An Optimal Lower Bound on the Communication Complexity of GAP-HAMMING-DISTANCE

Amit Chakrabarti* Oded Regev†

September 20, 2010

Abstract

We prove an optimal $\Omega(n)$ lower bound on the randomized communication complexity of the much-studied GAP-HAMMING-DISTANCE problem. As a consequence, we obtain essentially optimal multi-pass space lower bounds in the data stream model for a number of fundamental problems, including the estimation of frequency moments.

The GAP-HAMMING-DISTANCE problem is a communication problem, wherein Alice and Bob receive n-bit strings x and y, respectively. They are promised that the Hamming distance between x and y is either at least $n/2 + \sqrt{n}$ or at most $n/2 - \sqrt{n}$, and their goal is to decide which of these is the case. Since the formal presentation of the problem by Indyk and Woodruff (FOCS, 2003), it had been conjectured that the naïve protocol, which uses n bits of communication, is asymptotically optimal. The conjecture was shown to be true in several special cases, e.g., when the communication is deterministic, or when the number of rounds of communication is limited.

The proof of our aforementioned result, which settles this conjecture fully, is based on a new geometric statement regarding correlations in Gaussian space, related to a result of C. Borell (1985). To prove this geometric statement, we show that random projections of not-too-small sets in Gaussian space are close to a mixture of translated normal variables.

*Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA. Supported by NSF Grant IIS-0916565 and a McLane Family Fellowship.

†Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. Supported by the Israel Science Foundation, by the Wolfson Family Charitable Trust, and by a European Research Council (ERC) Starting Grant. Part of this work done while a DIGITEO visitor in LRI, Orsay.

1 Introduction

Communication complexity is a much-studied topic in computational complexity, deriving its importance both from the basic nature of the questions it asks and the wide range of applications of its results, covering, for instance, lower bounds on circuit depth (see, e.g., [KW90]) and on query times for static data structures (see, e.g., [MNSW98, Păt08]). In the basic setup, which is all that concerns us here, each of two players, Alice and Bob, receives a binary string as input. Their goal is to compute some function of the two strings, using a protocol that involves exchanging a *small* number of bits. Since communication complexity is often applied as a lower bound technique, much of the work in the area attempts to rule out the existence of a nontrivial protocol. For many functions, this amounts to proving an $\Omega(n)$ lower bound on the number of bits any successful protocol must exchange, n being the common length of Alice's and Bob's input strings. Proofs tend to be considerably more challenging, and more broadly applicable, when the protocol is allowed to be *randomized* and err with some small constant probability (such as $1/3$) on each input.

For a detailed coverage of the basics of the field, as well as a number of applications, we refer the reader to the textbook of Kushilevitz and Nisan [KN97]. Here, we only recap the most basic notions, in Section 2.

In this paper, we focus specifically on the Gap-Hamming-Distance problem (henceforth abbreviated as GHD), which was first formally studied by Indyk and Woodruff [IW03] in the context of proving space lower bounds for the Distinct Elements problem in the data stream model. We also consider some closely related variants of GHD.

The Problem and the Main Result. In the Gap-Hamming-Distance problem $\text{GHD}_{n,t,g}$, Alice and Bob receive binary strings $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$, respectively. They wish to decide whether x and y are "close" or "far" in the Hamming sense, with a certain *gap* separating the definitions of "close" and "far." Specifically, the players must output 0 if $\Delta(x,y) \leq t - g$ and 1 if $\Delta(x,y) > t + g$, where Δ denotes Hamming distance; if neither of these holds, they may output either 0 or 1. Clearly, this problem becomes easier as the gap g increases. Of special interest is the case when $t = n/2$ and $g = \Theta(\sqrt{n})$; these parameters are natural, and as we shall show later using elementary reductions, understanding the complexity of the problem with these parameters leads to a complete understanding of the problem for essentially all other gap sizes and threshold locations. Furthermore, applications of GHD, such as the ones considered by Indyk and Woodruff [IW03], need precisely this natural setting of parameters. Henceforth, we shall simply write "GHD" to denote $\text{GHD}_{n,n/2,\sqrt{n}}$.

Our main result states, simply, that this problem does not have a nontrivial protocol. Here is a somewhat informal statement; a fully formal version appears as Theorem 2.6.

Theorem 1.1 (Main Theorem, Informal)**.** *If a randomized protocol solves* GHD*, then it must communicate a total of* $\Omega(n)$ *bits.*

Relation to Prior Work. Theorem 1.1 is the logical conclusion of a moderately long line of research. This was begun in the aforementioned work of Indyk and Woodruff [IW03], who showed a linear lower bound on the communication complexity of a somewhat artificial variant of GHD in the *one-way* model, i.e., in the model where the communication is required to consist of just one message from Alice to Bob. Woodruff [Woo04] soon followed up with an $\Omega(n)$ bound for GHD itself, still in the one-way model; the proof used rather intricate combinatorial constructions and computations. Jayram et al. [JKS07] later provided a rather different and much simpler proof, by a reduction from the INDEX problem. Their reduction was geometric, in the sense that they exploited

a natural correspondence between Hamming space and Euclidean space; this correspondence has proved fruitful in further work on the problem, including this work. Recently, Woodruff [Woo09] and Brody and Chakrabarti [BC09] gave direct combinatorial proofs of the $\Omega(n)$ one-way bound.

All of this work left open an important question: *what can be said about the complexity of* GHD *when two-way communication is allowed?* It has been conjectured, at least since the formalization of the problem in 2003, that $\Omega(n)$ is still the right answer, i.e., that GHD has *no* nontrivial protocol, irrespective of the communication pattern.

Until 2009, our understanding of this matter was limited to two "folklore" results. Firstly, the *deterministic* communication complexity of $\text{GHD}_{n,n/2,g}$ can be shown to be $\Omega(n)$, even allowing two-way communication and a gap as large as $g = cn$, for a small enough constant c. This follows by directly demonstrating that its communication matrix contains no large monochromatic rectangles (see, e.g., [Woo07]). Secondly, a simple reduction from DISJOINTNESS to $\text{GHD}_{n,n/2,g}$ shows that its randomized (two-way) communication complexity is $\Omega(n/g)$; notice that the corresponding bound for GHD (where $g = \sqrt{n}$) is $\Omega(\sqrt{n})$. Meanwhile, we have an *upper* bound of $O(n^2/g^2)$, via the simple (and one-way) protocol that samples sufficiently many coordinates of x and y to give the right answer with high probability. It remained a significant challenge to improve upon either tradeoff, even for just two rounds of communication.

Recently, Brody and Chakrabarti [BC09] made progress on the conjecture, proving it for randomized protocols with two-way communication, but only a constant number of rounds of communication. In fact, they showed that in a k-round protocol, at least one message must have length $n/2^{O(k^2)}$. They achieved this via a round elimination argument. At a high level, they showed that if the first message in a GHD protocol is too short, the work done by the rest of the messages can be used to solve a "smaller" instance of GHD, by exploiting some combinatorial properties of Hamming space. More recently, Brody et al. [BCR$^+$10] improved the bound to $\Omega(n/(k^2 \log k))$, still using a round elimination argument, but exploiting geometric properties of Hamming and Euclidean space instead. We refer the reader to the discussion in [BCR$^+$10] for details, including a comparison of the two arguments.

Our main theorem completes this picture, confirming the main outstanding conjecture about GHD. Moreover, a straightforward reduction (Lemma 4.4) yields the more general result that the randomized complexity of $\text{GHD}_{n,n/2,g}$ is $\Theta(\min\{n, n^2/g^2\})$. Our lower bound proof is significantly different in approach from all of the aforementioned ones. We now give a high-level overview.

The Technique. Part of the difficulty in establishing our result is that many of the known techniques for proving communication complexity lower bounds seem unable to prove bounds better than $\widetilde{\Omega}(\sqrt{n})$. These include the classic rectangle-based methods of discrepancy and corruption,[1] for reasons described below. They also include certain linear algebraic approaches, such as the factorization norms method of Linial and Shraibman [LS07] and the pattern matrix method of Sherstov [She08], because these methods lower bound *quantum* communication complexity. The trouble is that GHD does have a constant-error $O(\sqrt{n} \log n)$ quantum communication protocol, as can be seen by combining a query complexity upper bound due to Nayak and Wu [NW99] with a communication-to-query reduction, as in Buhrman et al. [BCW98] or Razborov [Raz02].

