qquad = \sum_{i=1}^{m} \Pr_{x,Z}\big[\psi_{A_j}\big(E(x)+Z\big) = x \;\;\&\;\; \mathcal{E}_j\big] \\
&\qquad \le m \cdot \max_{1 \le j \le m} \Pr_{x,Z}\big[\psi_{A_j}\big(E(x)+Z\big) = x\big] \\
&\qquad \le p^{L_{d+1} r (\log r)^d} \cdot \max_{\deg(\psi) < d} \Pr_{x,Z}\big[\psi\big(E(x)+Z\big) = x\big],
\end{aligned}$$

where the last maximum is over all polynomial maps $\psi : \mathbb{F}_p^n \to \mathbb{F}_p^k$ of degree at most $d-1$. By the induction hypothesis we have that this maximum probability is at most $C' e^{-c'k/(\log k)^{(d-1)^2}}$, where $C' = C(d-1, \rho)$ and $c' = c(d-1, \rho)$; we conclude that

$$\begin{aligned}
&\Pr_{x \in \mathbb{F}_p^k, Z \sim \mathcal{N}_\rho(0)}\big[\phi\big(E(x)+Z\big) = x\big] \\
&\qquad \le C' \exp\left((\log p) L_{d+1} r (\log r)^d - \frac{c'k}{(\log k)^{(d-1)^2}}\right).
\end{aligned}$$

Taking

$$r = \frac{c'}{2L_{d+1}} \frac{k}{(\log k)^{d^2}},$$

and using our assumptions $k \geq p$ and $d \geq 2$, we have that

$$\begin{aligned}(\log p) L_{d+1} r (\log r)^d &\leq (\log k) L_{d+1} \frac{c'}{2L_{d+1}} \frac{k}{(\log k)^{d^2}} (\log k)^d \\ &= \frac{c'}{2} \frac{k}{(\log k)^{d^2-d-1}} \\ &\leq \frac{c'}{2} \frac{k}{(\log k)^{(d-1)^2}}.\end{aligned}$$

We conclude that

$$\Pr_{x \in \mathbb{F}_p^k, Z \sim \mathcal{N}_\rho(0)} \left[\phi\big(E(x) + Z\big) = x \right] \leq C' \exp\left(- \frac{c'k}{2(\log k)^{(d-1)^2}} \right)$$

in this case, and the theorem follows.

Appendix A. MAJORITY from list decoding

This appendix shows how to compute the MAJORITY function when given oracle access to circuits capable of list decoding the Hadamard code.

A.1. Classical circuits.

We start by considering the case of classical circuits, in particular proving Theorem 2.6, which we recall below for convenience. Our proof of this result follows the arguments exposed in [Vio06, Section 6.2].

Theorem 2.6 (MAJORITY from list-Hadamard)**.** *Let $\mathcal{C}$ be a probabilistic circuit that solves the list-Hadamard problem* $\mathrm{LH}_n(\varepsilon)$ *with probability at least* $3/4$. *There exists a (deterministic) oracle* AC^0 *circuit $\mathcal{D}$ of size* $\mathrm{poly}(n, 1/\varepsilon)$ *which, when given oracle access to $\mathcal{C}$ and the ability to fix its random bits, computes MAJORITY on* $\Omega(1/\varepsilon)$ *bits.*

Let Maj_t denote the MAJORITY function on t bits. We first introduce a promise problem called IsBal_t, which asks to determine whether a given binary string is balanced. We then show that a (possibly probabilistic) circuit that solves IsBal_t can be turned into a deterministic circuit that computes Maj_t. Finally, we show how a circuit for $\mathrm{LH}_n(\varepsilon)$ can be used to solve IsBal_t for $t = \Omega(1/\varepsilon)$.

