

A Unified Lower Bound on the Noisy Query Complexity of Boolean Functions

Yuzhou Gu

New York University

SEVENKPLUS.G@GMAIL.COM

Xin Li

Johns Hopkins University

LIXINTS@CS.JHU.EDU

Yinzhan Xu

University of California, San Diego

XYZZHAN@UCSD.EDU

Editors: Steve Hanneke and Tor Lattimore

Abstract

We study the query complexity of Boolean functions $f : \{0, 1\}^n \rightarrow \{0, 1\}$ in the noisy query model introduced by Feige, Raghavan, Peleg and Upfal [SICOMP 1994]. In this model, an algorithm can adaptively query the bits of an input vector, but each query result is independently flipped with constant probability $p \in (0, 1/2)$; repeated queries are allowed. The noisy query complexity $N_p(f)$ of a function f is defined as the minimum expected number of queries needed to compute $f(x)$ with error probability at most $1/3$, for the worst case input x .

We prove a general lower bound on $N_p(f)$ based on degree statistics of certain subgraphs of the Boolean hypercube. This is the first general lower bound beyond those implied by the simple observation that $N_p(f)$ is lower bounded by the randomized query complexity. We show that this recovers (up to a constant factor) most previously known lower bounds on the noisy query complexity of Boolean functions, providing a unified framework for understanding these results and simplifying the proofs in several cases. Furthermore, this resolves in the affirmative an open problem of Gu, Li and Xu [COLT 2025] that $N_p(f) = \Omega(I(f) \log I(f))$, where $I(f)$ denotes the total influence of f . We also apply our general lower bound to obtain tight bounds on the noisy query complexity for several new functions.

Keywords: noisy computation, total influence, boolean functions

1. Introduction

In modern computation architectures, especially in large systems, errors occur inevitably due to various reasons such as hardware failures, communication errors and environmental noise. It is thus crucial to design algorithms that are robust in the presence of errors. Feige et al. (1994) formally defined the following extremely natural and basic noisy query model that incorporates errors: Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a Boolean function. Given an unknown fixed input $x \in \{0, 1\}^n$, a (potentially randomized) algorithm can query each bit x_i , but the returned bit could be flipped with some fixed constant probability p , and repeated queries are allowed. The goal is to compute $f(x)$ while minimizing the number of queries. There has been a lot of work studying this model (Reischuk and Schmeltz, 1991; Feige, 1992; Kenyon and King, 1994; Evans and Pippenger, 1998; Zhu et al., 2023, 2024; Wang et al., 2025; Gu et al., 2025), as well as the related noisy comparison model (Horstein, 1963; Berlekamp, 1964; Burnashev and Zigangirov, 1974; Feige, 1992; Ben-Or and Hasidim, 2008; Emamjomeh-Zadeh et al., 2016; Dereniowski et al., 2019; Wang et al., 2022; Gu and Xu, 2023; Wang et al., 2023; Zhu et al., 2023; Dereniowski et al., 2025).

However, despite much attention, even such a basic model has not been fully understood. Let $N_p(f)$ denote the minimum expected number of noisy queries needed to compute f with error probability $\leq 1/3$ by an adaptive randomized algorithm. Clearly, $N_p(f) \geq \Omega(R(f))$, the randomized query complexity for computing f within error $1/3$, as computation with exact queries can only be easier than computation with noisy queries. Also, $N_p(f) = O(R(f) \log R(f))$ ¹ for the following reason: Given a randomized algorithm for computing f with exact queries, one can repeat each query $O(\log R(f))$ times in the noisy query model and take the majority results, so that each query will be correct with probability $1 - O(1/R(f))$. Hence, by union bound, all $R(f)$ queries can be simulated correctly with constant probability. However, this $[\Omega(R(f)), O(R(f) \log R(f))]$ range is not strong enough to fully determine $N_p(f)$, and there exist f where $N_p(f) = \Theta(R(f))$ and others where $N_p(f) = \Theta(R(f) \log R(f))$. For instance, it is known that PARITY (Feige et al., 1994) and Graph-Connectivity (Gu et al., 2025) have $N_p(f) = \Theta(R(f) \log R(f))$, and AND, OR have $N_p(f) = \Theta(R(f))$ (Feige et al., 1994). Hence, it is not possible to fully characterize $N_p(f)$ using only $R(f)$, and it is also important to study $N_p(f)$ with respect to other complexity measures.

Two important measures of Boolean functions are their sensitivity and influence. For any $x \in \{0, 1\}^n$, the sensitivity of f at x , $s(f, x)$, is defined as the number of y such that $f(x) \neq f(y)$ and x, y only differ at one bit; and the maximum sensitivity of f at any input x is denoted as $s(f)$. On the other hand, the total influence $I(f)$ is the same as the average sensitivity over all inputs, i.e., $I(f) = \mathbb{E}_{x \sim \{0, 1\}^n} [s(f, x)]$. Both sensitivity and influence are central notions in Boolean function analysis and have been used in various contexts such as hardness of approximation (Dinur and Safra, 2005), voting theory (Mossel, 2012) and random graph theory (Friedgut and Kalai, 1996). Sensitivity and its variants have also been shown to be closely related to many query complexity measures (e.g., (Nisan and Szegedy, 1994)).

It is well-known that $R(f) = \Omega(s(f))$ (Nisan, 1989) and hence we also have $N_p(f) = \Omega(s(f))$. In (Reischuk and Schmeltz, 1991), the authors proved that the *non-adaptive* noisy query complexity of any Boolean function f is at least $\Omega(s(f) \log s(f))$, so a natural question is whether this can be extended to adaptive queries. This is not possible. For instance, the OR function has sensitivity n and adaptive noisy query complexity $O(n)$ (Feige et al., 1994).

This motivates Gu et al. (2025) to study the relationship between $N_p(f)$ and $I(f)$, as $I(f) = \mathbb{E}_{x \sim \{0, 1\}^n} [s(f, x)]$ is the average sensitivity of f which can be smaller than $s(f)$. As before, the simple fact that $N_p(f) \geq \Omega(R(f))$ also implies $N_p(f) = \Omega(I(f))$. Can there be better bounds on $N_p(f)$ based on $I(f)$? Gu et al. (2025) conjectured that $N_p(f) = \Omega(I(f) \log I(f))$, and proved it for the special case where function f has near-maximum $\Theta(n)$ total influence $I(f)$. In this work, we resolve their conjecture in the affirmative.

Theorem 1 *For every constant $p \in (0, 1/2)$, there exists a constant $c = c(p) > 0$ such that $N_p(f) \geq cI(f) \log I(f)$ for every Boolean function f .*

This is the first lower bound on the noisy query complexity that is general enough to apply to an arbitrary function (besides the trivial $\Omega(R(f))$ lower bound). Previously, lower bounds are often studied for specific functions (Feige et al., 1994), restricted classes of functions such as random functions (Reischuk and Schmeltz, 1991; Feige, 1992; Evans and Pippenger, 1998) and functions with high total influence (Gu et al., 2025), or restricted algorithms such as non-adaptive algorithms (Reischuk and Schmeltz, 1991). This is also the best lower bound of $N_p(f)$ based purely on $I(f)$

1. Throughout the paper, we use the convention that $x \log x = 0$ for $x = 0$.

one could hope for, because for any value of $v \in [n]$, there are functions where $l(f) = v$ and $N_p(f) = \Theta(l(f) \log l(f))$ (e.g., PARITY on a subset of v indices).

1.1. Main Technical Theorem and Other Applications

Theorem 1 is a consequence of an even stronger result which we will explain next. Consider the Boolean hypercube H . For every edge $(x, y) \in E(H)$, we keep it if $f(x) \neq f(y)$ and remove it otherwise. Let us call the remaining graph H_f . Our main technical theorem that leads to **Theorem 1** is the following:

Theorem 2 *For every constant $p \in (0, 1/2)$ and $\gamma \in (0, 1]$, there exists $c = c(p, \gamma) > 0$ such that the following holds. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a Boolean function and let G be an induced subgraph of H_f with a nonempty edge set. Let $d_1 = \min_{x \in V(G)} \deg_G(x)$. Let μ be the distribution on $V(G)$ defined as $\mu(x) = \frac{\deg_G(x)}{2|E(G)|}$ and let $d_2 \geq d_1$ be any number satisfying $\mathbb{P}_{x \sim \mu}(\deg_G(x) \geq d_2) \geq \gamma$. Then $N_p(f) \geq c \cdot d_2 \log(1 + d_1)$.*

Note that the distribution μ is equivalent to the following distribution: We first sample a random edge from $E(G)$, and then sample a random endpoint of the edge. Using this interpretation, our theorem works for the case where G has a constant fraction of edges that have one endpoint of degree at least d_2 .

To see how **Theorem 2** implies **Theorem 1**, we consider the degeneracy of a Boolean function f , which may be a parameter of independent interest (even though degeneracy is a well-studied concept in graph theory, the degeneracy of H_f has not been studied to the best of our knowledge). The degeneracy of a graph is the smallest integer k such that every subgraph of the graph has at least one vertex with degree at most k . We use $\text{degen}(f)$ to denote the degeneracy of H_f . By definition of $\text{degen}(f)$, there exists an induced subgraph G of H_f where the minimum degree of G is $\text{degen}(f)$. Hence, we can apply **Theorem 2** with $d_1 = d_2 = \text{degen}(f)$ to obtain the following result:

Theorem 3 *For every constant $p \in (0, 1/2)$, there exists a constant $c = c(p) > 0$ such that $N_p(f) \geq c \text{degen}(f) \log \text{degen}(f)$ for every Boolean function f .*

Note that the average degree of H_f is exactly $l(f)$. A basic fact in graph theory is that the average degree of a graph is at most twice its degeneracy (see e.g., [Lick and White \(1970\)](#)). Applying this basic fact to H_f , we get that $l(f) \leq 2 \text{degen}(f)$. Therefore, **Theorem 3** implies **Theorem 1**.

We also remark that $\text{degen}(f)$ is related to sensitivity. More precisely, it can be seen that $\text{degen}(f) = \max_{\emptyset \subsetneq V \subseteq \{0, 1\}^n} \min_{x \in V} s(f_V, x)$, where f_V is the partial function which agrees with f on V but undefined everywhere else.

