Phase Transitions in Group Testing

Jonathan Scarlett and Volkan Cevher
Laboratory for Information and Inference Systems (LIONS), EPFL
Email: {jonathan.scarlett,volkan.cevher}@epfl.ch

Abstract

The group testing problem consists of determining a sparse subset of a set of items that are "defective" based on a set of possibly noisy tests, and arises in areas such as medical testing, fault detection, communication protocols, pattern matching, and database systems. We study the fundamental limits of any group testing procedure regardless of its computational complexity. In the noiseless case with the number of defective items k scaling with the total number of items p as $O(p^\theta)$ ($\theta \in (0,1)$), we show that the probability of reconstruction error tends to one when $n \le k \log_2 \frac{p}{k}(1+o(1))$, but vanishes when $n \ge c(\theta) k \log_2 \frac{p}{k}(1+o(1))$, for some explicit constant $c(\theta)$. For $\theta \le \frac{1}{3}$, we show that $c(\theta) = 1$, thus providing an exact threshold on the required number measurements, i.e. a phase transition, which was previously known only in the limit as $\theta \to 0$. Analogous necessary and sufficient conditions are derived for the noisy setting, and also for a relaxed partial recovery criterion.

1 Introduction

The group testing problem consists of determining a small subset S of "defective" items within a larger set of items $\{1,\dots,p\}$. This problem has a history in areas such as medical testing and fault detection, but has regained significant attention with following new applications in areas such as communication protocols [1], pattern matching [2], and database systems [3], and new connections with compressive sensing [4, 5]. Formally, the goal is to determine S via a number of tests, each taking the form

$$Y = \mathbb{1}\left\{ \bigcup_{i \in S} \{X_i = 1\} \right\} \oplus Z, \tag{1.1}$$

where the measurement vector $X = (X_1,\dots,X_p) \in \{0,1\}^p$ indicates which items are included in the test, Z is a noise term, Y is the resulting observation, and $\oplus$ denotes modulo-2 addition. In words, the output indicates, in a possibly noisy fashion, whether at least one defective item is present in the items corresponding to $X_i = 1$. One wishes to minimize the total number of tests n while still ensuring the reliable recovery of S.

The above model is simple but highly powerful, and itself comes with a variety of variations:

- The measurement vectors may be designed deterministically [6–8], or one may seek to characterize the behavior when they are generated randomly [9–12]. Among the latter class, the prevalent distribution is that in which the entries of X are independent and identically distributed (i.i.d.) Bernoulli variables; these have the best known theoretical guarantees in terms of the number of measurements.
- One may seek practical decoding techniques having low computational complexity and storage [10, 11, 13], whereas a complementary line of research considers *measurement-optimal* fundamental limits that hold regardless of such considerations. Such studies help to assess practical methods and determine the level of further improvement possible.

In this paper, we develop *phase transitions*, i.e., exact asymptotic thresholds on the required number of measurements including constant factors, for Bernoulli designs and measurement-optimal recovery algorithms. Early studies of this type were performed by Malyutov [9], and more recent studies include those of Atia and Saligrama [12], Aldridge *et al.* [11, 14], and Laarhoven [15]. In the case that the number of defective items k does not scale with p, the fundamental limits are well-understood for both the noiseless and noisy settings [9, 12, 15]; for example, in the noiseless case, the smallest possible number of measurements with vanishing error probability behaves as $\big(k \log_2 \frac{p}{k}\big)(1+o(1))$, which is in fact the same threshold as that for optimal adaptive measurements [16] (i.e., designs for which each test may depend on previous outcomes).

This work was supported in part by the European Commission under Grant ERC Future Proof, SNF 200021-146750 and SNF CRSII2-147633. The first author acknowledges support from the 'EPFL Fellows' fellowship programme co-funded by Marie Sklodowska-Curie, Horizon2020 Grant agreement no. 665667.

Surprisingly, there remain significant gaps in the best known upper and lower bounds on n when k scales with p, which is of considerable interest in applications where the number of defective items is "not too small". In this paper, we close these gaps in several regimes of interest. Our main contributions are as follows:

- We develop novel analysis techniques providing a significant departure from existing approaches based on tools such as maximum-likelihood decoding and Fano's inequality [9,12]. Specifically, we introduce *information densities* of the form $\imath(x;y) := \log\frac{P_{Y|X}(y|x)}{P_Y(y)}$ (defined formally in the sequel), and develop upper and lower bounds showing that the error probability of a measurement-optimal recovery algorithm (or "decoder") is precisely characterized by tail probabilities involving i.i.d. summations of these quantities, thus permitting the derivation of sample complexity bounds via concentration inequalities.
- Using these techniques, we show that the error probability (defined formally in the sequel) undergoes a *phase transition*, approaching one when n is slightly below a threshold, while approaching zero when n is slightly above the same threshold. Specifically, in the noiseless case, we prove the above-mentioned tightness of the threshold $\big(k\log_2\frac{p}{k}\big)(1+o(1))$ whenever $k = p^\theta$ for some $\theta \in \big(0, \frac{1}{3}\big)$, thus improving significantly on the previously-known condition $\theta \to 0$. Similarly, with additive modulo-2 Bernoulli noise, we obtain an analogous threshold for sufficiently small θ. In each of these settings, it immediately follows that *non-adaptive Bernoulli measurements yield the same phase transition as that of optimal adaptive measurements.* Moreover, we show that even when only a proportion $1-\alpha^*$ of the entries in S needs to be recovered, the corresponding threshold decreases by at most a factor of $1-\alpha^*$, and hence there is little to be gained by considering this relaxed criterion. This is in stark contrast with compressive sensing problems, where moving to partial recovery can lead to immense savings [17].

We note that the condition $\mathbb{P}[\text{error}] \to 1$ improves on the usual condition $\mathbb{P}[\text{error}] \not\to 0$ arising from Fano's inequality; while the stronger statement was previously given in [18] for Bernoulli designs, our approach has the key advantage of extending immediately to other sparsity problems in which the observations are continuous (*cf.* [19]).

Another particularly related work is that of Mézard *et al.* [20], who derived phase transitions for various random measurement designs in the noiseless setting, for certain scaling regimes. However, some of the arguments therein are based on a "no short loops" assumption that is only verified rigorously for $\theta \geq \frac{5}{6}$,[1] and non-rigorously for $\theta \geq \frac{2}{3}$. In contrast, in this paper we obtain phase transitions for $\theta \leq \frac{1}{3}$, which does not overlap with the range of interest in [20]. In fact, it is verified numerically in [20] that the assumption regarding short loops is *invalid* for $\theta = \frac{1}{3}$.

Finally, we briefly comment on practical decoders. In the noiseless setting with adaptive measurements, an algorithm by Hwang [7] is known to achieve the optimal phase transition. In the non-adaptive setting, several techniques have been shown to be optimal in terms of scaling laws [10, 11, 13], requiring $O(k\log p)$ measurements and polynomial space and time. However, the implied constants in the number of measurements are generally suboptimal. The results of this paper provide key insights into which of these gaps are fundamental; see Section 2 for details, as well as Figures 1–2.

1.1 Problem Statement

We consider both a noiseless and noisy variant of the model in (1.1). In the noiseless case we have $Z = 0$ deterministically, whereas in the noisy case we consider $Z \sim \text{Bernoulli}(\rho)$ for some $\rho \in (0, \frac{1}{2})$ not varying with p; i.e., each measurement is independently flipped with probability ρ.

We let S be uniform on $\mathcal{S}$, defined to contain the the $\binom{p}{k}$ subsets of $\{1,\dots,p\}$ of cardinality k. We consider Bernoulli measurements, where each entry of X is distributed as $\text{Bernoulli}\big(\frac{\nu}{k}\big)$ for some constant $\nu > 0$. The vector of n observations is denoted by $\mathbf{Y} \in \{0,1\}^n$, and the corresponding measurement matrix (each row of which contains a single measurement vector X) is denoted by $\mathbf{X} \in \{0,1\}^{n\times p}$. Given $\mathbf{X}$ and $\mathbf{Y}$, a *decoder* forms an estimate $\hat{S}$ of S. We consider two related performance measures. In the case of *exact* recovery, the error probability is given by

$$P_{\mathrm{e}} := \mathbb{P}[\hat{S} \neq S], \tag{1.2}$$

and is taken over the realizations of S, $\mathbf{X}$, and $\mathbf{Y}$ (the decoder is assumed to be deterministic). We assume that the decoder knows the system model, including $k := |S|$. This assumption is standard in the development of fundamental limits of the type considered in this paper [9, 12].

We also consider a less stringent performance criterion requiring that only $k - d_{\max}$ entries of S are successfully recovered, for some $d_{\max} \in \{1,\dots,k-1\}$. Following the study of an analogous criterion in compressive

[1] The quantity β in [20] corresponds to $1-\theta$ in our own notation.

sensing [17, 21], the error probability is given by

$$P_{\mathrm{e}}(d_{\max}) := \mathbb{P}\big[|S\backslash\hat{S}| > d_{\max} \cup |\hat{S}\backslash S| > d_{\max}\big]. \quad (1.3)$$

Notation We write $\mathbf{X}_S$ to denote the submatrix of $\mathbf{X}$ containing the columns indexed by S. The complement with respect to the set $\{1,\dots,p\}$ is denoted by $(\cdot)^c$. For a given joint distribution P_{XY}, the corresponding marginal distributions are denoted by P_X and P_Y, and similarly for conditional marginals (e.g., $P_{Y|X}$). We use usual notations for the entropy and mutual information (e.g. $H(X)$, $I(X;Y|Z)$). We define the binary entropy function in nats, $H_2(\rho) := -\rho\log\rho - (1-\rho)\log(1-\rho)$. We make use of the standard asymptotic notations $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$, $\Omega(\cdot)$ and $\omega(\cdot)$. We define the function $[\cdot]^+ = \max\{0,\cdot\}$, and write the floor function as $\lfloor\cdot\rfloor$. The function log has base e.

2 Main Results

2.1 Noiseless Case with Exact Recovery

Our main result for the noiseless case is as follows.

Theorem 1. *For the noiseless group testing problem with $k = \Theta(p^\theta)$ $(\theta \in (0,1))$ and an optimized measurement matrix parameter ν, there exists a decoder such that $P_{\mathrm{e}} \to 0$ as $p \to \infty$ provided that*

$$n \geq \inf_{\nu>0} \max\left\{\frac{\frac{\theta}{1-\theta}k\log\frac{p}{k}}{e^{-\nu}\nu}, \frac{k\log\frac{p}{k}}{H_2(e^{-\nu})}\right\}(1+\eta) \quad \text{(Achievability)} \quad (2.4)$$

for some $\eta > 0$. Conversely, we have $P_{\mathrm{e}} \to 1$ as $p \to \infty$ whenever

$$n \leq \frac{k\log\frac{p}{k}}{\log 2}(1-\eta) \qquad \text{(Converse)} \qquad (2.5)$$

for some $\eta > 0$.