Definition A.1 (The IsBal_t problem)**.** For an even positive integer t, define $\mathrm{IsBal}_t : \{x \in \mathbb{F}_2^t : |x| \leq t/2\} \to \mathbb{F}_2$ by

$$\mathrm{IsBal}_t(x) = \begin{cases} 1 & \text{if } |x| = t/2 \\ 0 & \text{otherwise.} \end{cases}$$

Given an arbitrary $x \in \mathbb{F}_2^t$, define the IsBal_t problem to be to return $\mathrm{IsBal}_t(x)$ if $|x| \leq t/2$ and an arbitrary bit otherwise.

Lemma A.2 (Derandomization lemma)**.** *Let $\mathcal{C}$ be a probabilistic circuit that solves* IsBal_t *with probability at least* $2/3$ *for every input. There exists a deterministic oracle* AC^0 *circuit* $\mathcal{C}'$ *that, when given oracle access to* $\mathcal{C}$ *and the ability to fix its random bits, solves* IsBal_t*.*

Proof: For some large enough constant $c \in \mathbb{N}$, consider ct parallel instances of $\mathcal{C}$. It follows from the Chernoff bound that, for any fixed $x \in \mathbb{F}_2^t$ given to all of these instances, with probability $1 - \exp(-10\,t)$ at least 55% of the instances solves the IsBal_t problem on input x.

By the union bound, one can fix the randomness in the instances of $\mathcal{C}$ in order to get a deterministic classical circuit that, for every input $x \in \mathbb{F}_2^t$ with $|x| \leq t/2$, returns a ct-bit string whose Hamming weight is at least $0.55t$ if $\mathrm{IsBal}_t(x) = 1$ and at most $0.45t$ if $\mathrm{IsBal}_t(x) = 0$. Distinguishing these two types of strings is known as the approximate majority problem, for which there is an AC^0 circuit [Ajt83]. Combining these circuits gives the result. □

We now show that a deterministic circuit that solves IsBal_t can be used to compute Maj_t.

Lemma A.3. *Let $\mathcal{C}$ be a deterministic circuit for* IsBal_t*. There exists an oracle* AC^0 *circuit $\mathcal{D}$ that, given oracle access to $\mathcal{C}$, computes* Maj_t*.*

Proof: For $x \in \mathbb{F}_2^t$ and $i \in \{0, 1, \ldots, t\}$, define x_i as the string x with the first i bits set to zero and the rest of the bits equal to those of x. So, for instance, $x_0 = x$ and x_t is the all-zeroes string. Let $\mathcal{D}$ be the circuit that runs $t + 1$ parallel instances of $\mathcal{C}$ with inputs $x_0, x_1, \ldots, x_t$, respectively, and returns the OR of the $t + 1$ outputs.

We claim that $\mathcal{D}$ computes Maj_t. Indeed, if x has fewer than $t/2$ ones then $\mathcal{C}$ returns 0 for each input x_i, as the number of 1s only decreases with i. If x has at least $t/2$ ones, then $\mathcal{C}$ returns 1 for at least one i, since x_0 has at least $t/2$ ones, whereas x_t is the all-zeroes string. This completes the proof. □

Towards turning a circuit $\mathcal{C}$ for $\mathrm{LH}_n(\varepsilon)$ into a circuit for IsBal_t, we associate with each input $x \in \mathbb{F}_2^t$ to IsBal_t a random error vector N_x over $\mathbb{F}_2^n$ as follows: independently, each coordinate of N_x is a uniformly random entry of x. In particular, for balanced x, the error vector N_x will correspond to an error rate of 1/2 and we refer to it as $N_{1/2}$. The next lemma shows that there is a message $m \in \mathbb{F}_2^k$ that has small probability of recovery by $\mathcal{C}$ under the error vector $N_{1/2}$.

Lemma A.4. *Let $\mathcal{C}$ be a probabilistic circuit which, on input $y \in \mathbb{F}_2^n$, returns a (random) list $L(y) \subseteq \mathbb{F}_2^k$ of at most $2^k/4$ elements. Then there exists $m \in \mathbb{F}_2^k$ such that*

$$\Pr[m \in L(H(m) + N_{1/2})] \leq 1/4, \tag{8}$$

where the probability is taken over L and $N_{1/2}$.