We also apply **Theorem 2** to obtain several new lower bounds for more specific functions. The first example is biased random function, which extends the previous lower bound on unbiased random functions ([Reischuk and Schmeltz, 1991](#); [Feige, 1992](#); [Evans and Pippenger, 1998](#)) to a more general setting (in the following, a biased random function means that for each input $x \in \{0, 1\}^n$, the value $f(x)$ is sampled independently, with $\Pr[f(x) = 1] = q$, where q need not equal $1/2$):

Theorem 4 *Let $\lambda \in (0, 1), p \in (0, 1/2), \gamma > 0$ be fixed constants. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a biased random function where for each $x \in \{0, 1\}^n$, $f(x) = 1$ independently with probability $q = \gamma n^{-\lambda}$. Then $N_p(f) \geq \Omega(n \log n)$ with probability $1 - \exp(-cn)$ for some constant $c > 0$ over the randomness of f .*

Our second example considers the complexity of counting the number of any fixed pattern $t = t_1 t_2 \cdots t_k$ in the bit sequence $x_1 x_2 \cdots x_n$ as a subsequence. This is a vast generalization of the problem on counting the number of 1's in the sequence $x_1 x_2 \cdots x_n$, which was known to require $\Omega(n \log n)$ noisy queries (Feige et al., 1994). Note that this subsequence counting problem does not have Boolean output, but our techniques can still apply to such functions, via reduction from a threshold version of the function which has Boolean output. Also, $N_p(f)$ for non-Boolean functions f can be defined analogously.

Theorem 5 *Let $t \in \{0, 1\}^k$ be a fixed pattern for some $k = O(1)$. Let $f : \{0, 1\}^n \rightarrow \mathbb{Z}$ be a function where $f(x)$ denotes the number of times t appears in x_1, \dots, x_n as a (not necessarily contiguous) subsequence. Let $p \in (0, 1/2)$ be a constant. Then $N_p(f) = \Omega(n \log n)$.*

Besides showing new lower bounds on $N_p(f)$, Theorem 2 is strong enough to unify *all* previous nontrivial (i.e. $N_p(f) \gg R(f)$) asymptotic noisy query lower bounds to the best of our knowledge. Previously, nontrivial lower bounds for $N_p(f)$ were known for PARITY, k -THRESHOLD, symmetric functions (Feige et al., 1994), random functions (Reischuk and Schmeltz, 1991; Feige, 1992; Evans and Pippenger, 1998), Graph-Connectivity and s - t -Connectivity (Gu et al., 2025), and functions with $\Omega(n)$ total influence (Gu et al., 2025). There are also examples whose lower bounds follow by directly reducing from these problems (e.g., s - t Shortest Path (Gu et al., 2025)). Besides such examples, this is a comprehensive list to the best of our knowledge. Among these problems, functions with $\Omega(n)$ total influence are a special case of Theorem 1. PARITY and random functions have $\Theta(n)$ total influence, so they are also special cases of Theorem 1. In Section 4, we will explain how Theorem 2 implies the lower bounds of noisy query complexity for k -THRESHOLD, symmetric functions, Graph-Connectivity and s - t -Connectivity. Because Theorem 2 is stronger than previous approaches, the new lower bound proofs based on it are also much simpler than previous proofs in some cases. For instance, to show the lower bound for Graph-Connectivity, Gu et al. (2025) had to analyze the structure of a uniform spanning tree of a complete graph, which is quite complicated and takes 5 pages of their proofs. Our proof based on Theorem 2 is quite simple and does not rely on Gu et al. (2025)'s structural results.

We remark that the lower bound obtained from Theorem 2 is not tight in general. The best lower bound one can hope to obtain using Theorem 2 is $\Omega(\Delta(H_f) \log \Delta(H_f))$, where $\Delta(H_f)$ denotes the maximum degree of H_f . As $\Delta(H_f)$ is exactly the sensitivity of f , $s(f)$, the best lower bound we can hope to obtain is $\Omega(s(f) \log s(f))$. Recall $N_p(f) = \Omega(R(f))$, so as long as $R(f) = \omega(s(f) \log s(f))$ (see e.g., (Rubinfeld, 1995) for examples), our lower bound is not tight. There are many known examples with $R(f) = \omega(s(f) \log s(f))$.

2. Preliminaries

For a simple undirected graph G , $V(G)$ denotes the set of vertices and $E(G)$ denotes the set of edges. For $x \in V(G)$, $\deg_G(x)$ denotes the degree of x in G .

For $x \in \{0, 1\}^n$, $|x|$ denotes the Hamming weight of x . For a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, H_f denotes the graph whose vertices are $\{0, 1\}^n$ and there is an edge between two vertices x and y if and only if x and y differ at exactly one bit and $f(x) \neq f(y)$.

$\text{Bin}(n, p)$ denotes the binomial distribution. The Chernoff bound states that for $X \sim \text{Bin}(n, p)$, $\mu = np$, and $\epsilon \geq 0$, we have

$$\begin{aligned}\mathbb{P}(X \geq (1 + \epsilon)\mu) &\leq \exp\left(-\frac{\epsilon^2}{2 + \epsilon}\mu\right), \\ \mathbb{P}(X \leq (1 - \epsilon)\mu) &\leq \exp\left(-\frac{\epsilon^2}{2}\mu\right).\end{aligned}$$

We also use the Azuma-Hoeffding inequality. Let $g : \{0, 1\}^n \rightarrow \mathbb{R}$ where

$$|g(x_1, \dots, x_n) - g(y_1, \dots, y_n)| \leq B$$

for some bound B , for every x and y that differ at exactly one bit. Then for every $\lambda > 0$,

$$\mathbb{P}_{x \sim \{0,1\}^n} (|g(x) - \mathbb{E}[g(x)]| \geq \lambda B) \leq 2 \exp(-2\lambda^2/n).$$

Markov's inequality states that for a non-negative random variable X , $\mathbb{P}(X \geq t\mathbb{E}[X]) \leq 1/t$ for all $t > 0$. We will also use the following corollary:

Corollary 6 *Let X be a random variable taking values in $[0, 1]$, then $\mathbb{P}(X \geq t) \geq \mathbb{E}[X] - t$ for $0 \leq t \leq \mathbb{E}[X]$.*

Proof We have

$$\mathbb{P}(X \geq t) = \mathbb{P}(1 - X \leq 1 - t) \geq 1 - \mathbb{P}(1 - X \geq 1 - t) \geq 1 - \frac{1 - \mathbb{E}[X]}{1 - t} \geq \mathbb{E}[X] - t,$$

where the third step is by Markov's inequality. ■

3. Technical Overview

In this section, we give a high-level overview of our proof for the main technical theorem [Theorem 2](#). The full proof is deferred to [Appendix A](#). Our proof uses the three-phase framework developed in [Feige \(1992\)](#); [Gu et al. \(2025\)](#). Given an input $x \in \{0, 1\}^n$, the goal of the three-phase problem is also to compute $f(x)$ by querying the input bits x_i . In the first phase, the algorithm queries every bit $c_1 \log d_1$ times in a nonadaptive way with noisy queries. In the second phase, some information is revealed to the algorithm for free. In the third phase, the algorithm can adaptively make $c_2 d_2$ exact queries to compute $f(x)$ (for simplicity, in this overview we assume a worst case $c_2 d_2$ query bound; in the actual proof, we relax it to an expected bound). As shown in [Gu et al. \(2025\)](#), if an algorithm cannot solve the three-phase problem with constant error probability, then any algorithm \mathcal{A} making $c_1 c_2 d_2 \log(d_1)$ noisy queries cannot compute $f(x)$ with constant error probability. Essentially, if there is such an algorithm \mathcal{A} for computing $f(x)$, then we can simulate it for the three-phase problem. As shown in [Gu et al. \(2025\)](#), there is some way to design the second phase, so that the observations made in the first two phases are equivalent to the following:

- For every i where $x_i = 1$, x_i is revealed independently with probability p_+ , where $p_+ = 1 - d_1^{-c_3 \pm o(1)}$ for some constant c_3 ;

- For every i where $x_i = 0$, x_i is revealed independently with probability p_- , where $p_- = o(1)$.

Let $z \in \{0, 1, \star\}^n$ denote the revealed information up to the second phase, i.e., $z_i = x_i$ if x_i is revealed, and $z_i = \star$ otherwise. Let ν_x be the distribution of z under x .

Previous methods using the three-phase framework (Gu et al., 2025) (or the predecessor two-phase framework in the case of Feige et al. (1994)) use ad-hoc analyses tailored to the specific functions. We instead provide a unified approach, based on degree statistics of a certain subgraph of the Boolean hypercube. Our approach first differs from Gu et al. (2025) by defining a distribution μ that applies to any Boolean function. Recall the distribution μ and a subgraph G of H_f from the statement of Theorem 2, where $\mu(x) = \frac{\deg_G(x)}{2|E(G)|}$ for $x \in V(G)$. We can extend μ to $\{0, 1\}^n$, simply by setting $\mu(x) = 0$ for $x \notin V(G)$. This will be the hard distribution for the three-phase problem.

Our goal is to show that, after the three phases, the probability of $f(x) = 0$ and the probability of $f(x) = 1$ are both likely lower bounded by a constant, where x is distributed with respect to the posterior distribution under all the observations. If this holds, then no matter what value the algorithm outputs for $f(x)$, there is always a constant error probability. Our approach for showing such balance conditions is different from previous approaches (Feige et al., 1994; Gu et al., 2025). Previously, they were shown either by ad-hoc analyses or, for the case of high-influence functions (Gu et al., 2025), by the fact that $I(f) = \Theta(n)$ implies this balance condition. Our approach is more refined and, as we will see shortly, uses a graph-theoretical view of this condition.

As a warm-up, we will first describe why the probability of $f(x) = 0$ and the probability of $f(x) = 1$ are both likely lower bounded by a constant after the first two phases. This is necessary for the lower bound, because otherwise, the algorithm can simply return the value that is more likely, and achieve sub-constant error probability without any queries in the third phase.

Let μ_z denote the posterior probability of x under the observation z . We need to show that $\min\{\mu_z(f^{-1}(0)), \mu_z(f^{-1}(1))\} \geq \Omega(1)$, with good probability. Let \mathcal{X}_z denote the support of μ_z . We can imagine a graph, whose vertices are \mathcal{X}_z , and there is an edge between two vertices x and y if $(x, y) \in E(G)$. Each vertex x has a weight, which equals $\mu_z(x)$, and we will also assign each edge (x, y) a weight $w_z(x, y)$. The degree of a vertex x is defined as

$$\text{wdeg}_z(x) = \mu_z(x)^{-1} \sum_{y:(x,y) \in E(G)} w_z(x, y).$$

This is a bipartite graph, where one side is $\mathcal{X}_z \cap f^{-1}(0)$, and the other side is $\mathcal{X}_z \cap f^{-1}(1)$. If we can show an upper bound D on the maximum degree, and show a lower bound M on the total edge weight, then the total vertex weights of either side of the bipartite graph can be lower bounded by M/D , i.e., $\min\{\mu_z(f^{-1}(0)), \mu_z(f^{-1}(1))\} \geq M/D$. The values for M and D we aim for are $M = \Omega(\sqrt{1-p_+})$ and $D = O(\sqrt{1-p_+})$, which can show $\min\{\mu_z(f^{-1}(0)), \mu_z(f^{-1}(1))\} \geq \Omega(1)$ via the previous inequality. More precisely, we will show an upper bound D on the degree of most probability mass of x under μ_z , which is still able to give us a lower bound.