Proof. See Section 3.4. □

By setting $\nu = \log 2$ in (2.4), it is readily verified that the condition coincides with (2.5) whenever $\theta \leq \frac{1}{3}$, and we thus have an exact threshold indicating a phase transition. The converse bound shown has been proved (using significantly different techniques) even for optimal adaptive measurements [16], and hence a key implication is that adaptivity provides no asymptotic gain over non-adaptive Bernoulli measurements when $k = O(p^{\frac{1}{3}})$.

Our upper bound improves on existing bounds in the literature, including those developed using Combinatorial Optimal Matching Pursuit (COMP) [10], Definite Defectives (DD) [11], and Almost-Separable Matrices (ASM) [14]; see Figure 1 for an illustration. For fairness, we note that the COMP and DD algorithms are computationally tractable and do not require knowledge of k. While the optimal threshold for $\theta > \frac{1}{3}$ remains unclear, it has been suggested that the DD curve in Figure 1 cannot be improved for $\theta > 0.5$ using Bernoulli measurements [11].

2.2 Noisy Case with Exact Recovery

We now consider the noisy group testing problem. We do not attempt to provide results with constants that are optimized to the same extent as the noiseless case, and we thus set $\nu = \log 2$, i.e., $P_X \sim \text{Bernoulli}\big(\frac{\log 2}{k}\big)$.

Theorem 2. *For the noisy group testing problem with $\rho \in (0, 0.5)$, $\nu = \log 2$, and $k = \Theta(p^\theta)$ $(\theta \in (0,1))$, we have $P_{\mathrm{e}} \to 0$ as $p \to \infty$ provided that*

$$n \geq \inf_{\delta_2 \in (0,1)} \max\left\{\zeta(\rho,\delta_2,\theta), \frac{1}{\log 2 - H_2(\rho)}\right\} \times \Big(k\log\frac{p}{k}\Big)(1+\eta) \quad \text{(Achievability)} \quad (2.6)$$

for some $\eta > 0$, where

$$\zeta(\rho,\delta_2,\theta) := \frac{2}{\log 2}\max\left\{\frac{2(1+\frac{1}{3}\delta_2(1-2\rho))\frac{\theta}{1-\theta}}{\delta_2^2(1-2\rho)^2}, \frac{\frac{1+2\theta}{1-\theta}}{(1-2\rho)\log\frac{1-\rho}{\rho}(1-\delta_2)}\right\}. \quad (2.7)$$

Conversely, we have $P_{\mathrm{e}} \to 1$ as $p \to \infty$ whenever

$$n \leq \frac{k\log\frac{p}{k}}{\log 2 - H_2(\rho)}(1-\eta) \qquad \text{(Converse)} \qquad (2.8)$$

for some $\eta > 0$.

Proof. The proof follows similar steps to Theorem 1; the differences are detailed in Appendix C. □

The second term in the maximum in (2.6) is dominant (thus matching (2.8)) for sufficiently small θ. To see this, we first note that the first term in the maximum in (2.7) tends to zero as $\theta \to 0$, and cannot be dominant in this limit. This implies that δ_2 may be arbitrarily close to zero when θ is sufficiently small. Assuming then that δ_2 and θ are small and the maximum in (2.7) is achieved by the second term, we have $\zeta(\rho,\delta_2,\theta) \approx \frac{2}{\log 2}\frac{1}{(1-2\rho)\log\frac{1-\rho}{\rho}}$. This is strictly smaller than $\frac{1}{\log 2 - H_2(\rho)}$; see Proposition 8 in Appendix C.

Once again, the converse is known to hold (at least in terms of the "weak" converse $P_{\mathrm{e}} \not\to 0$) even in the adaptive setting [10], and we have thus provided cases where non-adaptive Bernoulli measurements yield the same asymptotics as optimal adaptive measurements.

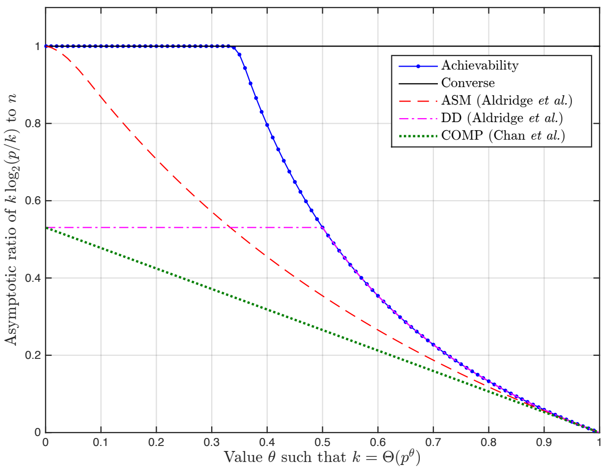

Figure 1: Asymptotic thresholds on the number of measurements required for noiseless group testing, with $k = \Theta(p^\theta)$ for some $\theta > 0$. The vertical axis represents the constant $c(\theta)$ such that the asymptotic number of measurements is $\frac{1}{c(\theta)} k \log_2 \frac{p}{k}$.

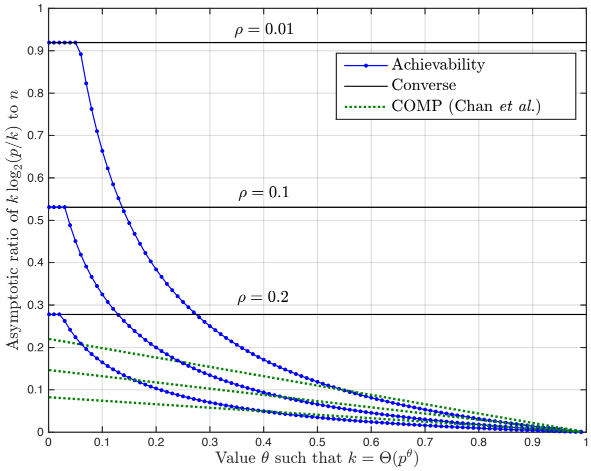

Figure 2: Asymptotic thresholds on the number of measurements required for noisy group testing, with $k = \Theta(p^\theta)$ for some $\theta > 0$. The vertical axis represents the constant $c(\theta)$ such that the asymptotic number of measurements is $\frac{1}{c(\theta)} k \log_2 \frac{p}{k}$.

Our upper bound improves on that of the COMP algorithm [10], and appears to be the first to provide a phase transition for small θ, or even in the limit as $\theta \to 0$. See Figure 2 for an illustration.

2.3 Partial Recovery

Next, we present our main result regarding the partial recovery criterion in (1.3).

Theorem 3. *For the group testing problem with $\rho \in [0, 0.5)$ (i.e., possibly noiseless), $\nu = \log 2$, $k \to \infty$, $k = o(p)$, and $d_{\max} = \lfloor \alpha^* k \rfloor$ for some $\alpha^* \in (0,1)$, we have $P_{\rm e}(d_{\max}) \to 0$ as $p \to \infty$ provided that*

$$n \geq \frac{k \log \frac{p}{k}}{\log 2 - H_2(\rho)}(1+\eta) \qquad \text{(Achievability)} \qquad (2.9)$$

for some $\eta > 0$. Conversely, $P_{\rm e}(d_{\max}) \to 1$ as $p \to \infty$ whenever

$$n \leq \frac{(1-\alpha^*)\big(k \log \frac{p}{k}\big)}{\log 2 - H_2(\rho)}(1-\eta) \qquad \text{(Converse)} \qquad (2.10)$$

for some $\eta > 0$.

Proof. The proof follows similar steps to Theorems 1–2, but is much simpler. See Appendix D for details. □

Theorem 3 shows that at least for sufficiently small θ (e.g., $k = O\big(p^{\frac{1}{3}}\big)$ in the noiseless case), there is not much to be saved by moving from exact recovery to partial recovery: Allowing for a fraction α^* of errors leads to at most a reduction in the number of measurements of a multiplicative factor $1 - \alpha^*$.

It may be tempting to take $\alpha^* \to 0$ in (2.9)–(2.10) to infer an exact threshold for all $\theta \in (0,1)$ with exact recovery. However, such a limit would be of the form $\lim_{\alpha^* \to 0} \lim_{p \to \infty}$, whereas a valid result for the exact recovery requires the opposite order $\lim_{p\to\infty} \lim_{\alpha^* \to 0}$.

3 Proofs

While our focus in this section is primarily on the noiseless case, we begin with some general definitions and non-asymptotic bounds on $P_{\rm e}$ that apply to both the noiseless and noisy cases. To this end, we let $\mathbf{X}$ be i.i.d. on some general distribution P_X, and let the observation vector $\mathbf{Y}$ be generated from the measurement matrix $\mathbf{X}$ and defective set S according to the n-fold product of some general distribution $P_{Y|X_S}$. All that we assume of this distribution is that it is the same for any realization of S, and that it is unchanged when the corresponding columns of X_S are permuted. Thus, our initial bounds will also apply to other noise models.

It will prove convenient to work with random variables that are implicitly conditioned on a fixed value of S, say $s = \{1, \dots, k\}$. We write $P_{Y|X_s}$ in place of $P_{Y|X_S}$ to emphasize that $S = s$, and we define

$$P_{X_sY}(x_s, y) := P_X^k(x_s) P_{Y|X_s}(y|x_s) \qquad (3.11)$$

$$P_{\mathbf{X}_s\mathbf{Y}}(\mathbf{x}_s, \mathbf{y}) := P_X^{n\times k}(\mathbf{x}_s) P_{Y|X_s}^n(\mathbf{y}|\mathbf{x}_s), \qquad (3.12)$$

where $P^n_{Y|X_s}(\cdot|\cdot)$ is the n-fold product of $P_{Y|X_s}(\cdot|\cdot)$.

If it is not stated otherwise, the random variables (X_s, Y) and $(\mathbf{X}_s, \mathbf{Y})$ are distributed as

$$(X_s, Y) \sim P_{X_sY} \qquad (3.13)$$

$$(\mathbf{X}_s, \mathbf{Y}) \sim P_{\mathbf{X}_s\mathbf{Y}}, \qquad (3.14)$$

with the remaining entries of the measurement matrix being distributed as $\mathbf{X}_{s^c} \sim P_X^{n\times(p-k)}$. That is, we condition on a fixed $S = s$ except where stated otherwise.

3.1 Preliminary Definitions As in [12,22], we consider partitions of the defective set $s \in \mathcal{S}$ into two sets $s_{\mathrm{dif}} \neq \emptyset$ and s_{eq}. One can think of s_{eq} as corresponding to an overlap $s \cap \bar{s}$ between the true set s and some incorrect set $\bar{s}$, with s_{dif} corresponding to the indices $s \backslash \bar{s}$ in one set but not the other. There are $2^k - 1$ ways of performing such a partition (the subtraction of one being due to the condition that s_{dif} is non-empty).