Instead, what does work is a suitable generalization of the corruption method. Recall that the standard corruption method proceeds as follows. First, one observes that every protocol

[1] We assume that the reader has some familiarity with these basic techniques in communication complexity, which are discussed in detail in the textbook of Kushilevitz and Nisan [KN97]. Some authors use terms like "one-sided discrepancy" and "rectangle bound" when describing the technique that we (following Beame et al. [BPSW06]) have termed "corruption."

that communicates c bits induces a partition of the communication matrix into 2^c disjoint near-monochromatic rectangles. In order to show a lower bound of c, one then needs to prove that any rectangle containing at least a 2^{-c} fraction of the 1-inputs must also contain (or be "corrupted" by) a not-much-smaller fraction of the 0-inputs (or vice versa). In other words, one shows that large near-monochromatic rectangles do not exist, from which the desired lower bound follows. It should be noted that proving such a property could be a challenging task. Indeed, this is the main technical contribution of Razborov's celebrated proof of the $\Omega(n)$ lower bound on the randomized communication complexity of the disjointness problem [Raz92].

This idea appears not to give a lower bound better than $\Omega(\sqrt{n})$ on the randomized communication complexity of GHD because its communication matrix *does* contain "annoying" rectangles that are both large and near-monochromatic. This can be seen, e.g., by considering all inputs (x, y) with $x_i = 0$, $y_i = 1$ for $i \in \{1, 2, \ldots, 100\sqrt{n}\}$: the resulting rectangle contains a $2^{-\Theta(\sqrt{n})}$ fraction of all 1-inputs (it is large), but a much smaller fraction of 0-inputs (it is nearly monochromatic).

Our generalization considers not just 0-inputs and 1-inputs, but also a carefully selected set of "joker" inputs, whose corresponding outputs are immaterial. Loosely speaking, we show that if a large rectangle contains many more 1-inputs than 0-inputs, then the fraction of joker inputs it contains must be *even larger* than the fraction of 1-inputs it contains (by some constant factor, say $3/2$). This property — call it the "joker property" — implies that even though annoying rectangles exist, their union cannot contain more than a constant fraction of the 1-inputs (say, $2/3$). In particular, there is no way to partition the 1-inputs into 2^c near-monochromatic rectangles, and a lower bound of c follows.

This simple-sounding idea seems to have considerable power. Indeed, the method we have presented above can be seen as a special case of the ideas behind the "smooth rectangle bound" recently introduced by Klauck [Kla10] and systematized by Jain and Klauck [JK10]. Formally, when we prove a communication lower bound using corruption-with-jokers as above, we are essentially lower bounding the smooth rectangle bound of the underlying function. For a careful understanding of this matter, based on linear programming duality, we refer the reader to Jain and Klauck [JK10].

Of course, there remains the task of proving the joker property referred to above. It turns out that the statement we need boils down to roughly the following: for arbitrary sets $A, B \subseteq \{0,1\}^n$ that are not too small (say, of size at least $2^{0.99n}$), if $x \in_R A$ and $y \in_R B$, then $\Delta(x, y)$ is not too concentrated around $n/2$; a precise statement appears as Corollary 3.8. The proof uses a *Gaussian noise correlation inequality* (Theorem 3.5, proved using analytic methods); this inequality and its proof are the main technical contributions of the paper and should be of independent interest. We note that Vidick [Vid10] recently communicated to us an alternative proof of the joker property.

Data Stream and Other Consequences. The original motivation for studying GHD was a specific application to the Distinct Elements problem on data streams. Specifically, given a stream (sequence) of m elements, each from $[n] := \{1, 2, \ldots, n\}$, we wish to estimate, to within a $1 \pm \varepsilon$ approximation, the number of distinct elements in it, while using space sublinear in m and n. A long line of research has culminated in a randomized algorithm [BJK$^+$04] that computes such an estimate (failing with probability at most $\frac{1}{3}$, say) in one pass over the stream, using $\widetilde{O}(\varepsilon^{-2})$ bits of space, where the $\widetilde{O}$-notation suppresses factors logarithmic in m and n. An easy reduction (implicit in Indyk and Woodruff [IW03]) shows that a lower bound of $\Omega(\phi(n, k))$ on the maximum message length of a $(2k-1)$-round protocol for GHD would imply a $\Omega(\phi(\varepsilon^{-2}, k))$ space lower bound on k-pass algorithms for the Distinct Elements problem. Thus, the one-way $\Omega(n)$ lower bound for GHD implied a tight $\Omega(\varepsilon^{-2})$ lower bound for one-pass streaming algorithms. The results of Brody and

Chakrabarti [BC09] and Brody et al. [BCR$^+$10] extended this to p-pass algorithms, giving lower bounds of $\Omega(\varepsilon^{-2}/2^{O(p^2)})$ and $\Omega(\varepsilon^{-2}/(p^2 \log p))$, respectively.

Our main result improves this pass/space tradeoff, giving a space lower bound of $\Omega(\varepsilon^{-2}/p)$. As is easy to see, this is tight up to factors logarithmic in m and n.

Suitable reductions from GHD imply similar space lower bounds for several other data stream problems, such as estimating frequency moments [Woo04] and empirical entropy [CCM07]. One can also derive appropriate lower bounds for a certain class of distributed computing problems known as functional monitoring [ABC09].

2 Corruption, a Generalization, and the Main Theorem

2.1 Preliminaries

Consider a communication problem given by a (possibly partial) function $f : X \times Y \to \{0, 1, \star\}$; we let f take the value "$\star$" at inputs for which we do not care about the output given. For a communication protocol, P, involving two players, Alice and Bob, we write $P(x, y)$ to denote the output of P when Alice receives $x \in X$ and Bob receives $y \in Y$. If P is randomized, this is a random variable. We say that P computes f with error at most ε if

$$\forall\,(x,y) \in X \times Y :\ f(x,y) \neq \star \Rightarrow \Pr[P(x,y) \neq f(x,y)] \leq \varepsilon\,.$$

When the function f is understood from the context, we use $\mathrm{err}(P)$ to denote $\inf\{\varepsilon :\ P$ computes f with error at most $\varepsilon\}$. For a deterministic protocol P and a distribution μ on $X \times Y$, we define

$$\mathrm{err}_\mu(P) \ := \ \Pr_{(x,y)\sim\mu}[f(x,y) \neq \star \ \wedge\ P(x,y) \neq f(x,y)]\,.$$

For a protocol P, let $\mathrm{cost}(P)$ denote the worst-case number of bits communicated by P. We let $\mathrm{R}_\varepsilon(f)$ and $\mathrm{D}_{\mu,\varepsilon}(f)$ denote the ε-error randomized and ε-error μ-distributional communication complexities of f, respectively; i.e.,

$$\begin{aligned}
\mathrm{R}_\varepsilon(f) \ &= \ \min\{\mathrm{cost}(P) :\ P \text{ is a randomized protocol for } f \text{ with } \mathrm{err}(P) \leq \varepsilon\}\,;\\
\mathrm{D}_{\mu,\varepsilon}(f) \ &= \ \min\{\mathrm{cost}(P) :\ P \text{ is a deterministic protocol for } f \text{ with } \mathrm{err}_\mu(P) \leq \varepsilon\}\,.
\end{aligned}$$

We also put $\mathrm{R}(f) = \mathrm{R}_{1/3}(f)$ and $\mathrm{D}_\mu(f) = \mathrm{D}_{\mu,1/3}(f)$.

2.2 Rectangles and Corruption

Consider a two-player communication problem given by a function $f : X \times Y \to Z$. A set $R \subseteq X \times Y$ is said to be a rectangle if $R = \mathcal{X} \times \mathcal{Y}$ for some $\mathcal{X} \subseteq X$ and $\mathcal{Y} \subseteq Y$. A fundamental property of communication protocols is the following.

Fact 2.1 (Rectangle property; see, e.g., [KN97])**.** *Let P be a deterministic communication protocol that takes inputs in $X \times Y$, produces an output in Z, and communicates c bits. Then, for all $z \in Z$, there exist 2^c pairwise disjoint rectangles $R_{1,z}, \ldots, R_{2^c,z}$ such that*

$$\forall\,(x,y) \in X \times Y :\ P(x,y) = z \iff (x,y) \in \textstyle\bigcup_{i=1}^{2^c} R_{i,z}\,.$$

The rectangles $R_{1,z}, \ldots, R_{2^c,z}$ are called the z-rectangles of P.