Towards this goal, we define $w_z(x, y) = \sqrt{\frac{\mu_z(x)\mu_z(y)}{\deg_G(x)\deg_G(y)}}$. Let $N_+(x)$ denote the set of y where $(x, y) \in E(G)$ and $|y| = |x| + 1$, and let $N_-(x)$ denote the set of y where $(x, y) \in E(G)$ and $|y| = |x| - 1$. By some calculation, one can show the following equality (this is Equation (5) in Appendix A) that relates the degree of x with the number of vertices in $N_+(x)$ and in $N_-(x)$ that

also appears in the support of μ_z .

$$\text{wdeg}_z(x) = \deg_G(x)^{-1} \left(|N_+(x) \cap \mathcal{X}_z| \sqrt{\frac{1-p_+}{1-p_-}} + |N_-(x) \cap \mathcal{X}_z| \sqrt{\frac{1-p_-}{1-p_+}} \right). \quad (1)$$

Upper bound on degree of “most” vertices. Because $|N_+(x) \cap \mathcal{X}_z| \leq \deg_G(x)$, the contribution of the first term in Equation (1) is only $\sqrt{\frac{1-p_+}{1-p_-}} = O(\sqrt{1-p_+})$ (recall $p_- = o(1)$). Next, we consider the second term. For $z \sim \nu_x$, for any $y \in N_-(x)$, $y \in \mathcal{X}_z$ if $z_i = \star$, where i is the index that x and y differ. This happens with probability $1-p_+$, and is independent for different $y \in N_-(x)$. Thus, we can apply concentration bounds to show that it is likely that

$$|N_-(x) \cap \mathcal{X}_z| \leq O((1-p_+) \cdot |N_-(x)|) = O((1-p_+) \cdot \deg_G(x)).$$

If this holds, then the contribution of the second term in Equation (1) can also be bounded by $O(\sqrt{1-p_+})$. This is where we need the condition that the minimum degree of G is d_1 ; otherwise, the deviation probability of the above concentration becomes too large.

Lower bound on total edge weights. We will show that with constant probability (over the randomness of z), $\text{wdeg}_z(x) = \Omega(\sqrt{1-p_+})$ for a constant density of x under μ_z . If this holds, then the total edge weights can be lower bounded by

$$\frac{1}{2} \mathbb{E}_{x \sim \mu_z} [\text{wdeg}_z(x)] \geq \Omega(\sqrt{1-p_+})$$

as desired.

By the assumption in the statement of Theorem 2, for $x \sim \mu$, with constant probability, we have $\deg_G(x) \geq d_2$. Hence, either $|N_+(x)| \geq \deg_G(x)/2 \geq d_2/2$, or $|N_-(x)| \geq \deg_G(x)/2 \geq d_2/2$, with constant probability. We will assume without loss of generality that the former case holds. Fix any x where $|N_+(x)| \geq \deg_G(x)/2 \geq d_2/2$. As before, when $z \sim \nu_x$, some $y \in N_+(x)$ remains in \mathcal{X}_z when $z_i = \star$, where i is the index of the bit where x and y differ. Again, we can apply concentration bounds to show that it is likely

$$|N_+(x) \cap \mathcal{X}_z| \geq \Omega((1-p_-) \cdot |N_+(x)|) = \Omega((1-p_-) \cdot \deg_G(x)).$$

Hence, by Equation (1), $\text{wdeg}_z(x) \geq \Omega(\sqrt{1-p_+})$ as desired.

Handling the third phase. In the third phase, the algorithm can make $c_2 d_2$ adaptive exact queries. Each query reveals one bit of the input vector x , and the posterior distribution of x will change. In particular, the posterior probability of those x that are inconsistent with the revealed bit will become 0, and the posterior distribution of other x 's will be scaled properly. Let $\mu_{z,t}$ denote the posterior distribution of x after the first t queries and let $\mathcal{X}_{z,t}$ denote the support of $\mu_{z,t}$. We can consider the subgraph of the graph on \mathcal{X}_z considered above, on the subset of vertices $\mathcal{X}_{z,t}$. All the vertex weights should be scaled to $\mu_{z,t}(x)$, and all edge weights should also be scaled properly. Intuitively, because now we are considering a subgraph, the upper bound on the degrees of “most” vertices should still hold.

To show a lower bound for the total edge weights, let X denote the set of x where $|N_+(x)| \geq \deg_G(x)/2 \geq d_2/2$ and $|N_+(x) \cap \mathcal{X}_z| = \Omega(\deg_G(x))$. Previous steps can establish that $\mu_z(X) \geq \Omega(1)$ with constant probability. It is not difficult to check that $\mathbb{E}[\mu_{z,t}(X) \mid z] = \mu_z(X)$, so

$\mathbb{E}[\mu_{z,t}(X) \mid z] \geq \Omega(1)$ with constant probability (over z) as well. Hence, with constant probability, vertices in X still carry a constant fraction of the weights in the subgraph on $\mathcal{X}_{z,t}$. Because each query in the third phase reveals one bit, $|N_+(x) \cap \mathcal{X}_{z,t}| \geq \Omega(\deg_G(x)) - t$ for any $x \in X$. As $t \leq c_2 d_2$, for sufficiently small c_2 , we still have $|N_+(x) \cap \mathcal{X}_{z,t}| \geq \Omega(\deg_G(x)) - c_2 d_2 = \Omega(\deg_G(x))$. Hence, we can still use a formula similar to Equation (1) to obtain a lower bound of total edge weights.

4. Applications

In this section, we show various applications of Theorem 2. We start with the following lemma which will be used in several of the proofs.

Lemma 7 *Let $\lambda \in [0, 1)$, $\gamma_1, \gamma_2, \gamma_3 > 0$, $p \in (0, 1/2)$ be constants. Then there exists $c = c(\lambda, \gamma_1, \gamma_2, \gamma_3, p)$ such that the following holds. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a Boolean function and let G be an induced subgraph of H_f . Suppose there exists a set S of $\gamma_1 n^{-\lambda} \cdot |V(G)|$ vertices with degree at least $\gamma_2 n$ in G , and the total number of edges in G is at most $\gamma_3 n |S|$. Then $N_p(f) \geq cn \log n$.*

Proof We iteratively remove vertices whose degrees are smaller than $\frac{1}{4}\gamma_1\gamma_2 n^{1-\lambda}$ from G until there are no such vertices left. Let G' denote the remaining graph and let S' denote $V(G') \cap S$.

The total number of edges we delete from G is at most $\frac{1}{4}\gamma_1\gamma_2 n^{1-\lambda} \cdot |V(G)|$. Because initially, the total degree of all vertices from S is at least $|S| \cdot \gamma_2 n$, the total degree of all vertices from S' in G' is at least (removing one edge can decrease the degrees of two vertices)

$$|S| \cdot \gamma_2 n - 2 \cdot \frac{1}{4}\gamma_1\gamma_2 n^{1-\lambda} \cdot |V(G)| \geq \frac{1}{2}|S| \cdot \gamma_2 n.$$

Hence, if we sample a random $s \in S$, $\mathbb{E}[\deg_{G'}(s)/n] \geq \frac{1}{2}|S| \cdot \gamma_2 n / |S|/n \geq \gamma_2/2$ (here, if $s \notin S'$, we regard $\deg_{G'}(s)$ as 0). By Corollary 6, $\Pr(\deg_{G'}(s)/n \geq \gamma_2/4) \geq \gamma_2/4$, i.e., there are $\frac{1}{4}\gamma_2|S|$ vertices whose degrees are at least $\frac{1}{4}\gamma_2 n$ in G' . We call these vertices S_{good} .

Let μ be the distribution from Theorem 2's statement applied to G' , i.e., $\mu(x) = \frac{\deg_{G'}(x)}{2|E(G')|}$ for $x \in V(G')$. We have

$$\begin{aligned} \mu(S_{\text{good}}) &= \frac{\sum_{s \in S_{\text{good}}} \deg_{G'}(s)}{2|E(G')|} \\ &\geq \frac{(\frac{1}{4}\gamma_2|S|) \cdot (\frac{1}{4}\gamma_2 n)}{2|E(G)|} \\ &\geq \frac{(\frac{1}{4}\gamma_2|S|) \cdot (\frac{1}{4}\gamma_2 n)}{2 \cdot \gamma_3 n |S|} \\ &\geq \frac{\gamma_2^2}{32\gamma_3}. \end{aligned}$$

Therefore, we can apply Theorem 2 with $d_1 \geq \frac{1}{4}\gamma_1\gamma_2 n^{1-\lambda}$ and $d_2 \geq \frac{1}{4}\gamma_2 n$ to obtain

$$N_p(f) \geq c \frac{1}{4}\gamma_2 n \cdot \log \left(1 + \frac{1}{4}\gamma_1\gamma_2 n^{1-\lambda} \right) \geq c' n \log n$$

for some constant c' . ■

Next, we prove the lower bound for biased random functions.

Theorem 4 *Let $\lambda \in (0, 1), p \in (0, 1/2), \gamma > 0$ be fixed constants. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a biased random function where for each $x \in \{0, 1\}^n$, $f(x) = 1$ independently with probability $q = \gamma n^{-\lambda}$. Then $\mathbb{N}_p(f) \geq \Omega(n \log n)$ with probability $1 - \exp(-cn)$ for some constant $c > 0$ over the randomness of f .*

Proof First, by Chernoff bound, the number of x where $f(x) = 1$ is between $\frac{2}{3}q \cdot 2^n$ and $\frac{4}{3}q \cdot 2^n$ with probability $\geq 1 - \exp(-c_1 q 2^n)$ for some constant $c_1 > 0$. Let this event be \mathcal{E}_1 .

Next, for any fixed x , consider its neighbors in the Boolean hypercube. The number of them that evaluate to 1 is distributed as $\text{Bin}(n, q)$. Let us denote the random variable by Y . We assume n is sufficiently large so that $q \leq 1/2$. Then by Chernoff bound,

$$\Pr\left(Y \geq \frac{2}{3}n\right) \leq \exp(-c_2 n)$$

for some constant $c_2 > 0$. Let $X_{\text{bad}} \subseteq \{0, 1\}^n$ be the set of x whose neighborhood contains more than $\frac{2}{3}n$ elements that evaluate to 1. Hence, $\mathbb{E}[|X_{\text{bad}}|] \leq 2^n \cdot \exp(-c_2 n)$. By Markov's inequality, $\Pr(|X_{\text{bad}}| > 2^n \cdot \exp(-c_2 n/2)) \leq \exp(-c_2 n/2)$. Let \mathcal{E}_2 denote the event where $|X_{\text{bad}}| \leq 2^n \cdot \exp(-c_2 n/2)$.