For fixed $s \in \mathcal{S}$ and a corresponding pair $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$, we introduce the notation

$$P_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{dif}}}, x_{s_{\mathrm{eq}}}) := P_{Y|X_s}(y|x_s), \qquad (3.15)$$

where $P_{Y|X_s}$ is the marginal distribution of (3.12). While the left-hand side of (3.15) represents the same quantity for any such $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$, it will still prove convenient to work with this form in place of the right-hand side. In particular, this allows us to introduce the marginal distribution

$$P_{Y|X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{eq}}}) := \sum_{x_{s_{\mathrm{dif}}}} P_X^{\ell}(x_{s_{\mathrm{dif}}}) P_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{dif}}}, x_{s_{\mathrm{eq}}}), \qquad (3.16)$$

where $\ell := |s_{\mathrm{dif}}|$. Using the preceding definitions, we introduce the *information density* [23]

$$\imath^n(\mathbf{x}_{s_{\mathrm{dif}}}; \mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}}) := \sum_{i=1}^n \imath(x_{s_{\mathrm{dif}}}^{(i)}; y^{(i)}|x_{s_{\mathrm{eq}}}^{(i)}) \qquad (3.17)$$

$$\imath(x_{s_{\mathrm{dif}}}; y|x_{s_{\mathrm{eq}}}) := \log \frac{P_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{dif}}}, x_{s_{\mathrm{eq}}})}{P_{Y|X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{eq}}})} \qquad (3.18)$$

where $(\cdot)^{(i)}$ denotes the i-th entry (respectively, row) of a vector (respectively, matrix). Averaging (3.18) with respect to (X_s, Y) in (3.13) yields a conditional mutual information, which we denote by

$$I(\ell) := I(X_{s_{\mathrm{dif}}}; Y|X_{s_{\mathrm{eq}}}), \qquad (3.19)$$

where $\ell := |s_{\mathrm{dif}}|$ (by symmetry, the mutual information for each $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ depends only on this quantity). As in [12], these mutual informations will play a key role in our analysis.

We are not aware of previous statistical works using techniques based on bounding information densities; however, these have proved to be highly powerful in communication problems, and are the basis of information-spectrum methods [24]. Roughly speaking, the information densities can be used to characterize how likely it is for $(\mathbf{Y}|\mathbf{X}_s)$ to appear as if it was generated conditioned on $\mathbf{X}_{s_{\mathrm{eq}}}$ alone, or vice versa.

3.2 Non-Asymptotic Bounds on the Error Probability Our initial non-asymptotic bounds build on threshold-based techniques from the channel coding literature [24–27], but with suitable modifications leading to the partitions of s into $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$.

Theorem 4. *For any $\delta_1 > 0$, there exists a decoder such that*

$$P_{\mathrm{e}} \leq \mathbb{P}\Bigg[\bigcup_{(s_{\mathrm{dif}}, s_{\mathrm{eq}})\,:\, s_{\mathrm{dif}} \neq \emptyset} \bigg\{\imath^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \leq \log\binom{p-k}{|s_{\mathrm{dif}}|} + \log\bigg(\frac{k}{\delta_1}\binom{k}{|s_{\mathrm{dif}}|}\bigg)\bigg\}\Bigg] + \delta_1. \qquad (3.20)$$

Proof. We fix the constants $\gamma_1, \dots, \gamma_k$ in $\mathbb{R}$ arbitrarily, and consider a decoder that searches for the unique set $s \in \mathcal{S}$ such that

$$\imath^n(\mathbf{x}_{s_{\mathrm{dif}}}; \mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}}) > \gamma_{|s_{\mathrm{dif}}|} \qquad (3.21)$$

for all $2^k - 1$ partitions $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ of s with $s_{\mathrm{dif}} \neq \emptyset$. An error occurs if no such s exists, if multiple exist, or if such a set differs from the true value.

Since the joint distribution of $(\mathbf{X}_s, \mathbf{Y}_s \,|\, S = s)$ is the same for all s in our setup, and the decoder that we have chosen exhibits a similar symmetry, we can condition on a fixed and arbitrary value of S, say $s = \{1, \dots, k\}$. By the union bound, the error probability is upper bounded by

$$P_{\mathrm{e}} \leq \mathbb{P}\Bigg[\bigcup_{(s_{\mathrm{dif}}, s_{\mathrm{eq}})} \Big\{\imath^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \leq \gamma_{|s_{\mathrm{dif}}|}\Big\}\Bigg] + \sum_{\bar{s} \in \mathcal{S}\backslash\{s\}} \mathbb{P}\Big[\imath^n(\mathbf{X}_{\bar{s}\backslash s}; \mathbf{Y}|\mathbf{X}_{\bar{s}\cap s}) > \gamma_{|s_{\mathrm{dif}}|}\Big], \qquad (3.22)$$

where here and subsequently we let the condition $s_{\mathrm{dif}} \neq \emptyset$ remain implicit. The first term corresponds to the true set failing the threshold test, and the second term corresponds to some incorrect set $\bar{s}$ passing the threshold test. In the summand of the second term, we have upper bounded the probability of an intersection of $2^k - 1$ events by just one such event, namely, the one corresponding to $s_{\mathrm{dif}} = \bar{s} \backslash s$ and $s_{\mathrm{eq}} = s \cap \bar{s}$.

Using the shorthand $\ell := |\bar{s} \backslash s|$, we can weaken the second probability in (3.22) as follows:

$$\begin{aligned}
&\mathbb{P}\big[\imath^n(\mathbf{X}_{\bar{s}\setminus s};\mathbf{Y}|\mathbf{X}_{\bar{s}\cap s}) > \gamma_\ell\big] \\
&\quad= \sum_{\mathbf{x}_{\bar{s}\cap s},\mathbf{x}_{\bar{s}\setminus s},\mathbf{y}} P_X^{n\times(k-\ell)}(\mathbf{x}_{\bar{s}\cap s}) P^n_{Y|X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\cap s}) \\
&\qquad \times P_X^{n\times\ell}(\mathbf{x}_{\bar{s}\setminus s})\mathbb{1}\bigg\{\log\frac{P^n_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\setminus s},\mathbf{x}_{\bar{s}\cap s})}{P^n_{Y|X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\cap s})} > \gamma_\ell\bigg\} \qquad (3.23) \\
&\quad\le \sum_{\mathbf{x}_{\bar{s}\cap s},\mathbf{x}_{\bar{s}\setminus s},\mathbf{y}} P_X^{n\times(k-\ell)}(\mathbf{x}_{\bar{s}\cap s}) P_X^{n\times\ell}(\mathbf{x}_{\bar{s}\setminus s}) \\
&\qquad\qquad \times P^n_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\setminus s},\mathbf{x}_{\bar{s}\cap s})e^{-\gamma_\ell} \qquad (3.24) \\
&\quad= e^{-\gamma_\ell}, \qquad (3.25)
\end{aligned}$$

where in (3.23) we used the fact that the output vector depends only on the columns of $\mathbf{x}_{\bar{s}}$ corresponding to entries of $\bar{s}$ that are also in s, and (3.24) follows by bounding $P_{\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}}$ using the event within the indicator function, and then upper bounding the indicator function by one. Substituting (3.25) into (3.22) gives

$$\begin{aligned}
P_{\mathrm{e}} \le \mathbb{P}\bigg[\bigcup_{(s_{\mathrm{dif}},s_{\mathrm{eq}})}\Big\{\imath^n(\mathbf{X}_{s_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \le \gamma_\ell\Big\}\bigg] \\
+\sum_{\ell=1}^{k}\binom{p-k}{\ell}\binom{k}{\ell}e^{-\gamma_\ell}, \qquad (3.26)
\end{aligned}$$

where the combinatorial terms arise from a standard counting argument [28]. Finally, the choice $\gamma_\ell = \log\big(\frac{k}{\delta_1}\binom{p-k}{\ell}\binom{k}{\ell}\big)$ makes the second term in (3.26) be upper bounded by δ_1, thus completing the proof. □

Theorem 4 bears some resemblance to a bound of Malyutov [9]; the latter can be obtained by applying the union bound and Chebyshev's inequality to (3.20). However, the key to obtaining Theorem 4 is using a more powerful concentration inequality; Chebyshev's inequality appears to be insufficient when $k = \Theta(p^\theta)$.

Theorem 5. *Fix $\delta_1 > 0$, and let $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ be an arbitrary partition of $s = \{1,\dots,k\}$ with $s_{\mathrm{dif}} \ne \emptyset$. For any decoder, we have*

$$\begin{aligned}
P_{\mathrm{e}} \ge \mathbb{P}\bigg[&\imath^n(\mathbf{X}_{s_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \\
&\le \log\binom{p-k+|s_{\mathrm{dif}}|}{|s_{\mathrm{dif}}|} + \log\delta_1\bigg] - \delta_1. \qquad (3.27)
\end{aligned}$$

Proof. As in [12], we consider an argument based on a genie. Letting ℓ denote the cardinality of s_{dif} in the theorem statement, the genie-aided setup is described as follows:

1. Generate S_{eq} uniformly on $\mathcal{S}_{\mathrm{eq}}(\ell)$, defined to contain the $\binom{p}{k-\ell}$ subsets of $\{1,\dots,p\}$ having cardinality $k-\ell$.
2. Generate S_{dif} uniformly on $\mathcal{S}_{\mathrm{dif}}(S_{\mathrm{eq}})$, defined to contain the $\binom{p-k+\ell}{\ell}$ subsets of $\{1,\dots,p\}\backslash S_{\mathrm{eq}}$ having cardinality ℓ.
3. Set $S = S_{\mathrm{dif}} \cup S_{\mathrm{eq}}$. The measurement matrix $\mathbf{X}$ is i.i.d. on P_X, and the observation vector $\mathbf{Y}$ is generated from S and $\mathbf{X}$ conditionally independently according to $P_{Y|X_S}$, as in the original setup.
4. Reveal the indices S_{eq} to the decoder (along with $\mathbf{X}$ and $\mathbf{Y}$). The decoder forms an estimate $\hat{S}_{\mathrm{dif}}$ of S_{dif}, and an error occurs if $\hat{S}_{\mathrm{dif}} \ne S_{\mathrm{dif}}$.

Clearly the distribution of S in this setup is uniform on $\mathcal{S}$, and hence the only difference compared to the original setup is that the decoder has additional information. It follows that any converse for this setup implies the same converse for the original setup.

Throughout the proof, we make use of the random variables defined in the preceding steps, departing from the notation implicitly conditioned on S equaling a fixed value s (see (3.14)) until the final step in obtaining (3.27).