Let us focus on problems with Boolean output, i.e., $Z = \{0,1\}$. The *discrepancy method* for proving lower bounds on $\mathrm{R}(f)$ consists of choosing a suitable distribution μ on $X \times Y$ and showing that for every rectangle R, the quantity $|\mu(R \cap f^{-1}(0)) - \mu(R \cap f^{-1}(1))|$ is "exponentially" small. For several functions, this method is unable to prove a strong enough lower bound; the canonical example is DISJ. A generalization that handles DISJ, and several other functions, is the *corruption method* [Raz92, Kla03, BPSW06] which consists of showing, instead, that for every "large" rectangle R, we have $\alpha\mu_1(R) \leq \mu_0(R)$, for a constant $\alpha > 0$, where μ_i is a probability distribution on $R \cap f^{-1}(i)$, for $i \in \{0,1\}$. Intuitively, we are arguing that any large rectangle that contains many 1s must be corrupted by the presence of many 0s. The largeness of R is often enforced indirectly by writing the inequality in the following manner, where m typically grows with $|X|$ and $|Y|$:

$$\exists\, \alpha_0, \alpha_1 > 0 \;\; \forall\, R \text{ rectangular}: \;\; \alpha_1\mu_1(R) \;\leq\; \alpha_0\mu_0(R) + 2^{-m}\,. \tag{1}$$

An inequality of this form allows us to conclude an $\Omega(m)$ lower bound on $\mathrm{D}_{\nu,\varepsilon}(f)$ for a suitable distribution ν and sufficiently small error $\varepsilon > 0$. (Rather than present a full proof, we note that this follows as a special case of Theorem 2.2, below.) By the easy direction of Yao's lemma, this implies $\mathrm{R}_\varepsilon(f) = \Omega(m)$.

2.3 Corruption With Jokers, and the Smooth Rectangle Bound

We now introduce a suitable generalization of the corruption method, which, as we shall soon see, implies that $\mathrm{D}_{\mu,\varepsilon}(\mathrm{GHD}) = \Omega(n)$, for suitable μ and ε. The corresponding technical challenge is met using a new Gaussian noise correlation inequality that we prove in Section 3. Our generalization can be captured within the very recent *smooth rectangle bound* framework [Kla10, JK10]. However, we believe that there is merit in singling out the method we use, because it appears wieldier than the smooth rectangle bound, which is more technically involved.

The key idea is that, in addition to the distributions μ_0 and μ_1 on the 0-inputs and 1-inputs to f, we consider an auxiliary distribution μ_+ on "joker" inputs. Strictly speaking, we just have a "joker distribution" μ_+,[2] and it does not matter how μ_+ relates to μ_0 and μ_1, but it is crucial that the inequality below gives a negative weight to μ_+, and is therefore a weakening of (1).

$$\alpha_1\mu_1(R) - \alpha_+\mu_+(R) \;\leq\; \alpha_0\mu_0(R) + 2^{-m}\,. \tag{2}$$

We shall in fact allow a little flexibility in our choice of μ_0 and μ_1 by requiring only that these be supported "mostly" on 0-inputs and 1-inputs. Also, we shall extend our theory to partial functions, since GHD is one. The next theorem captures our lower bound technique.

Theorem 2.2. *For all $\alpha_0, \alpha_1, \alpha_+, \varepsilon > 0$ such that $\varepsilon < (\alpha_1 - \alpha_+)/(\alpha_0 + \alpha_1)$, there exist $\beta \in \mathbb{R}$ and $\varepsilon' > 0$ such that the following holds. Let $f : X \times Y \to \{0,1,\star\}$ be a partial function. Let $A_0 = f^{-1}(0)$ and $A_1 = f^{-1}(1)$. Suppose that there exist distributions μ_0, μ_1, μ_+ on $X \times Y$, and a real number $m > 0$ such that*

(1) for $i \in \{0,1\}$, μ_i is mostly supported on A_i, i.e., $\mu_i(A_i) \geq 1 - \varepsilon$, and

(2) inequality (2) holds for all rectangles $R \subseteq X \times Y$.

Then, for the distribution $\nu := (\alpha_0\mu_0 + \alpha_1\mu_1)/(\alpha_0 + \alpha_1)$, we have $\mathrm{D}_{\nu,\varepsilon'}(f) \geq m + \beta$. In particular, we have $\mathrm{R}_{\varepsilon'}(f) \geq m + \beta$.

[2] In the sequel, when we apply the technique to GHD, μ_0, μ_1 and μ_+ will be sharply concentrated on pairwise disjoint sets of inputs, which we can think of as the interesting 0-inputs, the interesting 1-inputs, and the joker inputs, respectively.

Proof. Consider a deterministic protocol P that computes f with some error ε' (to be fixed later) under ν, and uses c bits of communication. Let $R_1, \ldots, R_{2^c} \subseteq X \times Y$ be the disjoint 1-rectangles of P, as given by Fact 2.1. Let $S_1 = \bigcup_{i=1}^{2^c} R_i$ and $S_0 = X \times Y \setminus S_1$. Notice that S_i is exactly the set of inputs on which P outputs i. Thus, for $i \in \{0,1\}$, we have

$$\begin{aligned} \mathrm{err}_{\mu_i}(P) &= \mu_i(S_i \cap A_{1-i}) + \mu_i(S_{1-i} \cap A_i) \\ &\geq \mu_i(S_{1-i} \cap A_i) \\ &\geq \mu_i(S_{1-i}) - \varepsilon\,, \end{aligned} \tag{3}$$

where the last step uses Condition (1).

Instantiating inequality (2) with each R_i and summing the resulting inequalities, we get

$$\alpha_1\mu_1(S_1) - \alpha_+\mu_+(S_1) \;\leq\; \alpha_0\mu_0(S_1) + 2^c \cdot 2^{-m}\,. \tag{4}$$

Noting that $\mu_1(S_1) = 1 - \mu_1(S_0)$ and applying (3) to the μ_0 and μ_1 terms in (4), we obtain

$$\alpha_1(1 - \mathrm{err}_{\mu_1}(P) - \varepsilon) - \alpha_+\mu_+(S_1) \;\leq\; \alpha_0(\mathrm{err}_{\mu_0}(P) + \varepsilon) + 2^{c-m}\,.$$

Further, noting that $\mu_+(S_1) \leq 1$, and rearranging terms, we obtain

$$\begin{aligned} \alpha_1 - \alpha_+ &\leq (\alpha_0 + \alpha_1)\varepsilon + (\alpha_0 \cdot \mathrm{err}_{\mu_0}(P) + \alpha_1 \cdot \mathrm{err}_{\mu_1}(P)) + 2^{c-m} \\ &= (\alpha_0 + \alpha_1)\varepsilon + (\alpha_0 + \alpha_1)\,\mathrm{err}_\nu(P) + 2^{c-m}\,. \end{aligned}$$

Using $\mathrm{err}_\nu(P) \leq \varepsilon'$ and rearranging further, we get

$$2^{c-m} \;\geq\; \alpha_1 - \alpha_+ - (\alpha_0 + \alpha_1)(\varepsilon + \varepsilon')\,.$$

By virtue of the upper bound on ε, we may choose ε' small enough to make the right-hand side of the above inequality positive, and equal to 2^β, say. Doing so gives us $c \geq m + \beta$, as desired.

Notice that the "hard distribution" ν is explicitly specified, once the distributions involved in Condition (2) are made explicit. □

We could, alternately, have proved Theorem 2.2 by demonstrating that the given conditions imply that the smooth rectangle bound of f is $\Omega(m)$. We have chosen to give the above proof instead, because it is more elementary, avoiding the technical details of the latter bound, and because it was discovered independently by the first named author.

2.4 Application to GHD: the Main Theorem

The Gap-Hamming-Distance problem is formalized as the computation of the partial function $\mathrm{GHD}_{n,t,g} : \{0,1\}^n \times \{0,1\}^n \to \{0,1,\star\}$ defined as follows.

$$\mathrm{GHD}_{n,t,g}(x,y) \;=\; \begin{cases} 0\,, & \text{if } \Delta(x,y) \leq t - g\,, \\ 1\,, & \text{if } \Delta(x,y) > t + g\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

It will be useful to have some flexibility in the choice of the location of the threshold, t, and the size of the gap, g. It is not hard to see that all settings with $t \in \Omega(n) \cap (n - \Omega(n))$ and $g = \Theta(\sqrt{n})$ lead to "equally hard" problems, asymptotically; we prove this formally in Lemma 4.2.