We will show that under $\mathcal{E}_1 \cap \mathcal{E}_2$ (which happens with probability $1 - \exp(-cn)$ for some constant c by union bound), $\mathbb{N}_p(f) \geq \Omega(n \log n)$.

For any $x \in f^{-1}(1) \setminus X_{\text{bad}}$, it has at least $n/3$ neighbors in H_f . Furthermore, under $\mathcal{E}_1 \cap \mathcal{E}_2$, $|f^{-1}(1) \setminus X_{\text{bad}}| \geq \frac{2}{3}q \cdot 2^n - |X_{\text{bad}}| \geq (\frac{2}{3} - o(1))q \cdot 2^n$. Finally, under $\mathcal{E}_1 \cap \mathcal{E}_2$, the total number of edges in H_f is at most $\frac{4}{3}q \cdot n \cdot 2^n \leq (2 + o(1))n|f^{-1}(1) \setminus X_{\text{bad}}|$.

Therefore, we can apply [Lemma 7](#) where $G := H_f$ and $S := f^{-1}(1) \setminus X_{\text{bad}}$. This finishes the proof. ■

Next, we show our lower bound for the subsequence counting problem:

Theorem 5 *Let $t \in \{0, 1\}^k$ be a fixed pattern for some $k = O(1)$. Let $f : \{0, 1\}^n \rightarrow \mathbb{Z}$ be a function where $f(x)$ denotes the number of times t appears in x_1, \dots, x_n as a (not necessarily contiguous) subsequence. Let $p \in (0, 1/2)$ be a constant. Then $\mathbb{N}_p(f) = \Omega(n \log n)$.*

Proof Assume without loss of generality that $t_k = 1$, by complementing all bits if $t_k = 0$. Also, assume without loss of generality $t_j = 0$ for some $j \in [k]$, as otherwise the lower bound is implied by the known lower bound of computing $|x|$ ([Feige et al., 1994](#)).

Let C be a sufficiently large constant. Let $\ell = \lfloor n/(k + (C - 1)(k - |t|)) \rfloor$ and $r = n - \ell((k + (C - 1)(k - |t|)))$. We will consider a subset of all possible inputs, where all but the last ℓ bits are fixed. More precisely, let L_1 be the concatenation of ℓ bits of 1's, and let L_0 be the concatenation of $C\ell$ bits of 0's. Let P be r copies of the bit $1 - t_1$. For $x \in \{0, 1\}^\ell$, define

$$g(x) = f(P, L_{t_1}, L_{t_2}, \dots, L_{t_{k-1}}, x).$$

Note that the first r bits cannot be involved in any subsequence t . Clearly, an $\Omega(\ell \log \ell)$ lower bound for computing g will imply an $\Omega(\ell \log \ell) = \Omega(n \log n)$ lower bound for computing f .

For any $x_i = 1$, the number of subsequences t using x_i in $L_{t_1}, L_{t_2}, \dots, L_{t_{k-1}}, x$ is at least $\prod_{j=1}^{k-1} |L_{t_j}| = C^{k-|t|} \cdot \ell^{k-1}$. On the other hand, for any $x_i = 0$, the number of subsequences involving x_i is at most $k^{k-1} C^{k-|t|-1} \ell^{k-1}$ (by using the fact that the number of 1's in the whole sequence is at most $k\ell$ and the number of 0's in the whole sequence is at most $Ck\ell$). Hence, by setting $C > 2k^{k-1}$, we can guarantee that $g(y) - g(x) \geq c_1 \cdot \ell^{k-1}$ for some constant $c_1 > 0$, where y and x differ at exactly one bit and $y_i = 1, x_i = 0$ on that bit.

On the other hand, if y and x differ at one bit, we have $|g(y) - g(x)| = O(\ell^{k-1})$, because each bit can only be involved in $O(\ell^{k-1})$ subsequences. Hence, we can apply the Azuma-Hoeffding inequality to obtain that,

$$\mathbb{P}_{x \sim \{0,1\}^\ell} \left(|g(x) - \mathbb{E}[g(x)]| \leq c_2 \ell^{k-0.5} \right) \geq 0.9 \quad (2)$$

for some constant $c_2 > 0$.

Furthermore, let B_{bad} denote the set of $x \in \{0,1\}^\ell$ where the number of 0's is less than $\ell/3$. By Chernoff bound, $|B_{\text{bad}}| \leq \exp(-c_3\ell) \cdot 2^\ell$ for some constant $c_3 > 0$.

Next, we split $\mathbb{Z}_{\geq 0}$ into intervals of length $c_1 \cdot \ell^{k-1}$, and we denote the intervals by I_1, I_2, \dots . Let I_m be the interval maximizing $|g^{-1}(I_m)|$. Let S denote $g^{-1}(I_m) \setminus B_{\text{bad}}$. Let T be the largest element in I_m , and consider the threshold version g_T of g defined as $g_T(x) = [g(x) \leq T]$.

- By Equation (2), there exist $O(\sqrt{\ell})$ intervals I_a, I_{a+1}, \dots, I_b such that $\sum_{i=a}^b |g^{-1}(I_i)| \geq 0.9 \cdot 2^\ell$. Hence, $|g^{-1}(I_m)| \geq 0.9 \cdot 2^\ell / O(\sqrt{\ell}) = \Omega(2^\ell / \sqrt{\ell})$ as m maximizes $|g^{-1}(I_m)|$. By the upper bound on B_{bad} , this also implies $|S| \geq |g^{-1}(I_m)| - \exp(-c_3\ell) \cdot 2^\ell = \Omega(|g^{-1}(I_m)|) = \Omega(2^\ell / \sqrt{\ell})$.
- For every $x \in S$, if we flip any $x_i = 0$ to 1 and obtain y , $g(y) - g(x) \geq c_1 \cdot \ell^{k-1}$ by previous discussion. By definition of intervals and T , $T - c_1 \ell^{k-1} < g(x) \leq T$. Hence, $g(y) > T$. This means that $g_T(x) \neq g_T(y)$, and y is a neighbor of x in H_{g_T} . As $x \notin B_{\text{bad}}$, the number of such neighbors is $\Omega(\ell)$.
- Finally, let (x, y) be any edge in H_{g_T} , with $g_T(x) = 1$ and $g_T(y) = 0$, i.e., $g(x) \leq T$ and $g(y) > T$. As discussed earlier, $g(y) - g(x) = O(\ell^{k-1})$, so we must have $g(x) > T - O(\ell^{k-1})$. This means that x is from $g^{-1}(I_i)$ for some $i = m - O(1)$. Hence, the total number of edges of H_{g_T} is upper bounded by

$$\sum_{i=m-O(1)}^m \ell \cdot |g^{-1}(I_i)| = O(\ell |g^{-1}(I_m)|) = O(\ell |S|).$$

Because of the three bullet points, we can apply Lemma 7 with input dimension ℓ , $\lambda = 1/2$, $G = H_{g_T}$ and the above set S to obtain $\mathbb{N}_p(g_T) = \Omega(\ell \log \ell) = \Omega(n \log n)$, which implies $\mathbb{N}_p(g) = \Omega(n \log n)$ and $\mathbb{N}_p(f) = \Omega(n \log n)$. \blacksquare

4.1. New Proofs of Previous Results

In this section, we show that our main technical theorem is general enough to imply all previous known nontrivial lower bounds on the noisy query complexity. As discussed in Section 1, it suffices to consider k -THRESHOLD, symmetric functions, Graph-Connectivity and s - t -Connectivity.

We first consider symmetric functions. A Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is symmetric if for any permutation π on $[n]$ and any $x \in \{0, 1\}^n$, $f(x) = f(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)})$. In other words, the value of $f(x)$ only depends on the Hamming weight of x .

Theorem 8 (Feige et al. (1994)) *Let f be a symmetric function. Let $1 \leq k \leq n$ be any integer such that there exists y_0 with Hamming weight $k - 1$ and y_1 with Hamming weight k such that $f(y_0) \neq f(y_1)$. Then the noisy query complexity of f is $\Omega(n \log \min\{k, n - k + 1\})$.*

Proof For every $0 \leq i \leq n$, let W_i be the set of all $x \in \{0, 1\}^n$ with Hamming weight i . Consider the induced subgraph of H_f on vertices $W_{k-1} \cup W_k$. Because f is symmetric, $f(x) = f(y_0)$ for every $x \in W_{k-1}$ and $f(x) = f(y_1)$ for every $x \in W_k$. Hence, the degree of every vertex $x \in W_{k-1}$ in this induced subgraph is $n - k + 1$, and the degree of every $x \in W_k$ is k . The minimum degree is thus $d_1 \geq \min\{k, n - k + 1\}$. Let $\mu(x)$ be the distribution defined in [Theorem 2](#) where $\mu(x) = \frac{\deg_G(x)}{2|E(G)|}$. Note that

$$\Pr_{x \sim \mu}[x \in W_k] = \frac{|W_k| \cdot k}{2|W_k| \cdot k} = 1/2.$$

Therefore, a random x sampled with respect to distribution μ has degree k with $1/2$ probability, and has degree $n - k + 1$ with $1/2$ probability. Hence, letting $d_2 = \max\{k, n - k + 1\}$

$$\Pr_{x \sim \mu}[\deg_G(x) \geq d_2] \geq 1/2.$$

Therefore, [Theorem 2](#) implies that $N_p(f) = \Omega(d_2 \log d_1) = \Omega(\max\{k, n - k + 1\} \log \min\{k, n - k + 1\}) = \Omega(n \log \min\{k, n - k + 1\})$. \blacksquare

In the k -THRESHOLD problem where $1 \leq k \leq n$, we are given $x \in \{0, 1\}^n$, and we need to determine whether $|x| \geq k$. As k -THRESHOLD is a special case of symmetric functions, [Theorem 8](#) immediately implies an $\Omega(n \log \min\{k, n - k + 1\})$ lower bound for the noisy query complexity of k -THRESHOLD, which was originally shown in [Feige et al. \(1994\)](#).

In Graph-Connectivity for N -vertex graphs, the input is a Boolean vector $x \in \{0, 1\}^{\binom{[N]}{2}}$, where each $x_{\{u,v\}}$ for an unordered pair $\{u, v\}$ indicates whether the edge (u, v) exists in the graph. The goal is to determine whether the graph is connected.

Theorem 9 (Gu et al. (2025)) *The noisy query complexity for Graph-Connectivity for N -vertex graphs is $\Omega(N^2 \log N)$.*

Proof Let f denote Graph-Connectivity throughout this proof.