We first study the error probability for the genie-aided setting conditioned on $S_{\mathrm{eq}} = s_{\mathrm{eq}}$, which we denote by $P_{\mathrm{e}}(s_{\mathrm{eq}})$. By the simple identity $\mathbb{P}[\mathcal{A}] = \mathbb{P}[\mathcal{A}\cap\mathcal{E}] + \mathbb{P}[\mathcal{A}\cap\mathcal{E}^c]$, we have for any event $\mathcal{A}(s_{\mathrm{eq}})$ that

$$P_{\mathrm{e}}(s_{\mathrm{eq}}) \ge \mathbb{P}[\mathcal{A}(s_{\mathrm{eq}})] - \mathbb{P}[\mathcal{A}(s_{\mathrm{eq}}) \cap \text{no error}]. \qquad (3.28)$$

We fix a constant $\gamma \in \mathbb{R}$ (different in general from $\gamma_1,\dots,\gamma_k$ above) and choose

$$\mathcal{A}(s_{\mathrm{eq}}) = \big\{\imath^n(\mathbf{X}_{S_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \le \gamma\big\}. \qquad (3.29)$$

Using the definitions in (3.17)–(3.18), and defining $\mathcal{D}(s_{\mathrm{dif}}|s_{\mathrm{eq}})$ to be the set of pairs $(\mathbf{x},\mathbf{y})$ such that the decoder outputs s_{dif} given s_{eq}, we obtain

$$\begin{aligned}
&\mathbb{P}[\mathcal{A}(s_{\mathrm{eq}}) \cap \text{no error}] \\
&\quad= \sum_{s_{\mathrm{dif}}\in\mathcal{S}_{\mathrm{dif}}(s_{\mathrm{eq}})}\frac{1}{\binom{p-k+\ell}{\ell}}\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{D}(s_{\mathrm{dif}}|s_{\mathrm{eq}})} P_X^{n\times p}(\mathbf{x}) \\
&\qquad\times P^n_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{dif}}},\mathbf{x}_{s_{\mathrm{eq}}}) \\
&\qquad\times \mathbb{1}\bigg\{\log\frac{P^n_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{dif}}},\mathbf{x}_{s_{\mathrm{eq}}})}{P^n_{Y|X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}})} \le \gamma\bigg\} \qquad (3.30) \\
&\quad\le \frac{1}{\binom{p-k+\ell}{\ell}}\sum_{s_{\mathrm{dif}}\in\mathcal{S}_{\mathrm{dif}}(s_{\mathrm{eq}})}\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{D}(s_{\mathrm{dif}}|s_{\mathrm{eq}})} P_X^{n\times p}(\mathbf{x}) \\
&\qquad\times P^n_{Y|X_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}})e^{\gamma} \qquad (3.31) \\
&\quad= \frac{e^{\gamma}}{\binom{p-k+\ell}{\ell}}, \qquad (3.32)
\end{aligned}$$

where (3.30) follows since an error occurs if and only if $(\mathbf{x},\mathbf{y}) \notin \mathcal{D}(s_{\mathrm{dif}}|s_{\mathrm{eq}})$, (3.31) follows by upper bounding $P^n_{Y|X_{s_{\mathrm{eq}}}}$ using the event in the indicator function, and (3.32) follows since the sets $\mathcal{D}(s_{\mathrm{dif}}|s_{\mathrm{eq}})$ are disjoint, and their union (over s_{dif}) is the entire space of $(\mathbf{x},\mathbf{y})$ pairs.

Averaging (3.28) over S_{eq} and applying (3.32), we obtain

$$P_{\mathrm{e}} \ge \sum_{s_{\mathrm{eq}} \in S_{\mathrm{eq}}(\ell)} \sum_{s_{\mathrm{dif}} \in S_{\mathrm{dif}}(s_{\mathrm{eq}})} \frac{1}{\binom{p}{k-\ell}} \frac{1}{\binom{p-k+\ell}{\ell}} \times \Bigg(\mathbb{P}\Big[\imath^n(\mathbf{X}_{s_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \le \gamma_\ell \,\Big|\, S_{\mathrm{dif}} = s_{\mathrm{dif}}, S_{\mathrm{eq}} = s_{\mathrm{eq}}\Big] - \frac{e^{\gamma}}{\binom{p-k+\ell}{\ell}} \Bigg). \quad (3.33)$$

Finally, we claim that this bound recovers (3.27) (which is written in terms of the joint distribution in (3.14) with a fixed $S = s$) upon setting $\gamma_\ell = \log\binom{p-k+\ell}{\ell} + \log\delta_1$. This immediately follows from the fact that all of the terms in the summations over s_{dif} and s_{eq} in (3.33) are equal, due to the symmetry of $P_{Y|X_S}$ with respect to S assumed in our setup (as well as the fact that $\mathbf{X}$ is i.i.d. and hence exhibits a similar symmetry). □

While we will use $(s_{\mathrm{dif}}, s_{\mathrm{eq}}) = (s, \emptyset)$ in Theorem 5 when obtaining (2.5) and (2.8), we have presented the more general form since (i) it provides a natural counterpart to Theorem 4, (ii) it is useful for comparison with [12], and (iii) the more general form is crucial in the extension to other support recovery problems [19].

3.3 Procedure for Applying Theorems 4 and 5

The bounds presented in the preceding theorems do not directly reveal the number of measurements required to achieving a vanishing error probability. In this subsection, we present the steps that can be used to obtain such conditions. The idea is to use a concentration inequality to bound the first term in (3.20) (or (3.27)), which is possible due to the fact that each $\imath^n$ is an i.i.d. summation. We provide the details of these steps separately for the achievability and converse (i.e., the upper and lower bound). We start with the former.

1. Observe that the mean of $\imath^n(\mathbf{X}_{s_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}})$ is $nI(|s_{\mathrm{dif}}|)$, where $I(\ell)$ is defined in (3.19).

2. Fix the constants $\delta_{2,1},\dotsc,\delta_{2,k}$ with $\delta_{2,\ell} \in (0,1)$, and suppose that we have for all ℓ that

$$\log\binom{p-k}{\ell} + \log\left(\frac{k}{\delta_1}\binom{k}{\ell}\right) \le n(1-\delta_{2,\ell})I(\ell), \quad (3.34)$$

$$\mathbb{P}\big[\imath^n(\mathbf{X}_{s_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \le n(1-\delta_{2,\ell})I(\ell)\big] \le \psi_\ell(n,\delta_{2,\ell}) \quad (3.35)$$

for some functions $\{\psi_\ell\}_{\ell=1}^k$, and any $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ with $|s_{\mathrm{dif}}| = \ell$. Combining these conditions with the union bound, we obtain

$$\mathbb{P}\Bigg[\bigcup_{(s_{\mathrm{dif}},s_{\mathrm{eq}})\,:\,s_{\mathrm{dif}}\neq\emptyset} \bigg\{\imath^n(\mathbf{X}_{s_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \le \log\binom{p-k}{|s_{\mathrm{dif}}|} + \log\left(\frac{k}{\delta_1}\binom{k}{|s_{\mathrm{dif}}|}\right)\bigg\}\Bigg] \le \sum_{\ell=1}^k \binom{k}{\ell}\psi_\ell(n,\delta_{2,\ell}). \quad (3.36)$$

3. Observe that the condition in (3.34) can be written as

$$n \ge \frac{\log\binom{p-k}{\ell} + \log\left(\frac{k}{\delta_1}\binom{k}{\ell}\right)}{I(\ell)(1-\delta_{2,\ell})}. \quad (3.37)$$

We summarize the preceding findings in the following theorem.

Theorem 6. *For any constants $\delta_1 > 0$ and $\{\delta_{2,\ell}\}_{\ell=1}^k$ ($\delta_{2,\ell} \in (0,1)$), and functions $\{\psi_\ell\}_{\ell=1}^k$ ($\psi_\ell : \mathbb{Z}\times\mathbb{R} \to \mathbb{R}$) such that* (3.35) *and* (3.37) *hold for all $\ell = 1,\dotsc,k$, we have*

$$P_{\mathrm{e}} \le \sum_{\ell=1}^k \binom{k}{\ell}\psi_\ell(n,\delta_{2,\ell}) + \delta_1. \quad (3.38)$$

With this result, it only remains to use a concentration inequality to characterize $\{\psi_\ell\}$, and then choose n, δ_1 and $\delta_{2,\ell}$ (as functions of p) such that (3.37) holds and the right-hand side of (3.38) vanishes.

The application of Theorem 5 is done using similar steps, so we provide slightly less detail. Fix $\delta_2 > 0$, and suppose that the pair $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ and value $\ell := |s_{\mathrm{dif}}|$ are such that

$$\log\binom{p-k+\ell}{\ell} - \log\delta_1 \ge n(1+\delta_2)I(\ell), \quad (3.39)$$

$$\mathbb{P}\big[\imath^n(\mathbf{X}_{s_{\mathrm{dif}}};\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}) \le n(1+\delta_2)I(\ell)\big] \ge 1 - \psi'(n,\delta_2) \quad (3.40)$$

for some function ψ'. Combining these, the first probability in (3.27) is lower bounded by $1-\psi'(n,\delta_2)$.

Next, we observe that (3.39) holds if and only if

$$n \le \frac{\log\binom{p-k+\ell}{\ell} - \log\delta_1}{I(\ell)(1+\delta_2)}. \quad (3.41)$$

Since the partition $(s_{\text{dif}}, s_{\text{eq}})$ is arbitrary, we can choose the pair that maximizes the right-hand side.

We summarize the preceding observations in the following.

Theorem 7. *For any constants $\delta_1 > 0$ and $\delta_2 > 0$, partition $(s_{\text{dif}}, s_{\text{eq}})$ ($s_{\text{dif}} \neq \emptyset$), and function $\psi' : \mathbb{Z}\times\mathbb{R} \to \mathbb{R}$ such that* (3.40) *and* (3.41) *hold, we have*

$$P_{\text{e}} \geq 1 - \psi'(n, \delta_2) - \delta_1. \tag{3.42}$$

3.4 Proof of Theorem 1

We now use Theorems 6 and 7 to obtain (2.4) and (2.5) respectively. We focus primarily on the former, which is considerably more difficult.

Step 1: Auxiliary Results and Concentration Inequalities We begin with the following proposition characterizing the mutual information.