Proof: Note that, for any $y \in \mathbb{F}_2^n$, the vector $y + N_{1/2}$ is uniformly distributed over $\mathbb{F}_2^n$; in particular, it has the same distribution as $N_{1/2}$. Let $M \in \mathbb{F}_2^k$ be a uniformly distributed random element. Then, by independence of M, $L(y)$ and $N_{1/2}$, get that

$$\begin{aligned}\Pr_{M,L,N_{1/2}}[M \in L(H(M) + N_{1/2})] &= \Pr_{M,L,N_{1/2}}[M \in L(N_{1/2})] \\ &\leq \frac{1}{2^k}\,\mathbb{E}_{L,N_{1/2}}\,|L(N_{1/2})| \\ &\leq 1/4.\end{aligned}$$

Hence, there exists a value m of M such that (8) holds. □

Finally, we prove that the circuit $\mathcal{C}$ in Theorem 2.6 can solve IsBal_t.

Lemma A.5. *Let $\mathcal{C}$ be a probabilistic circuit as in Theorem 2.6. There exists a probabilistic oracle* AC^0 *circuit $\mathcal{D}$ of size* $\mathrm{poly}(n, 1/\varepsilon)$ *that, when given oracle access to $\mathcal{C}$, solves* IsBal_t *with probability at least 3/4 for $t = \Omega(1/\varepsilon)$.*

Proof: We may assume without loss of generality that $\varepsilon \leq 1/4$. Let $\delta \in [\varepsilon, 1/4]$ be minimized such that $t = 1/(2\delta)$ is an even integer; note that, since $\delta \geq \varepsilon$, the circuit $\mathcal{C}$ also solves $\mathrm{LH}_n(\delta)$ with probability at least 3/4. Fix a message m as in Lemma A.4, and let $x \in \mathbb{F}_2^t$ be any given string (which serves as input to $\mathcal{D}$).

The circuit $\mathcal{D}$ has three layers. The first layer has the string $H(m)$ hardwired into it and uses n independent uniform samples to the coordinates of x to compute the random string $H(m) + N_x$. This layer is a probabilistic circuit using n parallel two-bit XOR gates. The second layer consists of the circuit $\mathcal{C}$, which produces a random list $L(H(m) + N_x)$ of size at most $n/4$. The third layer consists of an AC^0 circuit of size $\mathrm{poly}(n)$ that returns 0 if and only if $m \in L(H(m) + N_x)$. This can be done by checking equality between m and the $O(n)$ elements of the list. We claim that this solves IsBal_t.

If x is balanced then it follows from Lemma A.4 that $\mathcal{D}$ correctly returns 1 with probability at least 3/4. If x has Hamming weight strictly less than $t/2$, then each coordinate of N_x is 1 with probability at most $1/2 - 1/t = 1/2 - 2\delta$. By the Chernoff bound, N_x has Hamming weight at most $(1/2 - \delta)n$ with probability $1 - \exp(-\Omega(\delta^2 n))$. Hence, the properties of the circuit $\mathcal{C}$ imply that in this case $\mathcal{D}$ correctly outputs 0 with probability at least 3/4. □

Theorem 2.6 now follows directly by combining Lemma A.2, Lemma A.3 and Lemma A.5.

A.2. **The quantum case.** Now we sketch how the above proof can be used to turn our $\mathrm{QNC}^0[\oplus]$ circuit for decoding the Hadamard code into one that computes MAJORITY with polynomially small error. We first recall the following result.