Let S_1 denote the set of all spanning trees on N vertices, and let S_0 denote the set of all disjoint union of two spanning trees $\{T_1, T_2\}$ (T_1 and T_2 are not ordered) where $|T_1| + |T_2| = N$ and $|T_1|, |T_2| \geq N/3$. Clearly, $f(x) = 1$ for $x \in S_1$ and $f(x) = 0$ for $x \in S_0$. We use the following facts to estimate the sizes of S_0 and S_1 :

Fact 10 *The number of labeled N -vertex trees is N^{N-2} .*

Fact 11 (Stirling's formula) *There is a constant $\varepsilon > 0$, so that for all $N \geq k \geq 1$,*

$$\binom{N}{k} \geq \varepsilon \cdot \frac{N^{N+1/2}}{k^{k+1/2}(N-k)^{N-k+1/2}}.$$

We immediately obtain $|S_1| = N^{N-2}$ and

$$\begin{aligned}
 |S_0| &\geq \sum_{k=\lceil N/3 \rceil}^{\lfloor N/2-1 \rfloor} \binom{N}{k} k^{k-2} (N-k)^{N-k-2} \\
 &\geq \sum_{k=\lceil N/3 \rceil}^{\lfloor N/2-1 \rfloor} \Omega \left(\frac{N^{N+1/2}}{k^{k+1/2} (N-k)^{N-k+1/2}} \cdot k^{k-2} (N-k)^{N-k-2} \right) \\
 &\geq \sum_{k=\lceil N/3 \rceil}^{\lfloor N/2-1 \rfloor} \Omega(N^{N-4.5}) \\
 &\geq \Omega(N^{N-3.5}) \geq \Omega(|S_1|/N^{1.5}).
 \end{aligned}$$

Consider the induced subgraph of H_f on the vertex set $S_0 \cup S_1$, which we call G . For every $x = T_1 \sqcup T_2 \in S_0$, $\deg_G(x) \geq \frac{2N^2}{9}$, because we can connect any edge between T_1 and T_2 to obtain a spanning tree. Because G is bipartite where S_0 is one side of the bipartition,

$$|E(G)| \leq N^2 \cdot |S_0|.$$

Hence, we have obtained a subgraph G of H_f where there exists $S_0 \subseteq V(G)$ such that

- $\deg_G(s) = \Omega(N^2)$ for every $s \in S_0$;
- $|S_0| \geq \Omega(N^{-1.5})|V(G)|$;
- $|E(G)| \leq N^2|S_0|$.

Hence, we can apply [Lemma 7](#) with $n := \binom{N}{2}$, $S := S_0$ and $\lambda = 3/4$ to obtain the $\Omega(N^2 \log N)$ lower bound. ■

In the related s - t -Connectivity problem, instead of determining whether a given graph is connected, we need to determine whether two fixed vertices s and t in the graph are connected. Via a similar proof, we can also show the lower bound for s - t -Connectivity.

Theorem 12 (Gu et al. (2025)) *The noisy query complexity for s - t -Connectivity for N -vertex graphs is $\Omega(N^2 \log N)$.*

Proof Sketch As the proof is very similar to the proof of the lower bound for Graph-Connectivity, we only highlight the difference here. The initial subgraph of H_f we will look at is on $S_0 \cup S_1$, where S_1 still denotes the set of all spanning trees, but S_0 denotes the set of all disjoint union of two spanning trees $\{T_1, T_2\}$ where $|T_1| + |T_2| = N$, $|T_1|, |T_2| \geq N/3$, and additionally $s \in T_1, t \in T_2$. This affects the size of S_0 . Now we have

$$|S_0| \geq \sum_{k=\lceil N/3 \rceil}^{\lfloor N/2-1 \rfloor} \binom{N-2}{k-1} k^{k-2} (N-k)^{N-k-2}.$$

Because $\binom{N-2}{k-1} = \binom{N}{k} \cdot \frac{k(N-k)}{N(N-1)} = \Omega\left(\binom{N}{k}\right)$ for k in the specified range, we still obtain the same asymptotic lower bound for $|S_0|$. Hence, the rest of the proof follows verbatim. ■

Acknowledgments

X. L. is supported by NSF Award CCF-2127575. Y. X. is supported by NSF HDR TRIPODS Phase II grant 2217058 (EnCORE Institute).

References

- Michael Ben-Or and Avinatan Hassidim. The bayesian learner is optimal for noisy binary search (and pretty good for quantum as well). In *Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 221–230, 2008. doi: 10.1109/FOCS.2008.58.
- Elwyn R. Berlekamp. *Block coding with noiseless feedback*. PhD thesis, Massachusetts Institute of Technology, 1964.
- M. V. Burnashev and K. Sh. Zigangirov. An interval estimation problem for controlled observations. *Probl. Peredachi Inf.*, 10(3):51–61, 1974.
- Dariusz Dereniowski, Stefan Tiegel, Przemysław Uznański, and Daniel Wolleb-Graf. A framework for searching in graphs in the presence of errors. In *Proceedings of the 2nd Symposium on Simplicity in Algorithms (SOSA)*, pages 4:1–4:17, 2019.
- Dariusz Dereniowski, Aleksander Lukasiewicz, and Przemyslaw Uznanski. Noisy (binary) searching: Simple, fast and correct. In *Proceedings of the 42nd International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 29:1–29:18, 2025.
- Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Ann. Math.*, pages 439–485, 2005.
- Ehsan Emamjomeh-Zadeh, David Kempe, and Vikrant Singhal. Deterministic and probabilistic binary search in graphs. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 519–532, 2016.
- William Evans and Nicholas Pippenger. Average-case lower bounds for noisy boolean decision trees. *SIAM J. Comput.*, 28(2):433–446, 1998.
- Uriel Feige. On the complexity of finite random functions. *Inf. Process. Lett.*, 44(6):295–296, 1992.
- Uriel Feige, Prabhakar Raghavan, David Peleg, and Eli Upfal. Computing with noisy information. *SIAM J. Comput.*, 23(5):1001–1018, 1994.
- Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proc. Am. Math. Soc.*, 124(10):2993–3002, 1996.
- Yuzhou Gu and Yinzhan Xu. Optimal bounds for noisy sorting. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1502–1515, 2023.
- Yuzhou Gu, Xin Li, and Yinzhan Xu. Tight bounds for noisy computation of high-influence functions, connectivity, and threshold. In *Proceedings of 38th Conference on Learning Theory (COLT)*, pages 2540–2591, 2025.

- Michael Horstein. Sequential transmission using noiseless feedback. *IEEE Trans. Inf. Theory*, 9(3): 136–143, 1963.
- Claire Kenyon and Valerie King. On boolean decision trees with faulty nodes. *Random Struct. Algorithms*, 5(3):453–464, 1994.
- Don R Lick and Arthur T White. k-degenerate graphs. *Canadian Journal of Mathematics*, 22(5): 1082–1096, 1970.
- Elchanan Mossel. A quantitative arrow theorem. *Probab. Theory Relat. Fields*, 154(1):49–88, 2012.
- Noam Nisan. Crew prams and decision trees. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 327–335, 1989.
- Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Comput. Complex.*, 4:301–313, 1994.
- Rüdiger Reischuk and Bernd Schmeltz. Reliable computation with noisy circuits and decision trees—a general $n \log n$ lower bound. In *Proceedings 32nd Annual Symposium of Foundations of Computer Science (FOCS)*, pages 602–611, 1991.
- David Rubinfeld. Sensitivity vs. block sensitivity of boolean functions. *Combinatorica*, 15(2): 297–299, 1995.
- Ziao Wang, Nadim Ghaddar, and Lele Wang. Noisy sorting capacity. In *Proceedings of the 2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2541–2546. IEEE, 2022.
- Ziao Wang, Nadim Ghaddar, Banghua Zhu, and Lele Wang. Variable-length insertion-based noisy sorting. In *Proceedings of the 2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1782–1787, 2023.
- Ziao Wang, Nadim Ghaddar, Banghua Zhu, and Lele Wang. Noisy computing of the threshold function. In *Proceedings of the 36th International Conference on Algorithmic Learning Theory*, 2025.
- Banghua Zhu, Ziao Wang, Nadim Ghaddar, Jiantao Jiao, and Lele Wang. On the optimal bounds for noisy computing. In *Proceedings of the 2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1788–1793, 2023.
- Banghua Zhu, Ziao Wang, Nadim Ghaddar, Jiantao Jiao, and Lele Wang. Noisy computing of the OR and MAX functions. *IEEE J. Sel. Areas Inf. Theory*, 5:302–313, 2024.

Appendix A. Proof of the Main Technical Theorem

In this section, we prove [Theorem 2](#), which we recall below:

Theorem 2 *For every constant $p \in (0, 1/2)$ and $\gamma \in (0, 1]$, there exists $c = c(p, \gamma) > 0$ such that the following holds. Let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be a Boolean function and let G be an induced subgraph of H_f with a nonempty edge set. Let $d_1 = \min_{x \in V(G)} \deg_G(x)$. Let μ be the distribution*

on $V(G)$ defined as $\mu(x) = \frac{\deg_G(x)}{2|E(G)|}$ and let $d_2 \geq d_1$ be any number satisfying $\mathbb{P}_{x \sim \mu}(\deg_G(x) \geq d_2) \geq \gamma$. Then $N_p(f) \geq c \cdot d_2 \log(1 + d_1)$.

We will show that there exists a constant D so that theorem statement holds whenever $d_1 \geq D$ (this implies $d_2 \geq D$ as well because $d_2 \geq d_1$). For $d_1 \leq D$, the result follows from $N_p(f) = \Omega(R(f)) = \Omega(\max_x s(f, x)) = \Omega(d_2)$. In this proof, asymptotic notations (such as $o(\cdot)$) are with respect to $d_1 \rightarrow \infty$. In particular, $o(1)$ denotes a function that goes to 0 as $d_1 \rightarrow \infty$. By taking D to be sufficiently large, we can assume these $o(1)$ terms are smaller than an arbitrary constant.

A.1. The Three-Phase Framework

We use the three-phase framework developed in [Feige et al. \(1994\)](#); [Gu et al. \(2025\)](#). Consider the following three-phase problem.

- Phase 1: The algorithm nonadaptively queries each bit $m_1 = c_1 \log d_1$ times using noisy queries for some constant $c_1 > 0$.
- Phase 2: The oracle reveals certain input bits for free.
- Phase 3: The algorithm adaptively makes $m_2 = c_2 d_2$ exact queries in expectation for some constant $c_2 > 0$.

After the three phases, the algorithm needs to output $f(x)$ with success probability at least $1 - \delta$ for some constant $\delta > 0$.

The following reduction comes from [Gu et al. \(2025\)](#).