Proposition 1. *For the noiseless group testing problem, consider arbitrary sequences of sparsity levels $k \to \infty$ and $\ell \in \{1, \dots, k\}$ (both indexed by p). If $\frac{\ell}{k} = o(1)$, then*

$$I(\ell) = \left(e^{-\nu}\nu\frac{\ell}{k}\log\frac{k}{\ell}\right)(1 + o(1)). \tag{3.43}$$

Moreover, if $\frac{\ell}{k} \to \alpha \in (0, 1]$, then

$$I(\ell) = e^{-(1-\alpha)\nu}H_2\big(e^{-\alpha\nu}\big)(1 + o(1)). \tag{3.44}$$

Proof. The proof uses standard asymptotic expansions, and is given in Appendix B. □

Fix $\delta_2^{(1)} \in (0,1)$ and $\delta_2^{(2)} \in (0,1)$ and set $\delta_{2,\ell} = \delta_2^{(1)}$ for $\ell \leq \lfloor\frac{k}{\log k}\rfloor$, and $\delta_{2,\ell} = \delta_2^{(2)}$ for $\ell > \lfloor\frac{k}{\log k}\rfloor$. In Appendix A, we give two concentration inequalities showing that we may fix $\epsilon > 0$ and set

$$\psi_\ell(n, \delta_{2,\ell}) = \begin{cases} \exp\left(-n\frac{\ell}{k}e^{-\nu}\nu\Big((1-\delta_2^{(1)})\log(1-\delta_2^{(1)}) + \delta_2^{(1)}\Big)(1-\epsilon)\right) & \ell \leq \lfloor\frac{k}{\log k}\rfloor \\ 2\exp\left(-\frac{(\delta_2^{(2)}I(\ell))^2 n}{4(8+\delta_2^{(2)}I(\ell))}\right) & \ell > \lfloor\frac{k}{\log k}\rfloor \end{cases} \tag{3.45}$$

for sufficiently large p (depending on ϵ, $\delta_2^{(1)}$ and $\delta_2^{(2)}$). For the converse, we only use one of the two concentration inequalities, setting $\psi'(n, \delta_{2,\ell}) = 2\exp\big(-\frac{(\delta_{2,\ell}I(\ell))^2 n}{4(8+\delta_{2,\ell}I(\ell))}\big)$.

Step 2: Control the Remainder Terms The next step is to find conditions on n and the free parameters (e.g., ϵ) such that the first term in (3.38) and the second term in (3.42) vanish. For the latter, we set $(s_{\text{dif}}, s_{\text{eq}}) = (s, \emptyset)$ in Theorem 7. From the above choice of ψ' and the growth of $I(k)$ in (3.44), we immediately obtain that $\psi'(n, \delta_{2,k}) \to 0$ whenever $n \to \infty$. The term (3.38) requires more effort; we summarize the findings in the following proposition.

Proposition 2. *Let $k = \Theta(p^\theta)$ for some $\theta \in (0, 1)$.*

(i) For any $\eta > 0$, there exists $\delta_2^{(1)} \in (0, 1)$ and a choice of $\epsilon > 0$ in (3.45) *such that $\sum_{\ell=1}^{\lfloor\frac{k}{\log k}\rfloor}\binom{k}{\ell}\psi_\ell(n, \delta_2^{(1)}) \to 0$ provided that*

$$n \geq \frac{\frac{\theta}{1-\theta}k\log\frac{p}{k}}{e^{-\nu}\nu}(1 + \eta). \tag{3.46}$$

(ii) For any $\delta_2^{(2)} \in (0, 1)$, we have $\sum_{\lfloor\frac{k}{\log k}\rfloor+1}^{k}\binom{k}{\ell}\psi_\ell(n, \delta_2^{(2)}) \to 0$ provided that $n = \Omega\big(k\log\frac{p}{k}\big)$.

Proof. These claims follow using (3.45) and simple algebraic manipulations. See Appendix B for details. □

The idea here is that for the smaller values of ℓ, it is the concentration inequality that dominates the final bound, so we let $\delta_{2,\ell} = \delta_2^{(1)}$ be closer to one to provide better concentration behavior. For large values of ℓ, the opposite is true, so we let $\delta_{2,\ell} = \delta_2^{(2)}$ be close to zero.

Step 3: Combine and Simplify We are now in a position to prove Theorem 1. We immediately obtain the first term in the maximum in (2.4) from (3.46), so it remains to derive the second term. We start with (3.37); by taking $\delta_1 \to 0$ sufficiently slowly (so that the second term in (3.38) vanishes), we obtain the condition

$$n \geq \max_{\ell=1,\dots,k}\frac{\log\binom{p-k}{\ell} + 2\log\big(k\binom{k}{\ell}\big)}{I(\ell)(1-\delta_{2,\ell})}\big(1 + o(1)\big). \tag{3.47}$$

Using (3.43)–(3.44) and the identity $\log\binom{p-k}{\ell} = \Theta\big(\ell\log\frac{p}{\ell}\big)$ we see that the objective in (3.47) has growth rate

$$\Theta\left(\frac{k\log\frac{p}{\ell}}{1 + \log\frac{k}{\ell}}\right) \tag{3.48}$$

whenever the constants $\{\delta_{2,\ell}\}$ are bounded away from one. This behaves as $\Theta\big(k\log\frac{p}{k}\big)$ when $\frac{\ell}{k} = \Theta(1)$, and as $\Theta\big(\frac{k\log\frac{p}{k}}{\log\frac{k}{\ell}} + k\big)$ when $\frac{\ell}{k} = o(1)$ (the latter of these is seen by writing $\log\frac{p}{\ell} = \log\frac{p}{k} + \log\frac{k}{\ell}$). Thus, the maximum in (3.47) can only be achieved by a sequence such that

$\frac{\ell}{k} = \Theta(1)$. Moreover, with $\frac{\ell}{k} = \Theta(1)$, we see from the assumption $k = o(p)$ that the term $2 \log\big(k\binom{k}{\ell}\big) = O(k)$ is dominated by $\log\binom{p-k}{\ell} = \Theta\big(k \log \frac{p}{k}\big)$, and can thus be factored into the $o(1)$ remainder term in (3.47). This yields the condition

$$n \ge \max_{\ell=1,\dots,k} \frac{\ell \log \frac{p}{\ell}}{I(\ell)(1-\delta_{2,\ell})}\big(1+o(1)\big). \tag{3.49}$$

Since the maximum can only be achieved asymptotically with $\frac{\ell}{k} = \Theta(1)$, we proceed by considering $\frac{\ell}{k} \to \alpha$ for some arbitrary $\alpha \in (0,1]$. Under this scaling, $\ell \log \frac{p}{\ell}$ behaves as $\big(\alpha k \log \frac{p}{k}\big)(1+o(1))$. Moreover, according to Proposition 2, we can choose $\delta_{2,\ell}$ to be arbitrarily small for all ℓ values except those below $\lfloor \frac{k}{\log k} \rfloor$. Such values behave as $o(k)$, and thus do achieve the maximum in (3.49). Combining these observations with (3.44), the right-hand side of (3.49) yields the condition

$$n \ge \max_{\alpha \in (0,1]} \frac{\alpha k \log \frac{p}{k}}{e^{-(1-\alpha)\nu} H_2\big(e^{-\alpha\nu}\big)}\big(1+\eta\big), \tag{3.50}$$

where η may be arbitrarily small. By a change of variable $\lambda = e^{-\alpha\nu}$, the coefficient to $k \log \frac{p}{k}$ can be written as $\frac{1}{\nu} e^{\nu} \frac{\lambda \log \frac{1}{\lambda}}{H_2(\lambda)}$. This is easily verified to be decreasing in $\lambda \in [0,1]$, which implies that the maximizing value of α is one, and yields the second term in (2.4).

The proof of the converse is similar but considerably simpler; setting $(s_{\mathrm{dif}}, s_{\mathrm{eq}}) = (s, \emptyset)$ in Theorem 6, we obtain $\alpha = 1$ immediately. The denominator $\log 2$ in (2.5) is obtained by maximizing $H_2(e^{-\nu})$ over ν, and the condition $n \to \infty$ stated before Proposition 2 is satisfied when (2.5) holds with equality. Such equality can be assumed without loss of generality, since the decoder may always choose to ignore measurements.

4 Conclusion

We have provided new techniques for studying limits on the required number of measurements for group testing, building on thresholding methods from the information-theoretic channel coding literature. In the noiseless case, we have provided an exact asymptotic threshold (phase transition) on the number of measurements for $k = O(p^{\frac{1}{3}})$, matching the corresponding threshold for adaptive measurements. In the noisy case, we have obtained similar thresholds holding for sufficiently small θ. Moreover, we have provided a new approach to developing strong converse results, stating that $P_{\mathrm{e}} \to 1$. An important challenge for future work is to devise *practical* recovery algorithms yielding the phase transitions developed in this paper.

Appendices

A Concentration Inequalities

Throughout this section, we make use of Bernstein's inequality, which is given as follows [29, Sec. 2.8].

Lemma 1. *Let $W_1, \dots, W_n$ be independent real-valued random variables such that*

$$\sum_{i=1}^{n} \mathbb{E}[W_i^2] \le \tau \tag{A.1}$$

$$\sum_{i=1}^{n} \mathbb{E}[|W_i|^q] \le \frac{q!}{2} \tau c^{q-2} \quad (q \ge 3) \tag{A.2}$$

for some $\tau, c > 0$. Then

$$\mathbb{P}\bigg[\sum_{i=1}^{n} \big(W_i - \mathbb{E}[W_i]\big) \ge t\bigg] \le \exp\bigg(\frac{t^2}{2(\tau + ct)}\bigg) \tag{A.3}$$

for all $t > 0$.

We proceed by presenting a concentration inequality that applies to both the noisy and noiseless cases, and another that is specific to the noiseless case.

Proposition 3. *For the noiseless and noisy group testing problems, the following holds for all $\ell = 1, \dots, k$ and $\delta > 0$:*

$$\mathbb{P}\Big[\big|\imath^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\mathrm{eq}}}) - nI(\ell)\big| \ge n\delta\Big] \le 2\exp\bigg(-\frac{\delta^2 n}{4(8+\delta)}\bigg), \tag{A.4}$$

where $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ is an arbitrary partition of s with $|s_{\mathrm{dif}}| = \ell$.