Corollary 2.3. *There is a family of* $\mathrm{QNC}^0[\oplus]$ *circuits* $(\mathcal{C}_n)_{n\in\mathbb{N}}$ *such that the following holds. Let* $k \in \mathbb{N}$, $n = 2^k$ *and* $\varepsilon \in [1/\sqrt{n}, 1/2]$. *Then, on any input* $y \in \mathbb{F}_2^n$, *with probability* $1-\varepsilon$ *the circuit* $\mathcal{C}_n$ *returns a list* $L(y)$ *of size* $O(\varepsilon^{-2}\log(1/\varepsilon))$ *which contains every* $x \in \mathbb{F}_2^k$ *with* $d\big(y, H(x)\big) \leq (\frac{1}{2}-\varepsilon)n$.

Let $\varepsilon = n^{-1/4}$ and let $\mathcal{C}$ be the circuit from Corollary 2.3. Since $\mathcal{C}$ returns lists of size at most $n^{3/4}$, a stronger version of Lemma A.4 holds where the probability (8) – taken additionally over the measurement outcomes of $\mathcal{C}$ – is bounded from above by $n^{3/4}/2^k = n^{-1/4}$.

The proof of Lemma A.5 then gives an oracle $\mathrm{QNC}^0[\oplus]$ circuit $\mathcal{D}$ of size $\mathrm{poly}(n)$ that, given oracle access to $\mathcal{C}$, solves IsBal_t with probability $1-O(n^{-1/4})$ for $t = \Omega(n^{1/4})$. Here, the AC^0 circuit used to check membership of m can be replaced with our $\mathrm{QNC}^0[\oplus]$ circuit for the OR function (see Section 6) applied to the entrywise sum of m with each element in the list.

Now let $t' = \lfloor n^{1/8} \rfloor$, and note that the same circuit $\mathcal{D}$ above can be used to solve $\mathrm{IsBal}_{t'}$ with probability $1-O(n^{-1/4})$: it suffices to pad the input with zeroes and ones in the same number until we have a string of the correct size. Finally, with the proof of Lemma A.3 and the union bound we obtain a $\mathrm{QNC}^0[\oplus]$ circuit $\mathcal{D}'$ that, given oracle access to $\mathcal{D}$, solves $\mathrm{Maj}_{t'}$ with probability $1-O(n^{-1/8})$ for $t' = \lfloor n^{1/8} \rfloor$. Here again we use our $\mathrm{QNC}^0[\oplus]$ circuit for the OR function as explained in Section 6.

Acknowledgements

We thank Richard Cleve for helpful discussions on the proof of Theorem 2.2, Emanuele Viola for helpful pointers to the literature and comments on an earlier version of this manuscript, Tamar Ziegler for encouragement to write up Section 7, as well as anonymous referees for helpful comments.

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity*. Cambridge University Press, Cambridge, 2009. A modern approach. doi:10.1017/CBO9780511804090.

[Ajt83] M. Ajtai. Σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983. doi:10.1016/0168-0072(83)90038-6.

[BCS22] Jop Briët and Davi Castro-Silva. Random restrictions of high-rank tensors and polynomial maps, 2022. arXiv:2212.13728. doi:10.48550/arXiv.2212.13728.

[BGK18] Sergey Bravyi, David Gosset, and Robert König. Quantum advantage with shallow circuits. *Science*, 362(6412):308–311, Oct 2018. doi:10.1126/science.aar3106.

[BGKT20] Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. Quantum advantage with noisy shallow circuits. *Nature Physics*, 16(10):1040–1045, 2020. Preliminary version in FOCS'19. doi:10.1038/s41567-020-0948-z.

[BV97] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. doi:10.1137/S0097539796300921.

[BWKST19] Adam Bene Watts, Robin Kothari, Luke Schaeffer, and Avishay Tal. Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, page 515–526. ACM, Jun 2019. doi:10.1145/3313276.3316404.

[CHTW04] Richard Cleve, Peter Høyer, Benjamin Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *Proceedings. 19th IEEE Annual Conference on Computational Complexity, 2004.*, pages 236–249. IEEE, 2004. doi:10.1109/CCC.2004.1313847.