Proposition 13 *If there is an algorithm that computes $f(x)$ with success probability at least $1 - \delta$ in the worst case using $c_1 c_2 d_2 \log d_1$ queries in expectation in the noisy query model, then it can solve the corresponding three-phase problem with success probability at least $1 - \delta$ in the worst case.*

Proof Sketch Let \mathcal{A} be an algorithm for computing $f(x)$ in the noisy query model. The three-phase problem can simulate the noisy query model and supply \mathcal{A} with noisy bits. Because of Phase 1, this simulation can already support the first $m_1 = c_1 \log d_1$ noisy queries to each bit x_i . If \mathcal{A} makes more than m_1 queries to some bit x_i , the three-phase problem makes an exact query for x_i in Phase 3, and can simulate all future noisy queries to x_i . Since \mathcal{A} makes $c_1 c_2 d_2 \log d_1$ queries in expectation, the expected number of x_i to which \mathcal{A} makes more than m_1 queries is at most $(c_1 c_2 d_2 \log d_1) / m_1 = c_2 d_2$. Hence, the three-phase problem makes at most $c_2 d_2$ queries in Phase 3 in expectation. ■

In the definition of $N_p(f)$, we require the algorithm to be correct with probability at least $2/3$. By repeating $O(\log(1/\delta)) = O(1)$ times and taking the majority answer, the success probability can be boosted to $1 - \delta$. Therefore, to prove [Theorem 2](#), it suffices to show the following hardness result for the three-phase problem.

Proposition 14 *There exists $c_1, c_2, \delta > 0$ such that no algorithm can solve the corresponding three-phase problem with success probability at least $1 - \delta$ in the worst case.*

Let G be the induced subgraph of H_f as in the statement of [Theorem 2](#) and μ be the corresponding distribution on $V(G)$ where $\mu(x) = \frac{\deg_G(x)}{2|E(G)|}$. We will show that the three-phase problem is hard even if the input x is drawn from the distribution μ .

A.2. Phase 1 and 2

Let $m_1 = c_1 \log d_1$. Let a_i denote the number of Phase 1 queries to bit i whose answers are 1. Then $a_i \sim \text{Bin}(m_1, 1-p)$ for $x_i = 1$ and $a_i \sim \text{Bin}(m_1, p)$ for $x_i = 0$. For $0 \leq j \leq m_1$, define

$$p_j = \mathbb{P}(\text{Bin}(m_1, 1-p) = j) = \binom{m_1}{j} (1-p)^j p^{m_1-j}.$$

Let $I = [pm_1 - m_1^{0.6}, pm_1 + m_1^{0.6}]$.

In Phase 2, the oracle reveals the following information:

- Step 2a: The oracle reveals all x_i with $a_i \notin I$.
- Step 2b: The oracle reveals every $x_i = 1$ independently with probability q_{a_i} , where $q_k = 1 - \frac{p_{m_1-k} p_{k_l}}{p_k p_{m_1-k_l}}$ (for integer $k \in I$) and $k_l = \lceil pm_1 - m_1^{0.6} \rceil$.

Step 2b is a reweighting process so that the posterior probability that each coordinate is 1 does not depend on a_i . Intuitively, if $x_i = 0$, then the probability that $a_i = k$ for some value of k and x_i is unrevealed after Step 2b is p_{m_1-k} . On the other hand, if $x_i = 1$, then the probability that $a_i = k$ and x_i is unrevealed after Step 2b is $p_k \cdot (1 - q_k) = \frac{p_{m_1-k} p_{k_l}}{p_{m_1-k_l}}$. The ratio between these two values is always the same value $\frac{p_{k_l}}{p_{m_1-k_l}}$ regardless of the value of k . Thus intuitively, a_i does not give information on x_i .

The difference between [Gu et al. \(2025\)](#)'s and our Phase 1 and 2 is that we choose $m_1 = c_1 \log d_1$, while [Gu et al. \(2025\)](#) has $m_1 = c_1 \log n$. The analysis remains largely the same, and we only need to replace their dependency on n with d_1 . Hence, by ([Gu et al., 2025](#), Definition 9), observations in Phase 1 and 2 are equivalent to the following process:

- Observe every $x_i = 1$ independently with probability

$$p_+ = 1 - \sum_{k \in I} p_k (1 - q_k) = 1 - d_1^{-c_3 \pm o(1)},$$

where $c_3 = c_1(1-2p) \log \frac{1-p}{p}$. By picking sufficiently small c_1 , we can guarantee $c_3 < 1$.

- Observe every $x_i = 0$ independently with probability

$$p_- = \mathbb{P}(\text{Bin}(m_1, p) \notin I) = o(1).$$

Let $z \in \{0, 1, \star\}^n$ denote the revealed information up to Phase 2. That is, $z_i = x_i$ if x_i is revealed and $z_i = \star$ otherwise. So $\mathbb{P}(z_i = 1 | x_i = 1) = p_+$, $\mathbb{P}(z_i = 0 | x_i = 0) = p_-$.

For fixed z , let μ_z denote the distribution of x conditioned on z . Let $\mathcal{X}_z = \text{supp } \mu_z$. Note that $\mathcal{X}_z = \{x : z_i \in \{x_i, \star\} \forall i \in [n]\} \cap \text{supp } \mu$.

That is, μ_z is the posterior distribution of x given the observations in Phase 1 and 2. Let ν_x denote the distribution of z conditioned on x and let ν denote the overall distribution of z . Note that by Bayes,

$$\nu_x(z) = \frac{\mu_z(x) \nu(z)}{\mu(x)} \propto \frac{\mu_z(x) \nu(z)}{\deg_G(x)}. \quad (3)$$

A.3. Balance Lemma

In this section, we establish a condition for a posterior distribution to be balanced (Lemma 15). That is, given the observations, one cannot guess $f(x)$ with error probability $o(1)$. As an example, we show that at the end of Phase 2, with probability $1 - o(1)$, μ_z is balanced.

For $(x, y) \in E(G)$, define edge weight

$$w_z(x, y) = \sqrt{\frac{\mu_z(x)\mu_z(y)}{\deg_G(x)\deg_G(y)}}.$$

For $x \in \mathcal{X}_z$, define weighted degree as

$$\text{wdeg}_z(x) = \mu_z(x)^{-1} \sum_{y:(x,y) \in E(G)} w_z(x, y).$$

(When $\mu_z(x) = 0$, define $\text{wdeg}_z(x) = 0$.) The weighted degree satisfies the nice property that

$$\sum_{x \in f^{-1}(0)} \mu_z(x) \text{wdeg}_z(x) = \sum_{x \in f^{-1}(1)} \mu_z(x) \text{wdeg}_z(x).$$

Define average weighted degree as

$$\overline{\text{wdeg}}_z = \mathbb{E}_{x \sim \mu_z}[\text{wdeg}_z(x)] = 2 \sum_{x \in f^{-1}(0)} \mu_z(x) \text{wdeg}_z(x) = 2 \sum_{x \in f^{-1}(1)} \mu_z(x) \text{wdeg}_z(x).$$

The following lemma shows that if the weighted degree is not too concentrated on a small set, then the posterior distribution is balanced. In the following, we use the convention that for any measure η and any nonnegative function h , the product $\eta(A) \max_{x \in A} h(x)$ is defined to be 0 when $A = \emptyset$.

Lemma 15 (Balance Lemma) *Let $A \subseteq \mathcal{X}_z$. Suppose*

$$\max_{x \in \mathcal{X}_z \setminus A} \text{wdeg}_z(x) \leq C \overline{\text{wdeg}}_z$$

for some $C > 0$. Then

$$\min \{ \mu_z(f^{-1}(0)), \mu_z(f^{-1}(1)) \} \geq \frac{1}{2C} - \frac{\mu_z(A) \max_{x \in A} \text{wdeg}_z(x)}{C \overline{\text{wdeg}}_z}.$$

Proof

$$\begin{aligned} \frac{1}{2} \overline{\text{wdeg}}_z &= \sum_{x \in f^{-1}(0)} \mu_z(x) \text{wdeg}_z(x) \\ &\leq \sum_{x \in A} \mu_z(x) \text{wdeg}_z(x) + \sum_{x \in f^{-1}(0) \setminus A} \mu_z(x) \text{wdeg}_z(x) \\ &\leq \mu_z(A) \max_{x \in A} \text{wdeg}_z(x) + \mu_z(f^{-1}(0)) \max_{x \in \mathcal{X}_z \setminus A} \text{wdeg}_z(x) \\ &\leq \mu_z(A) \max_{x \in A} \text{wdeg}_z(x) + C \overline{\text{wdeg}}_z \mu_z(f^{-1}(0)). \end{aligned}$$

Rearranging terms, we get

$$\mu_z(f^{-1}(0)) \geq \frac{1}{2C} - \frac{\mu_z(A) \max_{x \in A} \text{wdeg}_z(x)}{C \text{wdeg}_z}.$$

By the same argument, we get the bound for $\mu_z(f^{-1}(1))$. ■

Let $x, y \in \mathcal{X}_z$, $(x, y) \in E(G)$ differ at bit i . Then we must have $z_i = \star$ and

$$\begin{aligned} \frac{w_z(x, y)}{\mu_z(x)} &= \sqrt{\frac{\mu_z(y)}{\mu_z(x)} \cdot \frac{1}{\deg_G(x) \deg_G(y)}} \\ &= \frac{1}{\deg_G(x)} \sqrt{\frac{\mu_z(y)}{\mu_z(x)} \cdot \frac{\deg_G(x)}{\deg_G(y)}} \\ &= \frac{1}{\deg_G(x)} \sqrt{\frac{\nu_y(z)}{\nu_x(z)}} && \text{(Equation (3))} \\ &= \frac{1}{\deg_G(x)} \left(\frac{1 - p_+}{1 - p_-} \right)^{(y_i - x_i)/2}, \end{aligned}$$

In particular,

$$\max_{x \in \mathcal{X}_z} \text{wdeg}_z(x) \leq \sqrt{\frac{1 - p_-}{1 - p_+}}. \quad (4)$$

We now study the distribution of weighted degree. For $x \in V(G)$, let $N_+(x) := \{y : (x, y) \in E(G), |y| = |x| + 1\}$ denote the up-neighbors of x , and $N_-(x) := \{y : (x, y) \in E(G), |y| = |x| - 1\}$ denote the down-neighbors of x . Then for $x \in \mathcal{X}_z$, we have

$$\text{wdeg}_z(x) = \deg_G(x)^{-1} \left(|N_+(x) \cap \mathcal{X}_z| \sqrt{\frac{1 - p_+}{1 - p_-}} + |N_-(x) \cap \mathcal{X}_z| \sqrt{\frac{1 - p_-}{1 - p_+}} \right). \quad (5)$$

We have $|N_+(x) \cap \mathcal{X}_z| \leq |N_+(x)| \leq \deg_G(x)$, so the contribution of the first term is $\leq \sqrt{\frac{1 - p_+}{1 - p_-}}$. For the second term, we need to prove some concentration bounds on $|N_-(x) \cap \mathcal{X}_z|$.