Proof. To bound the moments in Bernstein's inequality, we follow the arguments of [24, Rmk. 3.1.1] and [30, App. D]. Recall the definition of the information density in (3.18). For any $q \ge 2$, we have from Minkowski's inequality that

$$\mathbb{E}\big[|\imath(X_{s_{\mathrm{dif}}}; Y | X_{s_{\mathrm{eq}}})|^q\big]^{1/q} \le \mathbb{E}\bigg[\bigg(\log \frac{1}{P_{Y|X_{s_{\mathrm{dif}}} X_{s_{\mathrm{eq}}}}(Y | X_{s_{\mathrm{dif}}}, X_{s_{\mathrm{eq}}})}\bigg)^q\bigg]^{1/q} + \mathbb{E}\bigg[\bigg(\log \frac{1}{P_{Y|X_{s_{\mathrm{eq}}}}(Y|X_{s_{\mathrm{eq}}})}\bigg)^q\bigg]^{1/q}. \tag{A.5}$$

For any given $(x_{s_{\mathrm{dif}}}, x_{s_{\mathrm{eq}}})$, the remaining averaging over

Y in the first term has the form

$$\sum_y P_{Y|X_{s_{\rm dif}}X_{s_{\rm eq}}}(y|x_{s_{\rm dif}},x_{s_{\rm eq}}) \times \bigg(\log\frac{1}{P_{Y|X_{s_{\rm dif}}X_{s_{\rm eq}}}(y|x_{s_{\rm dif}},x_{s_{\rm eq}})}\bigg)^q, \quad \text{(A.6)}$$

and is thus upper bounded by $2\big(\frac{q}{e}\big)^{1/q}$, since the observations are binary and the function $f(z) = z\log^q\frac{1}{z}$ has a maximum value of $\big(\frac{q}{e}\big)^{1/q}$ for $z \in [0,1]$. By handling the second term in (A.5) similarly, we obtain

$$\mathbb{E}\big[|\imath(X_{s_{\rm dif}};Y|X_{s_{\rm eq}})|^q\big]^{1/q} \le 2\Big(2\Big(\frac{q}{e}\Big)^q\Big)^{1/q}, \quad \text{(A.7)}$$

or equivalently

$$\begin{aligned}\mathbb{E}\big[|\imath(X_{s_{\rm dif}};Y|X_{s_{\rm eq}})|^q\big] &\le \Big(\frac{q}{e}\Big)^q 8\cdot 2^{q-2} && \text{(A.8)}\\ &\le \frac{q!}{2}16\cdot 2^{q-2}, && \text{(A.9)}\end{aligned}$$

where (A.9) follows since $\big(\frac{q}{e}\big)^q \le q!$. We obtain Proposition 3 using Lemma 1 with $c = 2$, $\tau = 16n$, and $t = \delta n$. □

Proposition 4. *For the noiseless group testing problem, consider sequences $k \to \infty$ and ℓ (indexed by p) such that $\frac{\ell}{k} \to 0$. For any $\epsilon > 0$ and $\delta_2 > 0$ (not depending on p), the following holds for sufficiently large p:*

$$\mathbb{P}\Big[\imath^n(\mathbf{X}_{s_{\rm dif}};\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}) \le nI(\ell)(1-\delta_2)\Big] \le \exp\bigg(-n\frac{\ell}{k}e^{-\nu}\nu\Big((1-\delta_2)\log(1-\delta_2)+\delta_2\Big)(1-\epsilon)\bigg) \quad \text{(A.10)}$$

for all $(s_{\rm dif}, s_{\rm eq})$ with $|s_{\rm dif}| = \ell$.

Proof. We begin by evaluating the information density in (3.18); for brevity, we write $\imath_\ell := \imath(X_{s_{\rm dif}};Y|X_{s_{\rm eq}})$ and $\imath_\ell^n := \imath^n(\mathbf{X}_{s_{\rm dif}};\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})$. Recalling the system model in (1.1) (with $Z = 0$) and the fact that $\mathbf{X}$ is i.i.d. on $P_X \sim \text{Bernoulli}\big(\frac{\nu}{k}\big)$ and $\ell = o(k)$, we obtain the following:

1. We have $X_{s_{\rm eq}} \ne \mathbf{0}$ with probability $1-\big(1-\frac{\nu}{k}\big)^{k-\ell} = (1-e^{-\nu})(1+o(1))$, and in this case we have $\imath_\ell = 0$.
2. Given $X_{s_{\rm eq}} = \mathbf{0}$, we have $X_{s_{\rm dif}} \ne \mathbf{0}$ with probability $1-\big(1-\frac{\nu}{k}\big)^\ell = \frac{\nu\ell}{k}(1+o(1))$, and in this case we have $\imath_\ell = \log\frac{1}{1-(1-\frac{\nu}{k})^\ell} = \big(\log\frac{k}{\ell}\big)(1+o(1))$.
3. Given $X_{s_{\rm eq}} = \mathbf{0}$, we have $X_{s_{\rm dif}} = \mathbf{0}$ with probability $\big(1-\frac{\nu}{k}\big)^\ell = 1+o(1)$, and in this case we have $\imath_\ell = \log\frac{1}{(1-\frac{\nu}{k})^\ell} = \frac{\nu\ell}{k}(1+o(1))$.

Note that the asymptotic identities given here follow from the assumption $\ell = o(k)$, along with standard asymptotic expansions.

Let N_1 (respectively, N_0) be the random number of measurements such that $X_{s_{\rm eq}} = \mathbf{0}$ and $X_{s_{\rm dif}} \ne \mathbf{0}$ (respectively, $X_{s_{\rm eq}} = \mathbf{0}$ and $X_{s_{\rm dif}} = \mathbf{0}$). For any $\epsilon_1 \in (0,1)$, the above observations imply the following with probability one when p is sufficiently large:

$$\begin{aligned}\imath_\ell^n &\ge N_1\Big(\log\frac{k}{\ell}\Big)(1-\epsilon_1) + N_0\nu\frac{\ell}{k}(1-\epsilon_1) && \text{(A.11)}\\ &\ge N_1\Big(\log\frac{k}{\ell}\Big)(1-\epsilon_1). && \text{(A.12)}\end{aligned}$$

We also have from (3.43) that $I(\ell) \le \big(e^{-\nu}\nu\frac{\ell}{k}\log\frac{k}{\ell}\big)(1+\epsilon_1)$ for sufficiently large p. Combining these, we conclude that

$$N_1 > n\frac{1+\epsilon_1}{1-\epsilon_1}e^{-\nu}\nu\frac{\ell}{k}(1-\delta_2) \implies \imath_\ell^n > nI(\ell)(1-\delta_2). \quad \text{(A.13)}$$

By considering the contrapositive statement, we have for any $\epsilon_2 > 0$ and sufficiently large p that

$$\begin{aligned}&\mathbb{P}\Big[\imath^n(\mathbf{X}_{s_{\rm dif}};\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}) \le nI(\ell)(1-\delta_2)\Big]\\ &\quad\le \mathbb{P}\Big[N_1 \le ne^{-\nu}\nu\frac{\ell}{k}(1-\delta_2)(1+\epsilon_2)\Big]. && \text{(A.14)}\end{aligned}$$

By the observations at the start of this subsection, we have $N_1 \sim \text{Binomial}(n,q)$ with $q = e^{-\nu}\nu\frac{\ell}{k}(1+o(1))$. We can thus further upper bound the right-hand side of (A.14) by

$$\mathbb{P}\big[N_1 \le nq(1-\delta_2(1-\epsilon_3))\big] \quad \text{(A.15)}$$

for any $\epsilon_3 \in (0,1)$ and sufficiently large p; here we have used the fact that $(1-\delta_2)(1+o(1)) = (1-\delta_2(1+o(1))$, since δ_2 is fixed. It follows from a standard Chernoff-based tail bound for Binomial random variables (e.g., see [31, Sec. 4.1]) that

$$\begin{aligned}&\mathbb{P}\Big[\imath^n(\mathbf{X}_{s_{\rm dif}};\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}) \le nI(\ell)(1-\delta_2)\Big]\\ &\quad\le e^{-nq\big((1-\delta_2(1-\epsilon_3))\log(1-\delta_2(1-\epsilon_3))\big)+\delta_2(1-\epsilon_3)}. && \text{(A.16)}\end{aligned}$$

The proof is concluded by substituting $q = e^{-\nu}\nu\frac{\ell}{k}(1+o(1))$ and noting that ϵ_3 may be arbitrarily small. □

B Proofs of Auxiliary Results for the Noiseless Case

B.1 Proof of Proposition 1 As stated in [12, Eq. (36)], we have $I(\ell) = \big(1-\frac{\nu}{k}\big)^{k-\ell}H_2\big(\big(1-\frac{\nu}{k}\big)^\ell\big)$, where $H_2(\cdot)$ is the binary entropy function. For $\frac{\ell}{k} \to \alpha$ (with $k \to \infty$), we immediately obtain (3.44) using the limits

$\left(1-\frac{\nu}{k}\right)^{k-\ell} \to e^{-(1-\alpha)\nu}$ and $\left(1-\frac{\nu}{k}\right)^{\ell} \to e^{-\alpha\nu}$, along with the continuity of the binary entropy function. In the case that $\frac{\ell}{k} \to 0$, the analogous limits are $\left(1-\frac{\nu}{k}\right)^{k-\ell} \to e^{-\nu}$ and $\left(1-\frac{\nu}{k}\right)^{\ell} = 1-\frac{\nu\ell}{k}(1+o(1))$, and we obtain (3.43) using the fact that $H_2(1-\epsilon) = (-\epsilon\log\epsilon)(1+o(1))$ as $\epsilon \to 0$ (note also that $\log\frac{k}{\nu\ell} = \left(\log\frac{k}{\ell}\right)(1+o(1))$ since $\frac{k}{\ell} \to \infty$).

B.2 Proof of Proposition 2 For the first part, we write $\sum_{\ell=1}^{\lfloor \frac{k}{\log k} \rfloor} \binom{k}{\ell}\psi_\ell(n,\delta_2^{(1)}) \triangleq T_1 + T_2$, where T_1 sums the terms from 1 to $\lfloor \log k \rfloor$, and T_2 sums the terms from $\lfloor \log k \rfloor + 1$ to $\lfloor \frac{k}{\log k} \rfloor$. For each of these, we upper bound the summation by the number of terms times the maximum term.

For T_1, there are at most $\log k$ terms, and we apply the first case in (3.45), with $\delta_{2,\ell} = \delta_2^{(1)}$. The term $(1-\delta_2^{(1)})\log(1-\delta_2^{(1)})+\delta_2^{(1)}$ can be made arbitrarily close to one by choosing $\delta_2^{(1)}$ to be sufficiently small. Writing $\log\binom{k}{\ell} = \left(\ell\log\frac{k}{\ell}\right)(1+o(1))$ and performing some simple rearrangements, we obtain the following condition for $T_1 \to 0$:

$$n \geq \frac{k\log\frac{k}{\ell} + \frac{k}{\ell}\log\log k}{e^{-\nu}\nu}(1+\eta_1), \tag{B.17}$$

where η_1 may be arbitrarily small. Note that $\log\log k$ arises as the logarithm of the number of terms in the summation. We obtain (3.46) by noting that this bound is minimized at $\ell = 1$ and writing $k\log k = \left(\frac{\theta}{1-\theta}k\log\frac{p}{k}\right)(1+o(1))$ (since $k = \Theta(p^\theta)$).