[CSV21] Matthew Coudron, Jalex Stark, and Thomas Vidick. Trading locality for time: certifiable randomness from low-depth circuits. *Communications in Mathematical Physics*, 382(1):49–86, 2021. doi:10.1007/s00220-021-03963-w.

[EJPP00] J. Eisert, K. Jacobs, P. Papadopoulos, and M. B. Plenio. Optimal local implementation of nonlocal quantum gates. *Phys. Rev. A*, 62:052317, Oct 2000. doi:10.1103/PhysRevA.62.052317.

[Eli57] P Elias. List decoding for noisy channels. In *IRE WESCON Convention Record, 1957*, volume 2, pages 94–104, 1957.

[Gal19] François Le Gall. Average-Case Quantum Advantage with Shallow Circuits. In Amir Shpilka, editor, *34th Computational Complexity Conference (CCC 2019)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:20, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2019.21.

[GHMP02] Frederic Green, Steven Homer, Cristopher Moore, and Christopher Pollett. Counting, fanout and the complexity of quantum *ACC*. *Quantum Info. Comput.*, 2(1):35–65, dec 2002. doi:10.26421/QIC2.1-3.

[GL89] O. Goldreich and L. A. Levin. A hard-core predicate for all one-way functions. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, STOC '89, page 25–32, New York, NY, USA, 1989. Association for Computing Machinery. doi:10.1145/73007.73010.

[GS20] Daniel Grier and Luke Schaeffer. Interactive shallow clifford circuits: Quantum advantage against nc^1 and beyond. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC'20)*, page 875–888, New York, NY, USA, 2020. Association for Computing Machinery. URL: https://doi.org/10.1145/3357713.3384332.

[GT09] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math.*, 4(2):1–36, 2009. doi:10.11575/cdm.v4i2.62086.

[GV10] Venkatesan Guruswami and Salil Vadhan. A lower bound on list size for list decoding. *IEEE Transactions on Information Theory*, 56(11):5681–5688, 2010. doi:10.1109/TIT.2010.2070170.

[GW11] W. T. Gowers and J. Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geom. Funct. Anal.*, 21(1):36–69, 2011. doi:10.1007/s00039-010-0106-3.

[Ham50] R. W. Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160, 1950. doi:10.1002/j.1538-7305.1950.tb00463.x.

[HHL19] Hamed Hatami, Pooya Hatami, and Shachar Lovett. Higher-order Fourier analysis and applications. *Found. Trends Theor. Comput. Sci.*, 13(4):247–448, 2019. doi:10.1561/0400000064.

[HR90] Torben Hagerup and Christine Rüb. A guided tour of Chernoff bounds. *Information Processing Letters*, 33(6):305–308, 1990. doi:10.1016/0020-0190(90)90214-I.

[HŠ05] Peter Høyer and Robert Špalek. Quantum fan-out is powerful. *Theory of Computing*, 1(5):81–103, 2005. doi:10.4086/toc.2005.v001a005.

[KL08] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS'08)*, pages 166–175, 2008. doi:10.1109/FOCS.2008.17.

[Lov19] Shachar Lovett. The analytic rank of tensors and its applications. *Discrete Anal.*, pages Paper No. 7, 10, 2019. doi:10.19086/da.

[LV17] Chin Ho Lee and Emanuele Viola. Some limitations of the sum of small-bias distributions. *Theory of Computing*, 13(16):1–23, 2017. doi:10.4086/toc.2017.v013a016.

[MN01] Cristopher Moore and Martin Nilsson. Parallel quantum computation and quantum codes. *SIAM Journal on Computing*, 31(3):799–815, 2001. doi:10.1137/S0097539799355053.

[Moo99] Cristopher Moore. Quantum circuits: Fanout, parity, and counting. *arXiv preprint arXiv:quant-ph/9903046*, 1999. arXiv:arXiv:quant-ph/9903046.