When $z \sim \nu_x$, each down-neighbor y of x is in \mathcal{X}_z independently with probability $1 - p_+$ (it happens when the corresponding bit of x is not revealed, which happens with probability $1 - p_+$). So $|N_-(x) \cap \mathcal{X}_z| \sim \text{Bin}(|N_-(x)|, 1 - p_+)$.

By Chernoff bound, $|N_-(x)| \leq \deg_G(x)$ and $\deg_G(x) \geq d_1$, we have

$$\begin{aligned} &\mathbb{P}_{z \sim \nu_x}(|N_-(x) \cap \mathcal{X}_z| \geq (1 + \varepsilon) \deg_G(x)(1 - p_+)) \\ &\leq \exp\left(-\frac{\varepsilon^2}{2 + \varepsilon} \deg_G(x)(1 - p_+)\right) \\ &\leq \exp\left(-\frac{\varepsilon^2}{2 + \varepsilon} d_1^{1 - c_3 \pm o(1)}\right), \end{aligned}$$

for any $\varepsilon > 0$.

For given z , let $A_{\text{bad},z} = \{x \in \mathcal{X}_z : |N_-(x) \cap \mathcal{X}_z| \geq 2 \deg_G(x)(1 - p_+)\}$. Then (we pick $\varepsilon = 1$, and the $\frac{\varepsilon^2}{2+\varepsilon}$ term can be absorbed to the $o(1)$ factor)

$$\mathbb{E}_{z \sim \nu} [\mu_z(A_{\text{bad},z})] = \mathbb{P}_{x \sim \mu} \mathbb{P}_{z \sim \nu_x}(x \in A_{\text{bad},z}) \leq \exp\left(-d_1^{1-c_3 \pm o(1)}\right). \quad (6)$$

By Markov's inequality, we have

$$\mathbb{P}_{z \sim \nu} \left(\mu_z(A_{\text{bad},z}) \geq \sqrt{\exp\left(-d_1^{1-c_3 \pm o(1)}\right)} \right) \leq \sqrt{\exp\left(-d_1^{1-c_3 \pm o(1)}\right)}.$$

Note that $\sqrt{\exp\left(-d_1^{1-c_3 \pm o(1)}\right)} = \exp\left(-d_1^{1-c_3 \pm o(1)}\right)$, for suitable choices of the $o(1)$ terms. Thus,

$$\mathbb{P}_{z \sim \nu} \left(\mu_z(A_{\text{bad},z}) \geq \exp\left(-d_1^{1-c_3 \pm o(1)}\right) \right) \leq \exp\left(-d_1^{1-c_3 \pm o(1)}\right). \quad (7)$$

For $x \in \mathcal{X}_z \setminus A_{\text{bad},z}$, we have

$$\begin{aligned} \text{wdeg}_z(x) &\leq \deg_G(x)^{-1} \left(\deg_G(x) \sqrt{\frac{1-p_+}{1-p_-}} + 2 \deg_G(x)(1-p_+) \sqrt{\frac{1-p_-}{1-p_+}} \right) \\ &\leq 4\sqrt{(1-p_+)(1-p_-)}. \end{aligned} \quad (8)$$

We now have bounds on $\mu_z(A_{\text{bad},z})$ and $\text{wdeg}_z(x)$ for $x \in \mathcal{X}_z \setminus A_{\text{bad},z}$. Next, we establish a lower bound on $\overline{\text{wdeg}}_z$.

Fix $x \in V(G)$. Then $\max\{|N_+(x)|, |N_-(x)|\} \geq \deg_G(x)/2 \geq d_1/2$. If $|N_-(x)| \geq \deg_G(x)/2$, then

$$\mathbb{P}_{z \sim \nu_x} \left(|N_-(x) \cap \mathcal{X}_z| \leq \frac{1}{4} \deg_G(x)(1-p_+) \right) \leq \exp\left(-d_1^{1-c_3 \pm o(1)}\right).$$

Similar to $|N_-(x) \cap \mathcal{X}_z|$, $|N_+(x) \cap \mathcal{X}_z| \sim \text{Bin}(|N_+(x)|, 1-p_-)$. So if $|N_+(x)| \geq \deg_G(x)/2$, then

$$\mathbb{P}_{z \sim \nu_x} \left(|N_+(x) \cap \mathcal{X}_z| \leq \frac{1}{4} \deg_G(x)(1-p_-) \right) \leq \exp\left(-d_1^{1-o(1)}\right). \quad (9)$$

Combining the two cases and recalling [Equation \(5\)](#), we get

$$\begin{aligned} &\mathbb{P}_{z \sim \nu} \mathbb{P}_{x \sim \mu_z} \left(\text{wdeg}_z(x) \leq \frac{1}{4} \sqrt{(1-p_+)(1-p_-)} \right) \\ &= \mathbb{P}_{x \sim \mu} \mathbb{P}_{z \sim \nu_x} \left(\text{wdeg}_z(x) \leq \frac{1}{4} \sqrt{(1-p_+)(1-p_-)} \right) \\ &\leq \exp\left(-d_1^{1-c_3 \pm o(1)}\right). \end{aligned}$$

By Markov's inequality, we have

$$\mathbb{P}_{z \sim \nu} \left(\mathbb{P}_{x \sim \mu_z} \left(\text{wdeg}_z(x) \leq \frac{1}{4} \sqrt{(1-p_+)(1-p_-)} \right) > 1/2 \right) \leq \exp\left(-d_1^{1-c_3 \pm o(1)}\right).$$

So

$$\mathbb{P}_{z \sim \nu} \left(\overline{\text{wdeg}}_z \leq \frac{1}{8} \sqrt{(1-p_+)(1-p_-)} \right) \leq \exp \left(-d_1^{1-c_3 \pm o(1)} \right). \quad (10)$$

Let \mathcal{E}_1 be the event that $\mu_z(A_{\text{bad},z}) \leq \exp \left(-d_1^{1-c_3 \pm o(1)} \right)$ (recall Equation (7)). Let \mathcal{E}_2 be the event that $\overline{\text{wdeg}}_z \geq \frac{1}{8} \sqrt{(1-p_+)(1-p_-)}$. Then by Equations (7) and (10), we have

$$\mathbb{P}_{z \sim \nu} (\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - \exp \left(-d_1^{1-c_3 \pm o(1)} \right). \quad (11)$$

Conditioned on $\mathcal{E}_1 \cap \mathcal{E}_2$, we apply Lemma 15. By Equation (8) and definition of \mathcal{E}_2 , we can take $C = 32$ in Lemma 15. So we have

$$\begin{aligned} \min \{ \mu_z(f^{-1}(0)), \mu_z(f^{-1}(1)) \} &\geq \frac{1}{64} - \frac{\mu_z(A_{\text{bad},z}) \max_{x \in A_{\text{bad},z}} \overline{\text{wdeg}}_z(x)}{32 \overline{\text{wdeg}}_z} \\ &\geq \frac{1}{64} - \frac{\exp \left(-d_1^{1-c_3 \pm o(1)} \right) \sqrt{\frac{1-p_-}{1-p_+}}}{32 \cdot \frac{1}{8} \sqrt{(1-p_+)(1-p_-)}} \\ &\geq \frac{1}{64} - o(1). \end{aligned}$$

This means that with probability at least $1 - \exp \left(-d_1^{1-c_3 \pm o(1)} \right)$ over z , the posterior distribution μ_z is balanced after Phase 2.

A.4. Phase 3

In Phase 3, the algorithm adaptively queries $m_2 = c_2 d_2$ bits in expectation. We will show that with constant probability, after the m_2 queries, $\overline{\text{wdeg}}_z$ decreases by at most a constant factor. Then we can use Lemma 15 again to show that the posterior distribution is still balanced.

Let $B = \{x \in V(G) : \deg_G(x) \geq d_2\}$. By the definition of d_2 , $\mu(B) \geq \gamma$. For every $x \in B$, we have $\max\{|N_+(x)|, |N_-(x)|\} \geq \deg_G(x)/2 \geq d_2/2$. Let $B_+ = \{x \in B : |N_+(x)| \geq \deg_G(x)/2\}$ and $B_- = \{x \in B : |N_-(x)| \geq \deg_G(x)/2\}$. Then $\max\{\mu(B_+), \mu(B_-)\} \geq \mu(B)/2 \geq \gamma/2$. WLOG assume that $\mu(B_+) \geq \gamma/2$. (If not, we can define $g(x) = f(\mathbf{1} - x)$ and prove a lower bound for g instead. Clearly, $N_p(f) = N_p(g)$.)

We show that after Phase 2, with constant probability, vertices in B_+ contribute a constant amount to $\overline{\text{wdeg}}_z$. Let $B_{\text{good},z} = \{x \in B_+ : |N_+(x) \cap \mathcal{X}_z| \geq \frac{1}{4} \deg_G(x)(1-p_-)\}$. For $x \in B_+$, by the same analysis as Equation (9) (and now we have $\deg_G(x) \geq d_2$, instead of the bound $\deg_G(x) \geq d_1$ used in Equation (9)),

$$\mathbb{P}_{z \sim \nu_x} (x \in B_{\text{good},z}) \geq 1 - \exp \left(-d_2^{1-c_3 \pm o(1)} \right) = 1 - o(1).$$

So

$$\mathbb{E}_{z \sim \nu} \mu_z(B_{\text{good},z}) = \mathbb{E}_{x \sim \mu} [\mathbb{1}_{x \in B_+} \mathbb{P}_{z \sim \nu_x} (x \in B_{\text{good},z})] \geq \gamma/2 - o(1).$$

By Corollary 6, we have

$$\mathbb{P}_{z \sim \nu} (\mu_z(B_{\text{good},z}) \geq \gamma/4) \geq \gamma/4 - o(1).$$

Let \mathcal{E}_3 be the event that after Phase 2, $\mu_z(B_{\text{good},z}) \geq \gamma/4$. Then $\mathbb{P}_{z \sim \nu}(\mathcal{E}_3) \geq \gamma/4 - o(1)$.

For any k , define $\mathcal{E}_{\leq k} = \bigcap_{j=1}^k \mathcal{E}_j$. Then $\mathbb{P}_{z \sim \nu}(\mathcal{E}_{\leq 3}) \geq \gamma/4 - o(1)$ by union bound.