Rather than working with $\mathrm{GHD}_{n,n/2,\sqrt{n}}$ directly, it proves convenient to consider the partial function $f_b = \mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{n}}$, for some large enough constant b to be determined later. We shall now come up with distributions and constants that satisfy the conditions of Theorem 2.2: Condition (1) turns out to be easy to verify, and verifying Condition (2), as mentioned above, is a significant technical challenge that we deal with in Section 3.

Definition 2.3. For $p \in [-1,1]$, let ξ_p denote the distribution of $(x,y) \in \{0,1\}^n \times \{0,1\}^n$ defined by the following randomized procedure: pick $x \in_R \{0,1\}^n$ uniformly at random, and then pick y by independently flipping each bit of x with probability $(1-p)/2$. Notice that ξ_0 is the uniform distribution on $\{0,1\}^n \times \{0,1\}^n$.

We shall need the following two lemmas. The first of these follows easily from standard tail estimates for the binomial distribution; we omit its proof. The second is formally proved at the end of Section 3.

Lemma 2.4. *For all $\varepsilon > 0$ there exists $b > 0$ such that, for large enough n, we have*

$$\begin{aligned} \xi_{2b/\sqrt{n}}(A_0) &= \Pr_{(x,y)\sim\xi_{2b/\sqrt{n}}}\left[\Delta(x,y) \le \frac{n}{2} - (b+1)\sqrt{n}\right] \ge 1-\varepsilon, \text{ and} \\ \xi_0(A_1) &= \Pr_{(x,y)\sim\xi_0}\left[\Delta(x,y) \ge \frac{n}{2} - (b-1)\sqrt{n}\right] \ge 1-\varepsilon, \end{aligned}$$

where $A_0 = f_b^{-1}(0)$ and $A_1 = f_b^{-1}(1)$. □

Lemma 2.5. *For all $b > 0$ there exists $\delta > 0$ such that, for large enough n, we have*

$$\forall\, R \subseteq \{0,1\}^n \times \{0,1\}^n \text{ rectangular}: \ \tfrac{1}{2}\left(\xi_{-2b/\sqrt{n}}(R) + \xi_{2b/\sqrt{n}}(R)\right) \ge \tfrac{2}{3}\xi_0(R) - 2^{-\delta n}.$$

To derive the lower bound on $\mathrm{R}(\mathrm{GHD})$, we put $m = \delta n$, $\mu_0 = \xi_{2b/\sqrt{n}}$, $\mu_1 = \xi_0$, $\mu_+ = \xi_{-2b/\sqrt{n}}$, $\varepsilon = \frac{1}{8}$, $\alpha_1 = \frac{2}{3}$, and $\alpha_0 = \alpha_+ = \frac{1}{2}$. Note that this choice of constants satisfies $\varepsilon < (\alpha_1 - \alpha_+)/(\alpha_0 + \alpha_1)$. By Lemmas 2.4 and 2.5, we see that Conditions (1) and (2), respectively, of Theorem 2.2 are met; the inequality in Lemma 2.5 is easily seen to be the corresponding instantiation of (2).

Thus, applying Theorem 2.2, we conclude that there exist absolute constants $\varepsilon', \delta, b > 0$ and $\beta \in \mathbb{R}$ such that, for large enough n, we have $\mathrm{R}_{\varepsilon'}(f_b) \ge \delta n + \beta$. Combining this with Lemma 4.2 to adjust for the slightly off-center threshold and the size of the gap, and applying standard error reduction techniques, we obtain the following asymptotically optimal lower bound for GHD.

Theorem 2.6 (Main Theorem)**.** $\mathrm{R}(\mathrm{GHD}_{n,n/2,\sqrt{n}}) = \Omega(n)$. □

3 An Inequality on Correlation under Gaussian Noise

We now turn to the proof of Lemma 2.5, for which we need some technical machinery that we now develop. We begin with some preliminaries.

Some Probability Distributions. Let μ denote the uniform (Haar) distribution on $\mathbb{S}^{n-1}$, the unit sphere in $\mathbb{R}^n$. Let γ denote the standard Gaussian distribution on $\mathbb{R}$, with density function $(2\pi)^{-1/2}\mathrm{e}^{-x^2/2}$, and let γ^n denote the n-dimensional standard Gaussian distribution with density $(2\pi)^{-n/2}\mathrm{e}^{-\|x\|^2/2}$. For a set $A \subseteq \mathbb{R}^n$, when we write, e.g., $\gamma^n(A)$, we tacitly assume that A is measurable. For a set $A \subseteq \mathbb{R}^n$ we denote by $\gamma^n|_A$ the distribution γ^n conditioned on being in A. We

say that a pair (x, y) is an *η-correlated Gaussian pair* if its distribution is that obtained by choosing x from γ^n and then setting $y = \eta x + \sqrt{1-\eta^2}z$ where z is an independent sample from γ^n. It is easy to verify that if (x, y) is an η-correlated Gaussian pair, then so is (y, x); in particular, y is distributed as γ^n.

Relative Entropy. We recall some basic information theory for continuous probability distributions. For clarity, we eschew a fully rigorous treatment — which would introduce a considerable amount of extra complexity through its formalism — and instead refer the interested reader to the textbook of Gray [Gra90]. Given two probability distributions P and Q on a space Ω, we define the *relative entropy* of P with respect to Q as

$$D(P \,\|\, Q) \;=\; \int P(x) \ln(P(x)/Q(x)) \,\mathrm{d}x\,.$$

It is well known (and not difficult to show) that the relative entropy is always nonnegative and is zero iff the two distributions are essentially equal. We will also need Pinsker's inequality, which says that the statistical distance between two distributions P and Q is at most $\sqrt{D(P \,\|\, Q)/2}$ (see, e.g., [Gra90, Lemma 5.2.8]). Since we will only consider the relative entropy with respect to the Gaussian distribution, we introduce the notation

$$D_\gamma(X) \;:=\; D(P \,\|\, \gamma)$$

where X is a real-valued random variable with distribution P. We define D_{γ^n} similarly. These quantities can be thought of as measuring the "distance from Gaussianity." They can be seen, in some precise sense, as additive inverses of entropy, and as such satisfy many of the familiar properties of entropy. For two random variables X and Y, we use the notation $D_\gamma(X|Y)$ to denote the expectation over Y of the distance from Gaussianity of $X|Y$.

3.1 Projections of Sets in Gaussian Space

Our main technical result is a statement about the projections of sets in Gaussian space. More precisely, let $A \subseteq \mathbb{R}^n$ be any set of not too small measure, say, $\gamma^n(A) \geq \exp(-\delta n)$ for some constant $\delta > 0$. What can we say about the projections (or one-dimensional marginals) of $\gamma^n|_A$, i.e., the set of distributions of $\langle \gamma^n|_A, y\rangle$ as the (fixed) vector y ranges over the unit sphere $\mathbb{S}^{n-1}$?

Related questions have appeared in the literature. The first is in work by Sudakov [Sud78] and Diaconis and Freedman [DF84] (see also [Bob03] for a more recent exposition) who showed that for any random variable in $\mathbb{R}^n$ with zero mean and identity covariance matrix whose norm is concentrated around $\sqrt{n}$, almost all its projections are close to the standard normal distribution. A second related result is by Klartag [Kla07] who, building on the previous result but with considerable additional work, showed that almost all projections of the uniform distribution over a (properly normalized) convex body are close to the standard normal distribution. (For the special case of the cube $[-1, 1]^n$, this essentially follows from the central limit theorem.)

Our setting is different as we do not put any restrictions on the set A (such as convexity) apart from its measure not being too small (and clearly without any requirement on the measure one cannot say anything about its projections). Another important difference is that in our setting the projections are *not* necessarily normal. To see why, take $A = \{x : |x_1| > t\}$ for $t \approx \sqrt{\delta n}$, a set with Gaussian measure roughly $\exp(-\delta n)$, half of which is on vectors with $x_1 \approx t$ and the other half on vectors with $x_1 \approx -t$. It follows that the projection of $\gamma^n|_A$ on a unit vector y is distributed more or less like the mixture of two normal variables, one centered around ty_1 and the other centered

around $-ty_1$, both with variance 1. For unit vectors y with $|y_1| \geq 1/\sqrt{\delta n}$ (a set of measure about $\exp(-1/\delta)$), this distribution is very far from any normal distribution.