[MZ22] Guy Moshkovitz and Daniel G. Zhu. Quasi-linear relation between partition and analytic rank, 2022. arXiv:2211.05780. doi:10.48550/ARXIV.2211.05780.

[Nas20] Eric Naslund. The partition rank of a tensor and k-right corners in $\mathbb{F}_q^n$. *J. Combin. Theory Ser. A*, 174:105190, 25, 2020. doi:10.1016/j.jcta.2019.105190.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010. doi:10.1017/CBO9780511976667.

[PS13] Paul Pham and Krysta M. Svore. A 2D nearest-neighbor quantum architecture for factoring in polylogarithmic depth. *Quantum Info. Comput.*, 13(11–12):937–962, 2013. doi:10.26421/QIC13.11-12-3.

[Raz87] A. A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, April 1987. doi:10.1007/bf01137685.

[RU10] Atri Rudra and Steve Uurtamo. Two theorems on list decoding. In Maria Serna, Ronen Shaltiel, Klaus Jansen, and José Rolim, editors, *Approximation,* *Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 696–709, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

[Sho97] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal of Computing*, 26(5):1484–1509, 1997. Preliminary version in FOCS'94. doi:10.1137/S0097539795293172.

[Smo87] R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, page 77–82, New York, NY, USA, 1987. Association for Computing Machinery. doi:10.1145/28395.28404.

[STV99] M. Sudan, L. Trevisan, and S. Vadhan. Pseudorandom generators without the xor lemma. In *Proceedings. Fourteenth Annual IEEE Conference on Computational Complexity (Formerly: Structure in Complexity Theory Conference) (Cat.No.99CB36317)*, pages 4–, 1999. doi:10.1109/CCC.1999.766253.

[Tao12] Terence Tao. *Higher order Fourier analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012. doi:10.1090/gsm/142.

[Tre04] Luca Trevisan. Some applications of coding theory in computational complexity. In *Complexity of computations and proofs*, volume 13 of *Quad. Mat.*, pages 347–424. Dept. Math., Seconda Univ. Napoli, Caserta, 2004.

[TT13] Yasuhiro Takahashi and Seiichiro Tani. Collapse of the hierarchy of constant-depth exact quantum circuits. In *2013 IEEE Conference on Computational Complexity*, pages 168–178, 2013. doi:10.1109/CCC.2013.25.

[TV06] Terence Tao and Van H. Vu. *Additive Combinatorics*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2006. doi:10.1017/CBO9780511755149.

[TV07] Luca Trevisan and Salil Vadhan. Pseudorandomness and average-case complexity via uniform reductions. *Computational Complexity*, 16(4):331–364, 2007. doi:10.1007/s00037-007-0233-x.

[Vio06] Emanuele Viola. *The Complexity of Hardness Amplification and Derandomization*. PhD thesis, Harvard University, USA, 2006. AAI3217914. doi:10.5555/1195277.

[Vio09] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Comput. Complexity*, 18(2):209–217, 2009. Preliminary version in CCC'08. doi:10.1007/s00037-009-0273-5.

[Woz58] John M Wozencraft. List decoding. *Quarterly Progress Report*, 48:90–95, 1958.

[Yao77] Andrew Chi-Chin Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*, pages 222–227, 1977. doi:10.1109/SFCS.1977.24.

[YL04] Anocha Yimsiriwattana and Jr Lomonaco. Generalized GHZ states and distributed quantum computing. *Coding Theory and Quantum Computing*, 381, 03 2004. doi:10.1090/conm/381/07096.

CWI & QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands
Email address: j.briet@cwi.nl

QuSoft & University of Amsterdam & CWI, Science Park 123, 1098 XG Amsterdam, The Netherlands
Email address: buhrman@cwi.nl

CWI & QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands
Email address: davi.silva@cwi.nl

CWI & QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands, & The Netherlands Organisation for Applied Scientific Research (TNO)
Email address: niels.neumann@tno.nl