Note that for $x \in B_{\text{good},z}$, by Equation (5), we have

$$\text{wdeg}_z(x) \geq \frac{1}{4} \sqrt{(1-p_+)(1-p_-)}.$$

So at the beginning of Phase 3, conditioned on $\mathcal{E}_{\leq 3}$, we have

$$\overline{\text{wdeg}}_z \geq \sum_{x \in B_{\text{good},z}} \mu_z(x) \text{wdeg}_z(x) \geq \frac{\gamma}{16} \sqrt{(1-p_+)(1-p_-)}.$$

This bound is weaker than Equation (10), but the analysis is more robust under queries in Phase 3.

In the following, we will study the effect of queries in Phase 3. Let q be the number of queries in Phase 3. Note that $\mathbb{E}[q] = m_2$. Let i_1, \dots, i_q denote the queried bits in Phase 3 and b_1, \dots, b_q denote the corresponding query results. For $0 \leq t \leq q$, let $\mu_{z,t}$ denote the posterior distribution after t queries in Phase 3. Then $\mu_{z,0} = \mu_z$.

For the t -th query, the query result b_t is equal to $b \in \{0, 1\}$ with probability $\mathbb{P}_{x \sim \mu_{z,t-1}}(x_{i_t} = b)$. For a fixed x , $\mu_{z,t}(x)$ changes to 0 if $x_{i_t} \neq b_t$ and changes to $\frac{\mu_{z,t-1}(x)}{\mathbb{P}_{x' \sim \mu_{z,t-1}}(x'_{i_t} = b_t)}$ if $x_{i_t} = b_t$. The expectation is equal to $\mu_{z,t-1}(x)$. Therefore, $\mu_{z,t}(x)$ is a martingale over the query results. In particular, $\mathbb{E}[\mu_{z,q} | z] = \mu_z$ by the optional stopping theorem.

Conditioned on $\mathcal{E}_{\leq 3}$,

$$\mathbb{E}[\mu_{z,q}(B_{\text{good},z})] = \mathbb{E}[\mu_z(B_{\text{good},z})] \geq \gamma/4.$$

By Corollary 6, we have

$$\mathbb{P}(\mu_{z,q}(B_{\text{good},z}) \geq \gamma/8) \geq \gamma/8.$$

Let \mathcal{E}_4 be the event that $\mu_{z,q}(B_{\text{good},z}) \geq \gamma/8$. Then $\mathbb{P}(\mathcal{E}_4 | \mathcal{E}_{\leq 3}) \geq \gamma/8$.

Because $\mathbb{E}[\mu_{z,q} | z] = \mu_z$, we have $\mathbb{E}[\mu_{z,q}(A_{\text{bad},z})] = \mathbb{E}[\mu_z(A_{\text{bad},z})] \leq \exp(-d_1^{1-c_3 \pm o(1)})$ (recall Equation (6)). By Markov's inequality, we have (similar to Equation (7))

$$\mathbb{P}(\mu_{z,q}(A_{\text{bad},z}) \geq \exp(-d_1^{1-c_3 \pm o(1)})) \leq \exp(-d_1^{1-c_3 \pm o(1)}).$$

Let \mathcal{E}_5 be the event that $\mu_{z,q}(A_{\text{bad},z}) \leq \exp(-d_1^{1-c_3 \pm o(1)})$. Then

$$\mathbb{P}(\mathcal{E}_5 | \mathcal{E}_{\leq 3}) \geq 1 - \mathbb{P}(\neg \mathcal{E}_5 | \mathcal{E}_{\leq 3}) \geq 1 - \mathbb{P}(\neg \mathcal{E}_5) / \mathbb{P}(\mathcal{E}_{\leq 3}) \geq 1 - o(1).$$

By union bound, $\mathbb{P}(\mathcal{E}_{\leq 5}) = \mathbb{P}(\mathcal{E}_{\leq 3}) \mathbb{P}(\mathcal{E}_4 \cap \mathcal{E}_5 | \mathcal{E}_{\leq 3}) \geq (\gamma/4 - o(1))(\gamma/8 - o(1)) = \gamma^2/32 - o(1)$.

Because $\mathbb{E}[q] = m_2$, we have

$$\mathbb{E}[q | \mathcal{E}_{\leq 5}] \leq \frac{m_2}{\mathbb{P}(\mathcal{E}_{\leq 5})} \leq (1 + o(1)) \frac{32m_2}{\gamma^2}.$$

So conditioned on $\mathcal{E}_{\leq 5}$, with probability at least $1/2 - o(1)$, we have $q \leq \frac{64m_2}{\gamma^2} = \frac{64c_2d_2}{\gamma^2}$. Let \mathcal{E}_6 be the event that $q \leq \frac{64c_2d_2}{\gamma^2}$. Then $\mathbb{P}(\mathcal{E}_{\leq 6}) \geq \gamma^2/64 - o(1)$.

From now on we condition on $\mathcal{E}_{\leq 6}$. We would like to show that $\mu_{z,q}$ is balanced. For $0 \leq t \leq q$, define $\mathcal{X}_{z,t} = \text{supp } \mu_{z,t}$ and

$$\begin{aligned} w_{z,t}(x,y) &= \sqrt{\frac{\mu_{z,t}(x)\mu_{z,t}(y)}{\deg_G(x)\deg_G(y)}}, \\ \text{wdeg}_{z,t}(x) &= \mu_{z,t}(x)^{-1} \sum_{y:(x,y) \in E(G)} w_{z,t}(x,y), \\ \overline{\text{wdeg}}_{z,t} &= \mathbb{E}_{x \sim \mu_{z,t}}[\text{wdeg}_{z,t}(x)]. \end{aligned}$$

We have the following variation of [Lemma 15](#) for $\mu_{z,q}$.

Lemma 16 (Balance Lemma) *Let $A \subseteq \mathcal{X}_{z,q}$. Suppose*

$$\max_{x \in \mathcal{X}_{z,q} \setminus A} \text{wdeg}_{z,q}(x) \leq C \overline{\text{wdeg}}_{z,q}$$

for some $C > 0$. Then

$$\min \{ \mu_{z,q}(f^{-1}(0)), \mu_{z,q}(f^{-1}(1)) \} \geq \frac{1}{2C} - \frac{\mu_{z,q}(A) \max_{x \in A} \text{wdeg}_{z,q}(x)}{C \overline{\text{wdeg}}_{z,q}}.$$

The proof is essentially the same as [Lemma 15](#) and is omitted.

For $x \in \mathcal{X}_{z,t}$, we have

$$\begin{aligned} \text{wdeg}_{z,t}(x) &= \sum_{y:(x,y) \in E(G)} \sqrt{\frac{\mu_{z,t}(y)}{\mu_{z,t}(x)} \cdot \frac{1}{\deg_G(x)\deg_G(y)}} \\ &= \sum_{y:(x,y) \in E(G)} \mathbb{1}_{y \in \mathcal{X}_{z,t}} \sqrt{\frac{\mu_z(y)}{\mu_z(x)} \cdot \frac{1}{\deg_G(x)\deg_G(y)}} \\ &= \deg_G(x)^{-1} \left(|N_+(x) \cap \mathcal{X}_{z,t}| \sqrt{\frac{1-p_+}{1-p_-}} + |N_-(x) \cap \mathcal{X}_{z,t}| \sqrt{\frac{1-p_-}{1-p_+}} \right). \end{aligned}$$

For any $x \in B_{\text{good},z} \cap \mathcal{X}_{z,q}$, $|N_+(x) \cap \mathcal{X}_{z,t}|$ decreases by at most one after each query. So

$$|N_+(x) \cap \mathcal{X}_{z,q}| \geq |N_+(x) \cap \mathcal{X}_z| - q \geq \frac{1}{4} \deg_G(x)(1-p_-) - \frac{64c_2d_2}{\gamma^2}. \quad (12)$$

Because $\deg_G(x) \geq d_2$, for $c_2 = \frac{1}{1024}\gamma^2$, the last expression is at least $\frac{1}{8} \deg_G(x)(1-p_-)$ for sufficiently large d_1 as $p_- = o(1)$. Then

$$\begin{aligned} \overline{\text{wdeg}}_{z,q} &\geq \sum_{x \in B_{\text{good},z} \cap \mathcal{X}_{z,q}} \mu_{z,q}(x) \text{wdeg}_{z,q}(x) \\ &\geq \mu_{z,q}(B_{\text{good},z}) \cdot \frac{1}{8} \sqrt{(1-p_+)(1-p_-)} && \text{(by definition of } B_{\text{good},z}\text{)} \\ &\geq \frac{\gamma}{64} \sqrt{(1-p_+)(1-p_-)}. && \text{(by definition of } \mathcal{E}_4\text{)} \end{aligned}$$

On the other hand $\text{wdeg}_{z,q}(x) \leq \text{wdeg}_z(x)$ for all $x \in \mathcal{X}_{z,q}$. So

$$\max_{x \in \mathcal{X}_{z,q} \setminus A_{\text{bad},z}} \text{wdeg}_{z,q}(x) \leq \max_{x \in \mathcal{X}_z \setminus A_{\text{bad},z}} \text{wdeg}_z(x) \stackrel{\text{Equation (8)}}{\leq} 4\sqrt{(1-p_+)(1-p_-)} \leq \frac{256}{\gamma} \overline{\text{wdeg}_{z,q}}.$$

Finally, let $A := A_{\text{bad},z} \cap \mathcal{X}_{z,q}$, then

$$\begin{aligned} & \frac{\mu_{z,q}(A) \max_{x \in A} \text{wdeg}_{z,q}(x)}{\overline{\text{wdeg}_{z,q}}} \\ & \leq \frac{\mu_{z,q}(A) \max_{x \in A} \text{wdeg}_z(x)}{\frac{\gamma}{64} \sqrt{(1-p_+)(1-p_-)}} && \text{(by definition of } \mathcal{E}_5) \\ & \leq \frac{\exp\left(-d_1^{1-c_3 \pm o(1)}\right) \sqrt{\frac{1-p_-}{1-p_+}}}{\frac{\gamma}{64} \sqrt{(1-p_+)(1-p_-)}} && \text{(Equation (4))} \\ & = o(1). \end{aligned}$$

We now apply [Lemma 16](#) with $C = \frac{256}{\gamma}$ and A to get that $\mu_{z,q}$ is balanced. That is, the posterior distribution $\mu_{z,q}$ satisfies

$$\min \{ \mu_{z,q}(f^{-1}(0)), \mu_{z,q}(f^{-1}(1)) \} \geq \frac{\gamma}{512} - o(1)$$

conditioned on $\mathcal{E}_{\leq 6}$. Recall $\mathbb{P}(\mathcal{E}_{\leq 6}) \geq \gamma^2/64 - o(1)$. Therefore, for small enough $\delta > 0$, any algorithm cannot compute $f(x)$ with error probability at most δ in the three-phase problem. This finishes the proof.