Our main theorem below shows that the general situation is similar: for any set A of not too small measure, almost all projections of $\gamma^n|_A$ are close to being mixtures of translated normal variables of variance 1. One implication of this (which is essentially all we will use later) is that for any $A \subseteq \mathbb{R}^n$ of not too small measure, and $B \subseteq \mathbb{S}^{n-1}$ whose measure is also not too small, the inner product $\langle x, y \rangle$ for x chosen from $\gamma^n|_A$ and y chosen uniformly from B is not too concentrated around 0; in fact, it must be at least as "spread out" as γ (and possibly much more).

Theorem 3.1. *For all $\varepsilon, \delta > 0$ and large enough n, the following holds. Let $A \subseteq \mathbb{R}^n$ be such that $\gamma^n(A) \geq \mathrm{e}^{-\varepsilon^2 n}$. Then, for all but an $\mathrm{e}^{-\delta n/36}$ measure of unit vectors $y \in \mathbb{S}^{n-1}$, the distribution of $\langle x, y \rangle$ where $x \sim \gamma^n|_A$ is equal to the distribution of $\alpha X + Y$ for some $1 - \delta \leq \alpha \leq 1$ and random variables X and Y satisfying*

$$D_\gamma(X|Y) \;\leq\; \varepsilon\,.$$

The proof is based on the following two lemmas. The first one below shows that for any set A whose measure is not too small, and any orthonormal basis, most of the projections of $\gamma^n|_A$ on the basis vectors are close to normal. In fact, the statement is somewhat stronger, as it allows us to condition on previous projections (and this will be crucially used).

Lemma 3.2. *For all $\varepsilon > 0$ and large enough n the following holds. For all sets $A \subseteq \mathbb{R}^n$ with $\gamma^n(A) \geq \mathrm{e}^{-\varepsilon^2 n}$ and all orthonormal bases $y_1, \ldots, y_n$, at least a $1 - \varepsilon$ fraction of the indices $k \in [n]$ satisfy*

$$D_\gamma(P_k|P_1, \ldots, P_{k-1}) \;\leq\; \varepsilon\,,$$

where $P_i = \langle u, y_i \rangle$, with $u \sim \gamma^n|_A$.

Proof. By definition, $D_{\gamma^n}(\gamma^n|_A) = -\ln \gamma^n(A) \leq \varepsilon^2 n$. Thus, since $(P_1, \ldots, P_n)$ is the vector u written in the orthonormal basis $y_1, \ldots, y_n$, using the chain rule for relative entropy, we have

$$\varepsilon^2 n \;\geq\; D_{\gamma^n}(\gamma^n|_A) \;=\; D_{\gamma^n}(P_1, \ldots, P_n) \;=\; \sum_{k=1}^{n} D_\gamma(P_k|P_1, \ldots, P_{k-1})\,.$$

Hence, for at least a $1 - \varepsilon$ fraction of indices k, we have $D_\gamma(P_k|P_1, \ldots, P_{k-1}) \leq \varepsilon$. □

The second lemma is due to Raz [Raz99] and shows that any not-too-small subset of the sphere contains $n/2$ "nearly orthogonal" vectors. The idea of Raz's proof is the following. First, a simple averaging argument shows that there is a not-too-small measure of vectors $y' \in \mathbb{S}^{n-1}$ satisfying the property that the measure of B inside the unit sphere formed by the intersection of $\mathbb{S}^{n-1}$ and the subspace orthogonal to y' is not much smaller than $\mu(B)$. Second, by the isoperimetric inequality, almost all vectors in $\mathbb{S}^{n-1}$ are within distance δ of B. Together, we obtain a vector y' as above that is within distance δ of B. We take $y_{n/2}$ to be the closest vector in B to y' and repeat the argument recursively with the intersection of B and the subspace orthogonal to y'.

Definition 3.3. A sequence of unit vectors $y_1, \ldots, y_k \in \mathbb{S}^{n-1}$ is called *δ-orthogonal* if for all $i \in [k]$, the squared norm of the projection of y_i on $\mathrm{span}(y_1, \ldots, y_{i-1})$ is at most δ.

Lemma 3.4 ([Raz99, Lemma 4.4])**.** *For all $\delta > 0$ and large enough n, the following holds. Every $B \subseteq \mathbb{S}^{n-1}$ of Haar measure $\mu(B) \geq \mathrm{e}^{-\delta n/36}$ contains a δ-orthogonal sequence $y_1, \ldots, y_{n/2} \in B$.*

Proof of Theorem 3.1. Let $B \subseteq \mathbb{S}^{n-1}$ be an arbitrary set of unit vectors of measure at least $\mathrm{e}^{-\delta n/36}$. We will prove the theorem by showing that at least one vector $y \in B$ satisfies the condition stated in the theorem.

By Lemma 3.4, there is a sequence of $n/2$ vectors $y_1, \ldots, y_{n/2} \in B$ that is δ-orthogonal. Let $y_1^*, \ldots, y_{n/2}^*$ be their Gram-Schmidt orthogonalization, i.e., each y_k^* is defined to be the projection of y_k on the space orthogonal to $\mathrm{span}(y_1, \ldots, y_{k-1})$. Notice that, by definition, we can write each y_k as

$$y_k \;=\; y_k^* + \sum_{i=1}^{k-1} \alpha_{k,i} y_i^*$$

for some real coefficients $\alpha_{k,i}$. Moreover, by assumption, $\|y_k^*\|^2 \geq 1 - \delta$.

Let $P_1, \ldots, P_{n/2}$ be the random variables representing $\langle x, y_1^*/\|y_1^*\| \rangle, \ldots, \langle x, y_{n/2}^*/\|y_{n/2}^*\| \rangle$ when x is chosen from $\gamma^n|_A$. By applying Lemma 3.2 to any completion of $y_1^*/\|y_1^*\|, \ldots, y_{n/2}^*/\|y_{n/2}^*\|$ to an orthonormal basis, we see that there exists an index $k \in [n/2]$ for which

$$D_\gamma(P_k | P_1, \ldots, P_{k-1}) \;\leq\; \varepsilon\,.$$

(In fact, at least $1 - 2\varepsilon$ of the indices k satisfy this.) It remains to notice that we can write $\langle x, y_k \rangle$ as

$$\|y_k^*\| P_k + \sum_{i=1}^{k-1} \alpha_{k,i} \|y_i^*\| P_i\,,$$

which satisfies the condition in the theorem, with X taken to be P_k and Y taken to be the above sum. Here we are using the fact that Y is a function of $P_1, \ldots, P_{k-1}$, which implies that $D_\gamma(X|Y) \leq D_\gamma(P_k | P_1, \ldots, P_{k-1})$ since conditioning cannot decrease relative entropy. □

3.2 The Correlation Inequality

We now turn to our main technical result, which is given by the following theorem.

Theorem 3.5. *For all $c, \varepsilon > 0$ there exists a $\delta > 0$ such that for all large enough n and $0 \leq \eta \leq c/\sqrt{n}$ the following holds. For all sets $A, B \subseteq \mathbb{R}^n$ with $\gamma^n(A), \gamma^n(B) \geq \mathrm{e}^{-\delta n}$ we have that*

$$\frac{1}{2}\left(\Pr_{(x,y) \text{ is } \eta\text{-correlated}}[x \in A \wedge y \in B] + \Pr_{(x,y) \text{ is } -\eta\text{-correlated}}[x \in A \wedge y \in B] \right) \;\geq\; (1-\varepsilon)\gamma^n(A)\gamma^n(B).$$

As will become evident in the proof, pairs $(x, y) \in A \times B$ for which $|\langle x, y \rangle|$ is small contribute much less to the left hand side than to the right hand side. Hence the theorem essentially amounts to showing that $\langle x, y \rangle$ is not too concentrated around zero, and precisely such an anti-concentration statement is given by Theorem 3.1.

We point out the following easy corollary (which is in fact equivalent to Theorem 3.5).