For T_2, a similar argument yields (B.17) with $\frac{1}{\ell}\log k$ in place of $\frac{1}{\ell}\log\log k$ (this follows by upper bounding the number of terms in the summation by k). Since $\ell \geq \log k$, we have $\frac{1}{\ell}\log k = O(1)$, and we conclude that $T_2 \to 0$ provided that (3.46) holds.

Finally, for the second part of the proposition, we substitute the second of the cases in (3.45). By an analogous argument to that leading to (B.17), along with the scaling laws of $I(\ell)$ in (3.43)–(3.44), it is readily verified that it suffices that $n = \Omega\left(\frac{\ell\log\frac{k}{\ell}}{1+(\frac{\ell}{k}\log\frac{k}{\ell})^2}\right)$ with a sufficiently large implied constant. Using the fact that $\ell > \frac{k}{\log k}$ for this part, this reduces to $\Omega\left(\frac{k\log k}{\log\log k}\right)$. Thus, any $\Omega(k\log k)$ scaling suffices, and the proof is concluded by noting that $\log k = \Theta\left(\log\frac{p}{k}\right)$ (since $k = \Theta(p^\theta)$).

C Proof of Theorem 2 for the Noisy Case

Here we provide the relevant details for noisy group testing, leading to Theorem 2. We focus our attention on the parts that differ from the noiseless case. Throughout the section, we use the notation $q_1 \star q_2 := q_1q_2 + (1-q_1)(1-q_2)$. We work with an arbitrary Bernoulli distribution $P_X \sim \mathrm{Bernoulli}\left(\frac{\nu}{k}\right)$ to begin, and later substitute the specific value $\nu = \log 2$.

Before proceeding, we analyze the values taken by the information density $\imath_\ell := \imath(X_{s_{\mathrm{dif}}}; Y | X_{s_{\mathrm{eq}}})$ (with $\ell := |s_{\mathrm{dif}}|$) given in (3.18), under the model in (1.1):

1. We have $X_{s_{\mathrm{eq}}} \neq \mathbf{0}$ with probability $1-\left(1-\frac{\nu}{k}\right)^{k-\ell}$, and in this case we have $\imath_\ell = 0$.

2. Given $X_{s_{\mathrm{eq}}} = \mathbf{0}$, we have the following, where we define $\xi := \left(1-\frac{\nu}{k}\right)^{\ell}$:

   - $X_{s_{\mathrm{dif}}} = \mathbf{0} \cap Y = 0$ with probability $(1-\rho)\xi$, yielding $\imath_\ell = \log\frac{1-\rho}{(1-\rho)\xi+\rho(1-\xi)}$;
   - $X_{s_{\mathrm{dif}}} = \mathbf{0} \cap Y = 1$ with probability $\rho\xi$, yielding $\imath_\ell = \log\frac{\rho}{\rho\xi+(1-\rho)(1-\xi)}$;
   - $X_{s_{\mathrm{dif}}} \neq \mathbf{0} \cap Y = 0$ with probability $\rho(1-\xi)$, yielding $\imath_\ell = \log\frac{\rho}{(1-\rho)\xi+\rho(1-\xi)}$;
   - $X_{s_{\mathrm{dif}}} \neq \mathbf{0} \cap Y = 1$ with probability $(1-\rho)(1-\xi)$, yielding $\imath_\ell = \log\frac{1-\rho}{\rho\xi+(1-\rho)(1-\xi)}$.

In the case that $\ell = o(k)$, we can write $\xi = 1-\frac{\nu\ell}{k}(1+o(1))$, yielding the following simplifications:

1. The preceding four probabilities behave as $(1-\rho)\left(1-\frac{\nu\ell}{k}(1+o(1))\right)$, $\rho\left(1-\frac{\nu\ell}{k}(1+o(1))\right)$, $\rho\frac{\nu\ell}{k}(1+o(1))$, and $(1-\rho)\frac{\nu\ell}{k}(1+o(1))$.

2. The corresponding information densities behave as $\frac{1-2\rho}{1-\rho}\frac{\nu\ell}{k}(1+o(1))$, $-\frac{1-2\rho}{\rho}\frac{\nu\ell}{k}(1+o(1))$, $-\log\frac{1-\rho}{\rho}(1+o(1))$ and $\log\frac{1-\rho}{\rho}(1+o(1))$. For example, the first of these follows by writing $\log\frac{1-\rho}{(1-\rho)(1-\frac{\nu\ell}{k})+\rho\frac{\nu\ell}{k}} = \log\frac{1-\rho}{1-\rho-(1-2\rho)\frac{\nu\ell}{k}}$, dividing the numerator and denominator by $1-\rho$, and Taylor expanding the logarithm.

C.1 Analogs of Propositions 1–2 The analog of Proposition 1 is as follows.

Proposition 5. *For the noisy group testing problem, consider arbitrary sequences of sparsity levels $k \to \infty$ and $\ell \in \{1,\dots,k\}$ (both indexed by p). If $\frac{\ell}{k} = o(1)$, then*

$$I(\ell) = \left(e^{-\nu}\nu\frac{\ell}{k}(1-2\rho)\log\frac{1-\rho}{\rho}\right)(1+o(1)). \tag{C.18}$$

Moreover, if $\frac{\ell}{k} \to \alpha \in (0,1]$, then

$$I(\ell) = e^{-(1-\alpha)\nu}\left(H_2\left(e^{-\alpha\nu}\star\rho\right) - H_2(\rho)\right)(1+o(1)). \tag{C.19}$$

Proof. We obtain (C.18) by recalling that the mutual information is the average of the information density, and applying the above-given asymptotic expansions (along with $1-\big(1-\frac{\nu}{k}\big)^{k-\ell} \to e^{-\nu}$).

To prove (C.19), we write $I(X_{s_{\rm dif}};Y|X_{s_{\rm eq}}) = H(Y|X_{s_{\rm eq}}) - H(Y|X_{s_{\rm eq}},X_{s_{\rm dif}})$. The system model (1.1) immediately gives $H(Y|X_{s_{\rm eq}},X_{s_{\rm dif}}) = H_2(\rho)$. Moreover, a direct calculation reveals that $H(Y|X_{s_{\rm eq}} = x_{s_{\rm eq}})$ equals $H_2(\rho)$ if $x_{s_{\rm eq}}$ has an entry equal to one, and $H_2\big(\xi \star \rho\big)$ otherwise, where we again write $\xi := \big(1-\frac{\nu}{k}\big)^{\ell}$. The proof is concluded by noting that $\xi \to e^{-\alpha\nu}$ when $\frac{\ell}{k} \to \alpha$, and by similarly noting that $\mathbb{P}[X_{s_{\rm eq}} = \mathbf{0}] = \big(1-\frac{\nu}{k}\big)^{k-\ell} \to e^{-(1-\alpha)\nu}$. □

As in the noiseless case, we use Proposition 3 to characterize ψ_ℓ for $\ell > \lfloor \frac{k}{\log k} \rfloor$ (and ψ' with $(s_{\rm dif},s_{\rm eq}) = (s,\emptyset)$). For $\ell \le \lfloor \frac{k}{\log k} \rfloor$, we instead use the following.

Proposition 6. *For the noisy group testing problem, consider sequences $k \to \infty$ and ℓ (indexed by p) such that $\frac{\ell}{k} \to 0$. For any $\epsilon > 0$ and $\delta_2 > 0$ (not depending on p), the following holds for sufficiently large p:*

$$\mathbb{P}\Big[\imath^n(\mathbf{X}_{s_{\rm dif}};\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}) \le nI(\ell)(1-\delta_2)\Big] \le \exp\bigg(-n\frac{\ell}{k}e^{-\nu}\nu\Big(\frac{\delta_2^2(1-2\rho)^2}{2(1+\frac{1}{3}\delta_2(1-2\rho))}\Big)(1-\epsilon)\bigg). \tag{C.20}$$

for all $(s_{\rm dif},s_{\rm eq})$ with $|s_{\rm dif}| = \ell$.

Proof. We make use of the asymptotic identities for $\imath_\ell$ at the start of this appendix. We first note that by simple averaging analogous to that used to obtain (C.18), we have $v := \mathbb{E}[\imath_\ell^2] = e^{-\nu}\nu\frac{\ell}{k}\big(\log^2\frac{1-\rho}{\rho}\big)(1+o(1))$. Moreover, we have $\imath_\ell \le \big(\log\frac{1-\rho}{\rho}\big)(1+o(1))$ with probability one. Using the form of Bernstein's inequality based on Bennet's inequality [29, Sec. 2.7], we have $\mathbb{P}[\imath^n \le n(I(\ell)-\delta)] \exp\big(-n\frac{\delta^2}{2(v+\frac{1}{3}\delta M)}\big)$ (where M is any almost-sure upper bound on $\imath_\ell$). Setting $\delta = \delta_2 I(\ell)$, substituting (C.18) and the preceding expressions for v and M, and canceling the common terms in the numerator and denominator, we obtain (C.20). □

Letting ψ_ℓ equal the right-hand side of (C.20) for $\ell \le \lfloor \frac{k}{\log k} \rfloor$ (while being the same as in (3.45) for $\ell > \lfloor \frac{k}{\log k} \rfloor$), we obtain the following.

Proposition 7. *Let $k = \Theta(p^\theta)$ for some $\theta \in (0,1)$.*

(i) For any $\eta > 0$ and $\delta_2 \in (0,1)$, there exists a choice of $\epsilon > 0$ in (3.45) such that $\sum_{\ell=1}^{\lfloor \frac{k}{\log k} \rfloor}\binom{k}{\ell}\psi_\ell(n,\delta_2) \to 0$ provided that

$$n \ge \frac{2(1+\frac{1}{3}\delta_2(1-2\rho))\frac{\theta}{1-\theta}}{e^{-\nu}\nu\delta_2^2(1-2\rho)^2}\Big(k\log\frac{p}{k}\Big)(1+\eta). \tag{C.21}$$

(ii) For any $\delta_2 \in (0,1)$, we have $\sum_{\lfloor \frac{k}{\log k} \rfloor+1}^{k}\binom{k}{\ell}\psi_\ell(n,\delta_2) \to 0$ provided that $n = \Omega\big(k\log\frac{p}{k}\big)$.

Proof. The proof is nearly identical to that of Proposition 2, except that (C.20) is used in place of (A.10), and δ_2 is kept arbitrary in the first part, rather than being taken towards one. □

Note that the choices of δ_2 in the two cases above need not coincide, since $\delta_{2,\ell}$ can vary with ℓ.

C.2 Completion of the Proof of Theorem 2

Recall that we have chosen $\nu = \log 2$. This yields $e^{-\nu}\nu = \frac{\log 2}{2}$, and thus the first term in (2.7) follows directly from (C.21).