Corollary 3.6. *For all $c, \varepsilon > 0$ there exists a $\delta > 0$ such that for all large enough n and $0 \leq \eta \leq c/\sqrt{n}$ the following holds. For any sets $A, B \subseteq \mathbb{R}^n$ with $\gamma^n(A), \gamma^n(B) \geq \mathrm{e}^{-\delta n}$ where A (or B) is centrally symmetric (i.e., $A = -A$) we have that*

$$\Pr_{(x,y) \text{ is } \eta\text{-correlated}}[x \in A \wedge y \in B] \;\geq\; (1-\varepsilon)\gamma^n(A)\gamma^n(B)\,.$$

Remark. Without the symmetry assumption, this probability can be considerably smaller. For instance, take A and B to be two opposing half-spaces, i.e., $A = \{x : x_1 < -t\}$ and $B = \{x : x_1 > t\}$ for $t \approx \sqrt{\delta n}$. Then for $\eta = c/\sqrt{n}$, the probability above can be seen to be $\mathrm{e}^{-\Theta(\sqrt{n})}\gamma^n(A)\gamma^n(B)$. In fact, C. Borell [Bor85] showed that for any given $\gamma^n(A), \gamma^n(B)$ and any $0 \le \eta \le 1$, two opposing half-spaces A, B of the corresponding measures exactly achieve the minimum of the probability above. It would be interesting to obtain a strengthening of Corollary 3.6 of a similar tight nature. See [Bar01] for a short related discussion.

Recall that $\cosh(x) := \frac{1}{2}(\mathrm{e}^x + \mathrm{e}^{-x})$. The following technical claim shows that if the distribution of x is close to the normal distribution (in relative entropy) then the expectation of $\cosh(\alpha x + z)$ is at least $\mathrm{e}^{\alpha^2/2} - \varepsilon$. Notice that if x is normal, this expectation is

$$\mathbb{E}_{x\sim\gamma}[\cosh(\alpha x + z)] \;=\; \cosh(z)\,\mathbb{E}_{x\sim\gamma}[\cosh(\alpha x)] \;=\; \cosh(z)\,\mathrm{e}^{\alpha^2/2} \ge \mathrm{e}^{\alpha^2/2},$$

where in the first equality we used the symmetry of γ.

Claim 3.7. *For all $\varepsilon, \alpha_0 > 0$ there exists a $\delta > 0$ such that for any probability distribution P on the reals satisfying $D(P \,\|\, \gamma) < \delta$, any $z \in \mathbb{R}$, and any $0 < \alpha \le \alpha_0$, we have*

$$\mathbb{E}_{x\sim P}[\cosh(\alpha x + z)] \;\ge\; \mathrm{e}^{\alpha^2/2} - \varepsilon.$$

Proof. Let $M = M(\varepsilon, \alpha_0)$ be big enough so that for all z and all $\alpha \le \alpha_0$,

$$\begin{aligned}\mathbb{E}_{x\sim\gamma}[\min(\cosh(\alpha x + z), M)] &= \mathbb{E}_{x\sim\gamma}[\min(\cosh(z)\cosh(\alpha x), M)] \\ &\ge \mathbb{E}_{x\sim\gamma}[\min(\cosh(\alpha x), M)] \\ &\ge \mathrm{e}^{\alpha^2/2} - \varepsilon/2\,.\end{aligned}$$

Such an M exists because

$$\mathbb{E}_{x\sim\gamma}[\cosh(\alpha x)^2] \;=\; \mathbb{E}_{x\sim\gamma}\left[\frac{1}{2}(1+\cosh(2\alpha x))\right] \;\le\; \mathbb{E}_{x\sim\gamma}\left[\frac{1}{2}(1+\cosh(2\alpha_0 x))\right]$$

and the latter expectation is finite. Next, since the statistical distance between P and γ is at most $\sqrt{2D(P\,\|\,\gamma)} < \sqrt{2\delta}$, we have that

$$\mathbb{E}_{x\sim P}[\cosh(\alpha x + z)] \;\ge\; \mathbb{E}_{x\sim P}[\min(\cosh(\alpha x + z), M)] \;\ge\; \mathrm{e}^{\alpha^2/2} - \varepsilon/2 - M\sqrt{2\delta} \;\ge\; \mathrm{e}^{\alpha^2/2} - \varepsilon$$

for small enough $\delta > 0$. □

Proof of Theorem 3.5. Let $\beta_1, \beta_2, \beta_3, \beta_4 > 0$ be small enough constants (depending only on c and ε) to be determined later. By choosing a small enough δ, we can guarantee that A', defined as

$$A' \;=\; \{x \in A :\, (1-\beta_1)n \le \|x\|^2 \le (1+\beta_1)n\}\,,$$

satisfies $\gamma^n(A') \ge \gamma^n(A) - \beta_2\mathrm{e}^{-\delta n} \ge (1-\beta_2)\gamma^n(A)$ and similarly for B'. We can write

$$\begin{aligned}&\Pr_{(x,y)\text{ is }\eta\text{-correlated}}[x \in A \wedge y \in B] \\ &\quad\ge \Pr_{(x,y)\text{ is }\eta\text{-correlated}}[x \in A' \wedge y \in B'] \\ &\quad= (2\pi)^{-n/2}(2\pi(1-\eta^2))^{-n/2}\int 1_{A'}(x)1_{B'}(y)\mathrm{e}^{-\|x\|^2/2}\mathrm{e}^{-\|y-\eta x\|^2/2(1-\eta^2)}\mathrm{d}x\mathrm{d}y \\ &\quad= (1-\eta^2)^{-n/2}\,\mathbb{E}_{x,y\sim\gamma^n}\left[1_{A'}(x)1_{B'}(y)\mathrm{e}^{-\eta^2\|x\|^2/2(1-\eta^2)}\mathrm{e}^{-\eta^2\|y\|^2/2(1-\eta^2)}\mathrm{e}^{\eta\langle x,y\rangle/(1-\eta^2)}\right] \\ &\quad= (1-\eta^2)^{-n/2}\,\mathbb{E}_{x\sim\gamma^n|_{A'},y\sim\gamma^n|_{B'}}\left[\mathrm{e}^{-\eta^2\|x\|^2/2(1-\eta^2)}\mathrm{e}^{-\eta^2\|y\|^2/2(1-\eta^2)}\mathrm{e}^{\eta\langle x,y\rangle/(1-\eta^2)}\right]\gamma^n(A')\gamma^n(B') \\ &\quad\ge (1-\eta^2)^{-n/2}\,\mathrm{e}^{-\eta^2(1+\beta_1)n/(1-\eta^2)}\,\mathbb{E}_{x\sim\gamma^n|_{A'},y\sim\gamma^n|_{B'}}\left[\mathrm{e}^{\eta\langle x,y\rangle/(1-\eta^2)}\right]\gamma^n(A')\gamma^n(B')\,.\end{aligned}$$

By averaging this inequality with the analogous one for $-\eta$ and recalling the definition of cosh, we obtain that the expression we wish to bound is at least

$$(1-\eta^2)^{-n/2}\,\mathrm{e}^{-\eta^2(1+\beta_1)n/(1-\eta^2)}\,\mathbb{E}_{x\sim\gamma^n|_{A'},y\sim\gamma^n|_{B'}}[\,\cosh(\eta\langle x,y\rangle/(1-\eta^2))\,]\gamma^n(A')\gamma^n(B')\,. \qquad (5)$$

Let $B'' \subseteq B'$ be the set of all $y \in B'$ for which

$$\mathbb{E}_{x\sim\gamma^n|_{A'}}[\,\cosh(\eta\langle x,y\rangle/(1-\eta^2))\,] \;\le\; (1-\beta_3)\mathrm{e}^{(\eta/(1-\eta^2))^2(1-\beta_1)n/2}\,.$$

We can now complete the proof by showing that $\gamma^n(B'') \le \beta_4\gamma^n(B')$, since this would imply that (5) is at least

$$\begin{aligned}&(1-\eta^2)^{-n/2}\,\mathrm{e}^{-\eta^2(1+\beta_1)n/(1-\eta^2)}(1-\beta_4)(1-\beta_3)\,\mathrm{e}^{(\eta/(1-\eta^2))^2(1-\beta_1)n/2}\,\gamma^n(A')\gamma^n(B')\\ &\;\ge\; (1-\varepsilon)\gamma^n(A)\gamma^n(B)\,,\end{aligned}$$

assuming β_1, β_2, β_3 and β_4 are chosen to be sufficiently small and n is large enough.

In order to complete the proof, assume by contradiction that $\gamma^n(B'') > \beta_4\gamma^n(B') \ge \beta_4(1-\beta_2)\mathrm{e}^{-\delta n}$. Let $\beta_5, \beta_6, \beta_7 > 0$ be small enough constants to be determined later. Let $\sqrt{(1-\beta_1)n} \le r \le \sqrt{(1+\beta_1)n}$ be such that the Haar measure $\mu((r\,\mathbb{S}^{n-1} \cap B'')/r)$ of points in B'' of norm r is at least $\gamma^n(B'')$. (The existence of such an r follows from the spherical symmetry of the Gaussian distribution.) We now apply Theorem 3.1 with ε taken to be β_5, δ taken to be β_6, and A taken to be A'. By taking (our) δ to be small enough, we obtain a vector $y \in B''$ for which the distribution of $\langle x,y\rangle$ where $x \sim \gamma^n|_{A'}$ is given by the distribution of $\alpha r X + rY$ for some $1-\beta_6 \le \alpha \le 1$ and random variables X and Y satisfying

$$D_\gamma(X|Y) \;\le\; \beta_5\,.$$

In particular, we have

$$\Pr_Y[\,D(X|Y\,\|\,\gamma) \le \sqrt{\beta_5}\,] \;\ge\; 1-\sqrt{\beta_5}\,.$$

Claim 3.7 now implies that

$$\begin{aligned}\mathbb{E}_{x\sim\gamma^n|_{A'}}[\,\cosh(\eta\langle x,y\rangle/(1-\eta^2))\,] \;&=\; \mathbb{E}[\,\cosh(\eta/(1-\eta^2)(\alpha r X + rY))\,]\\ &\ge\; (1-\sqrt{\beta_5})(\mathrm{e}^{(\eta/(1-\eta^2)\alpha r)^2/2}-\beta_7)\\ &\ge\; (1-\sqrt{\beta_5})(\mathrm{e}^{(\eta/(1-\eta^2))^2(1-\beta_6)^2(1-\beta_1)n/2}-\beta_7)\\ &>\; (1-\beta_3)\,\mathrm{e}^{(\eta/(1-\eta^2))^2(1-\beta_1)n/2}\,,\end{aligned}$$

assuming β_5, β_6 and β_7 are sufficiently small, in contradiction to the assumption that $y \in B''$. □

3.3 Corollary for the Boolean cube

The Gaussian noise correlation inequality we have just proved implies a similar statement for the Boolean cube, from which Lemma 2.5 follows easily. The statement involves the distribution ξ_p from Definition 2.3.

Corollary 3.8 (Stronger variant of Lemma 2.5)**.** *For all $c, \varepsilon > 0$ there exists a $\delta > 0$ such that for all large enough n and $0 \le p \le c/\sqrt{n}$ the following holds. For all sets $A, B \subseteq \{0,1\}^n$ with $|A|, |B| \ge 2^{(1-\delta)n}$, we have that*

$$\tfrac{1}{2}\left(\xi_{-p}(A\times B) + \xi_p(A\times B)\right) \;\ge\; (1-\varepsilon)\,\xi_0(A\times B)\,.$$

To derive Lemma 2.5, take $R = A \times B$, $\varepsilon = \frac{1}{3}$, and observe that if $\min\{|A|, |B|\} < 2^{(1-\delta)n}$, then $\xi_0(R) < 2^{-\delta n}$ and the inequality in that lemma holds trivially, because its right-hand side is negative.

A short calculation shows that the inequality in Corollary 3.8 is equivalent to

$$(1-p^2)^{n/2}\, \mathbb{E}_{x\in A, y\in B}\left[\, \cosh\left(\ln\left(\frac{1+p}{1-p}\right)\cdot(\Delta(x,y)-n/2)\right)\right] \;\geq\; 1-\varepsilon\,.$$

Hence the corollary can be interpreted as an anti-concentration statement, saying that for sets A, B that are not too small, the Hamming distance $\Delta(x,y)$ between $x \in_R A$ and $y \in_R B$ cannot be too concentrated around $n/2$. The quantification is delicate. Notice that already for sets of size $2^{n/2}$ this is no longer the case: take, for instance, the sets $A = \{0^{n/2}x : x \in \{0,1\}^{n/2} \text{ and } |x| = n/4\}$ and $B = \{x0^{n/2} : x \in \{0,1\}^{n/2} \text{ and } |x| = n/4\}$.

Proof. Given any $A, B \subseteq \{0,1\}^n$, define

$$A' \;=\; \{x \in \mathbb{R}^n : \operatorname{sign}(x) \in A\}$$

where $\operatorname{sign}(x) \in \{0,1\}^n$ is the vector indicating the sign of each coordinate of x, and define B' similarly. Then it is easy to check that $\gamma^n(A') = |A|/2^n$ and $\gamma^n(B') = |B|/2^n$, so that $\gamma^n(A')\gamma^n(B') = \xi_0(A \times B)$, and that for all η,

$$\Pr_{(x,y) \text{ is } \eta\text{-correlated}}[x \in A' \wedge y \in B'] \;=\; \xi_p(A \times B)$$

for $p = 1 - \frac{2}{\pi}\arccos\eta$ (since the probability that $\operatorname{sign}(x) \neq \operatorname{sign}(y)$ when $x, y \in \mathbb{R}$ are η-correlated can be computed to be $\frac{1}{\pi}\arccos\eta$). For small p, we get $p \approx \frac{2}{\pi}\eta$, and the corollary follows from Theorem 3.5. □

4 Reductions, Related Results and Generalizations

Recall that our argument in Section 2.4 gave an $\Omega(n)$ lower bound on $\mathrm{R}(\mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{n}})$, for a certain constant b. To obtain an $\Omega(n)$ bound for GHD itself (which, we remind the reader, is shorthand for $\mathrm{GHD}_{n,n/2,\sqrt{n}}$), we use a toolkit of simple reductions, given in the next lemma. Furthermore, using the toolkit, we can generalize the GHD bound to cover most parameter settings, and using similarly simple reductions, we can obtain optimal lower bounds for related problems.

Lemma 4.1. *For all integers n, k, ℓ, m and reals $t, g, g' \in [0, n]$, with $n, k > 0$ and $g' \geq g$, the following relations hold.*

(1) $\mathrm{R}(\mathrm{GHD}_{n,t,g'}) \leq \mathrm{R}(\mathrm{GHD}_{n,t,g})$.

(2) $\mathrm{R}(\mathrm{GHD}_{n,t,g}) \leq \mathrm{R}(\mathrm{GHD}_{kn,kt,kg})$.

(3) $\mathrm{R}(\mathrm{GHD}_{n,t,g}) \leq \mathrm{R}(\mathrm{GHD}_{n+\ell+m,t+\ell,g})$.

(4) $\mathrm{R}(\mathrm{GHD}_{n,t,g}) = \mathrm{R}(\mathrm{GHD}_{n,n-t,g})$.

Proof. We give brief sketches of the proofs of these statements.

(1) A correct protocol for $\mathrm{GHD}_{n,t,g}$ is also one for $\mathrm{GHD}_{n,t,g'}$.

(2) We can solve $\mathrm{GHD}_{n,t,g}$ by having Alice and Bob "repeat" their n-bit input strings k times each — which has the effect of also amplifying the gap by a factor of k — and then simulating a protocol for $\mathrm{GHD}_{kn,kt,kg}$.

(3) We can solve $\mathrm{GHD}_{n,t,g}$ by having Alice pad her input by appending the string $0^{\ell+m}$ to it, Bob pad his by appending $1^{\ell}0^{m}$ to it, and then simulating a protocol for $\mathrm{GHD}_{n+\ell+m,t+\ell,g}$.

(4) Alice flips each bit of her input and the parties then simulate a protocol for $\mathrm{GHD}_{n,n-t,g}$. □

As promised, using parts of the above lemma, we establish the following lemma, which formally completes the proof of the main theorem.

Lemma 4.2. *For all integers $n > 0$ and reals $b > 0$, with $n/2 \geq b\sqrt{n}$, we have* $\mathrm{R}(\mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{n}}) \leq \mathrm{R}(\mathrm{GHD}_{4n,2n,\sqrt{4n}})$.

Proof. For simplicity, we assume that $b\sqrt{n}$ is an integer. Applying part (3) of Lemma 4.1, with $\ell = n/2 + b\sqrt{n}$ and $m = n/2 - b\sqrt{n}$, we have $\mathrm{R}(\mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{n}}) \leq \mathrm{R}(\mathrm{GHD}_{2n,n,\sqrt{n}})$. To complete the proof, we then apply part (2) of that lemma, with $k = 2$. □

The previous lemma can in fact be generalized, by invoking the remaining parts of Lemma 4.1, to obtain a lower bound that handles all thresholds t that are not too close to either end of the interval $[0, n]$. We omit the details, which are routine, if somewhat tedious.

Proposition 4.3. *For all reals $a \in (0, \frac{1}{2}]$ and $b > 0$, and all large enough integers n, the following holds. Let t, g be reals with $t \in [an, (1-a)n]$ and $g \leq b\sqrt{n}$. Then* $\mathrm{R}(\mathrm{GHD}_{n,t,g}) = \Omega(n)$. □

The next result resolves the randomized complexity of $\mathrm{GHD}_{n,n/2,g}$ for a general gap size, g.

Proposition 4.4. *For integers n and g, with $1 \leq g \leq n$, we have* $\mathrm{R}(\mathrm{GHD}_{n,n/2,g}) = \Theta(\min\{n, n^2/g^2\})$.

Proof. For the upper bound, consider the protocol where Alice and Bob, on input $(x, y) \in \{0,1\}^n \times \{0,1\}^n$, use public randomness to select a subset $S \subseteq [n]$ uniformly at random, from amongst all subsets of a certain size, k, compute $d = |\{i \in S : x_i \neq y_i\}|$ by brute force (say, with Alice sending Bob the bits x_i for $i \in S$), and output 0 if $d \leq k/2$ and 1 if $d > k/2$. This protocol clearly communicates k bits, and an easy application of the Chernoff bound shows that this gives a $\frac{1}{3}$-error protocol if we choose $k = O(n^2/g^2)$.

For the lower bound, we may assume that $g > \sqrt{n}$, for otherwise the claim is obviously true. Applying part (2) of Lemma 4.1 with $k = g^2/n$ (for simplicity, we ignore divisibility issues), we obtain $\mathrm{R}(\mathrm{GHD}_{n^2/g^2,n^2/2g^2,n/g}) \leq \mathrm{R}(\mathrm{GHD}_{n,n/2,g})$. The result follows by applying Theorem 2.6 to the left-hand side of this inequality. □

We remark that results similar to those for GHD also hold for GAP-INTERSECTION-SIZE, where Alice and Bob have sets $x, y \subseteq [n]$ as inputs and are required to distinguish between the cases $|x \cap y| \leq t - g$ and $|x \cap y| > t + g$, for a threshold parameter t and gap size g. Let this problem be denoted by $\mathrm{GIS}_{n,t,g}$. We then have the following result, by an easy reduction from GHD.

Proposition 4.5. *Suppose $t \in \Omega(n) \cap (n - \Omega(n))$ and $g = \Theta(\sqrt{n})$. Then* $\mathrm{R}(\mathrm{GIS}_{n,t,g}) = \Omega(n)$. □

Finally, we also remark that results similar to those for GHD also hold for the closely related (in fact, essentially equivalent) problem GAP-INNER-PRODUCT. Here, Alice and Bob have d-dimensional unit vectors x, y as inputs and are trying to distinguish between the cases $\langle x, y \rangle \geq \varepsilon$ and $\langle x, y \rangle \leq -\varepsilon$. There is a simple $O(1/\varepsilon^2)$ protocol for this problem: the players use shared randomness to choose $O(1/\varepsilon^2)$ random hyperplanes and then compare which side of each hyperplane their inputs lie in. Our main theorem implies that this is tight assuming $d \geq 1/\varepsilon^2$, as can be seen by embedding the hypercube in the set $\{-1/\sqrt{n}, 1/\sqrt{n}\}^n$.

Acknowledgments

We thank Bo'az Klartag for referring us to [Bar01] and Thomas Vidick for comments on an earlier draft. Oded Regev thanks Hartmut Klauck for introducing him to the smooth rectangle bound. Amit Chakrabarti thanks Oded Regev for agreeing to include a complete proof of the Gaussian noise correlation inequality in this paper, and T. S. Jayram for many enlightening discussions about GHD, over the years.

References

[ABC09] C. J. Arackaparambil, J. Brody, and A. Chakrabarti. Functional monitoring without monotonicity. In *Proc. 36th International Colloquium on Automata, Languages and Programming*, pages 95–106. 2009.

[Bar01] F. Barthe. An isoperimetric result for the Gaussian measure and unconditional sets. *Bull. London Math. Soc.*, 33(4):408–416, 2001.

[BC09] J. Brody and A. Chakrabarti. A multi-round communication lower bound for gap hamming and some consequences. In *Proc. 24th Annual IEEE Conference on Computational Complexity*, pages 358–368. 2009.

[BCR+10] J. Brody, A. Chakrabarti, O. Regev, T. Vidick, and R. de Wolf. Better gap-hamming lower bounds via better round elimination. In *Proc. 14th International Workshop on Randomization and Approximation Techniques in Computer Science*. 2010.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proc. 30th Annual ACM Symposium on the Theory of Computing*, pages 63–68. 1998.

[BJK+04] Z. Bar-Yossef, T. S. Jayram, R. Kumar, D. Sivakumar, and L. Trevisan. Counting distinct elements in a data stream. In *Proc. 6th International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 128–137. 2004.

[Bob03] S. G. Bobkov. On concentration of distributions of random weighted sums. *Ann. Probab.*, 31(1):195–215, 2003.

[Bor85] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985.

[BPSW06] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Comput. Complexity*, 15(4):391–432, 2006.

[CCM07] A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for computing the entropy of a stream. In *Proc. 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 328–335. 2007.

[DF84] P. Diaconis and D. Freedman. Asymptotics of graphical projection pursuit. *Ann. Statist.*, 12(3):793–815, 1984.

[Gra90] R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, New York, NY, USA, 1990. Available online at `http://ee.stanford.edu/~gray/it.html`.

[IW03] P. Indyk and D. Woodruff. Tight lower bounds for the distinct elements problem. In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 283–289. 2003.

[JK10] R. Jain and H. Klauck. The partition bound for classical communication complexity and query complexity. In *Proc. 25th Annual IEEE Conference on Computational Complexity*, pages 247–258. 2010.

[JKS07] T. S. Jayram, R. Kumar, and D. Sivakumar. The one-way communication complexity of gap hamming distance, 2007. Manuscript.

[Kla03] H. Klauck. Rectangle size bounds and threshold covers in communication complexity. In *Proc. 18th Annual IEEE Conference on Computational Complexity*, pages 118–134. 2003.

[Kla07] B. Klartag. A central limit theorem for convex sets. *Invent. Math.*, 168(1):91–131, 2007.

[Kla10] H. Klauck. A strong direct product theorem for disjointness. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 77–86. 2010.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, 1997.

[KW90] M. Karchmer and A. Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. Disc. Math.*, 3(2):255–265, 1990. Preliminary version in *Proc. 20th Annual ACM Symposium on the Theory of Computing*, pages 539–550, 1988.

[LS07] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proc. 39th Annual ACM Symposium on the Theory of Computing*, pages 699–708. 2007.

[MNSW98] P. B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998. Preliminary version in *Proc. 27th Annual ACM Symposium on the Theory of Computing*, pages 103–111, 1995.

[NW99] A. Nayak and F. Wu. The quantum query complexity of approximating the median and related statistics. In *Proc. 31st Annual ACM Symposium on the Theory of Computing*, pages 384–393. 1999.

[Păt08] M. Pătraşcu. (Data) STRUCTURES. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 434–443. 2008.

[Raz92] A. A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992. Preliminary version in *Proc. 17th International Colloquium on Automata, Languages and Programming*, pages 249–253, 1990.

[Raz99] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proc. 31st Annual ACM Symposium on the Theory of Computing*, pages 358–367. 1999.

[Raz02] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Science, Mathematics*, 67:0204025, 2002.

[She08] A. A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 85–94. 2008.

[Sud78] V. N. Sudakov. Typical distributions of linear functionals in finite-dimensional spaces of high dimension. *Dokl. Akad. Nauk SSSR*, 243(6):1402–1405, 1978.

[Vid10] T. Vidick. Personal communication, 2010.

[Woo04] D. P. Woodruff. Optimal space lower bounds for all frequency moments. In *Proc. 15th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 167–175. 2004.

[Woo07] D. P. Woodruff. *Efficient and Private Distance Approximation in the Communication and Streaming Models*. Ph.D. thesis, MIT, 2007.

[Woo09] D. Woodruff. The average case complexity of counting distinct elements. In *Proc. 12th International Conference on Database Theory*, pages 284–295. 2009.