Next, we consider the condition in (3.37) with $\ell = |s_{\rm dif}| \le \lfloor \frac{k}{\log k} \rfloor$. Letting $\delta_1 \to 0$ sufficiently slowly, applying Stirling's approximation, and substituting (C.18), we obtain the condition

$$n \ge \frac{k\log\frac{p}{\ell} + k\log k + \frac{k}{\ell}\log k}{e^{-\nu}\nu(1-2\rho)\log\frac{1-\rho}{\rho}(1-\delta_2)}(1+o(1)). \tag{C.22}$$

This is maximized for $\ell = 1$, thus yielding the second term in (2.7) upon writing $k\log k = \frac{\theta}{1-\theta}\big(k\log\frac{p}{k}\big)(1+o(1))$ and $k\log p = \frac{1}{1-\theta}\big(k\log\frac{p}{k}\big)(1+o(1))$ (since $k = \Theta(p^\theta)$).

Finally, we consider (3.37) with $\ell > \lfloor \frac{k}{\log k} \rfloor$. In this case, the numerator is dominated by the first term, and for the case that $\frac{\ell}{k} \to \alpha \in (0,1]$, we obtain the condition

$$n \ge \frac{\alpha k\log\frac{p}{k}}{e^{-(1-\alpha)\nu}\big(H_2(e^{-\alpha\nu}\star\rho) - H_2(\rho)\big)(1-\delta_2)}(1+o(1)), \tag{C.23}$$

where we have used (C.19). For the case that $\frac{\ell}{k} \to 0$ with $\ell > \lfloor \frac{k}{\log k} \rfloor$, we obtain a condition of the form (C.22) where only the first term of the numerator is kept. Such a condition is clearly dominated by (C.22).

Using the result in [9, Thm. 3a] in the limiting case that the number of defective items grows large, we have for the worst-case choice of $\alpha \in [0,1]$ and an optimized choice of $\nu > 0$ that the minimax threshold resulting from (C.23) is obtained with $\alpha = 1$ and $\nu = \log 2$. Substituting these values yields the second term in (2.6).

C.3 An Auxiliary Result for Comparing the Terms

The following result allows us to compare the terms appearing in the upper bound of Theorem 2.

Proposition 8. *For all $\rho \in (0,0.5)$, we have*

$$(1-2\rho)\log\frac{1-\rho}{\rho} \ge 4\big(\log 2 - H_2(\rho)\big). \tag{C.24}$$

Proof. By some simple manipulations, the left-hand side can be written as $\log \frac{1}{\rho(1-\rho)} - 2H_2(\rho)$, and we may thus equivalently prove that $\log \frac{1}{\rho(1-\rho)} + 2H_2(\rho) \geq 4 \log 2$. This, in turn, can be verified by showing that the minimum of the function $\log \frac{1}{\rho(1-\rho)} + 2H_2(\rho)$ occurs at $\rho = 0.5$ (i.e., the point about which it is symmetric). □

D Proof of Theorem 3 for Partial Recovery

The proof for partial recovery is similar to that of exact recovery, but the steps following the analogues of Theorems 6 and 7 become considerably simpler. Intuitively, this is because the main difficulty for exact recovery was handling "small" values of $\ell := |s_{\mathrm{dif}}|$, which need not be handled for partial recovery. We only explain the key differences here; some additional details can be found in [19].

The analysis in the derivation of (3.20) extends immediately to handle the partial recovery criterion in (1.3), since we have already split the error events according to the amount of overlap between the true set and the incorrect set. The only difference is that the decoder searches for a set s such that (3.21) holds whenever $|s_{\mathrm{dif}}| > d_{\max}$ (as opposed to $s_{\mathrm{dif}} \neq \emptyset$). It follows that Theorem 4 remains true with $P_{\mathrm{e}}(d_{\max})$ in place of P_{e} when the union in (3.20) is restricted to $|s_{\mathrm{dif}}| \in \{d_{\max}+1, \dots, k\}$.

The extension of the analysis in the proof of Theorem 5 is less immediate, but still straightforward. We first recall the observation from [21] that the performance metric in (1.3) allows us to focus without loss of generality on decoders such that the estimated defective set $\hat{S}$ (or $\hat{S}_{\mathrm{dif}} \cup S_{\mathrm{eq}}$ in the genie-aided setting) has cardinality k almost surely. For any such decoder, the definition in (1.3) is unchanged when the second term in the union is removed.

We restrict the partition $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ of s to satisfy $|s_{\mathrm{dif}}| > d_{\max}$. In (3.30)–(3.31), we change the definition of $\mathcal{D}(s_{\mathrm{dif}}|s_{\mathrm{eq}})$ to be the set of pairs $(\mathbf{x}, \mathbf{y})$ such that the decoder outputs a sequence $\hat{s}_{\mathrm{dif}}$ such that $|s_{\mathrm{dif}} \backslash \hat{s}_{\mathrm{dif}}| \leq d_{\max}$. This means that the sets $\mathcal{D}(\cdot|s_{\mathrm{eq}})$ are no longer disjoint. However, we can easily count the number of such sets that each $(\mathbf{x}, \mathbf{y})$ pair falls into. For fixed $(s_{\mathrm{eq}}, s_{\mathrm{dif}})$ and $d \in \{0, \dots, d_{\max}\}$, the number of sets $\hat{s}_{\mathrm{dif}} \subseteq \{1, \dots, p\} \backslash s_{\mathrm{eq}}$ such that $|s_{\mathrm{dif}} \backslash \hat{s}_{\mathrm{dif}}| = d$ is $\binom{p-k}{d}\binom{|s_{\mathrm{dif}}|}{|s_{\mathrm{dif}}|-d} = \binom{p-k}{d}\binom{|s_{\mathrm{dif}}|}{d}$. Thus, each $(\mathbf{x}, \mathbf{y})$ pair is included in $\sum_{d=0}^{d_{\max}} \binom{p-k}{d}\binom{|s_{\mathrm{dif}}|}{d}$ of the sets $\mathcal{D}(\cdot|s_{\mathrm{eq}})$, and (3.32) is replaced by

$$\mathbb{P}[\mathcal{A}(s_{\mathrm{eq}}) \cap \text{no error}] \leq \frac{\sum_{d=0}^{d_{\max}} \binom{p-k}{d}\binom{\ell}{d}}{\binom{p-k+\ell}{\ell}} e^{-\gamma}. \qquad \text{(D.25)}$$

Thus, Theorem 5 remains true when the pair $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ is constrained to satisfy $|s_{\mathrm{dif}}| \in \{d_{\max}+1, \dots, k\}$, and $\log \binom{p-k+|s_{\mathrm{dif}}|}{|s_{\mathrm{dif}}|}$ is replaced by $\log \binom{p-k+|s_{\mathrm{dif}}|}{|s_{\mathrm{dif}}|} - \log \sum_{d=0}^{d_{\max}} \binom{p-k}{d}\binom{|s_{\mathrm{dif}}|}{d}$.

The remainder of the analysis follows that of Section 3.4, except that the "small" values of ℓ need not be handled. That is, we need only make use of the general concentration inequality in Proposition (3), and we end up with the single condition in (2.9). For the converse part, we again choose $(s_{\mathrm{dif}}, s_{\mathrm{eq}}) = (s, \emptyset)$ in the analog of Theorem 7, and the steps are again similar, with the multiplicative factor $1 - \alpha^*$ arising showing that the above-mentioned term $\log \sum_{d=0}^{d_{\max}} \binom{p-k}{d}\binom{|s_{\mathrm{dif}}|}{d}$ (with $d_{\max} = \lfloor \alpha^* k \rfloor$ and $|s_{\mathrm{dif}}| = k$) behaves as $\big(\alpha^* k \log \frac{p}{k}\big)(1 + o(1))$, in the same way that it was shown that $\log \binom{p-k+|s_{\mathrm{dif}}|}{|s_{\mathrm{dif}}|}$ (with $|s_{\mathrm{dif}}| = k$) behaves as $\big(k \log \frac{p}{k}\big)(1 + o(1))$.

References

[1] A. Fernández Anta, M. A. Mosteiro, and J. Ramón Muñoz, "Unbounded contention resolution in multiple-access channels," in *Distributed Computing*. Springer Berlin Heidelberg, 2011, vol. 6950, pp. 225–236.

[2] R. Clifford, K. Efremenko, E. Porat, and A. Rothschild, "Pattern matching with don't cares and few errors," *J. Comp. Sys. Sci.*, vol. 76, no. 2, pp. 115–124, 2010.

[3] G. Cormode and S. Muthukrishnan, "What's hot and what's not: Tracking most frequent items dynamically," *ACM Trans. Database Sys.*, vol. 30, no. 1, pp. 249–278, March 2005.

[4] A. Gilbert, M. Iwen, and M. Strauss, "Group testing and sparse signal recovery," in *Asilomar Conf. Sig., Sys. and Comp.*, Oct. 2008, pp. 1059–1063.

[5] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, "One sketch for all: Fast algorithms for compressed sensing," in *Proc. ACM Symp. Theory Comp.*, New York, 2007, pp. 237–246.

[6] A. G. D'yachkov and V. V. Rykov, "A survey of superimposed code theory," *Prob. Contr. Inf.*, vol. 12, no. 4, pp. 1–13, 1983.

[7] D.-Z. Du and F. K. Hwang, *Combinatorial group testing and its applications*, ser. Series on Applied Mathematics. World Scientific, 1993.

[8] H.-B. Chen and F. K. Hwang, "Exploring the missing link among d-separable, $\bar{d}$-separable and d-disjunct matrices," *Disc. App. Math.*, vol. 155, no. 5, pp. 662–664, 2007.

[9] M. Malyutov, "The separating property of random matrices," *Math. notes Acad. Sci. USSR*, vol. 23, no. 1, pp. 84–91, 1978.

[10] C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama, "Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms," 2011, http://arxiv.org/abs/1107.4540.

[11] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: Bounds and simulations," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, June 2014.

[12] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, March 2012.

[13] D. Malioutov and M. Malyutov, "Boolean compressed sensing: LP relaxation for group testing," in *IEEE Int. Conf. Acoust. Sp. Sig. Proc. (ICASSP)*, March 2012, pp. 3305–3308.

[14] M. Aldridge, L. Baldassini, and K. Gunderson, "Almost separable matrices," 2014, http://arxiv.org/abs/1410.1826.

[15] T. Laarhoven, "Asymptotics of fingerprinting and group testing: Tight bounds from channel capacities," 2014, http://arxiv.org/abs/1404.2576.

[16] L. Baldassini, O. Johnson, and M. Aldridge, "The capacity of adaptive group testing," in *IEEE Int. Symp. Inf. Theory*, July 2013, pp. 2676–2680.

[17] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3065–3092, May 2012.

[18] V. Tan and G. Atia, "Strong impossibility results for sparse signal processing," *IEEE Sig. Proc. Letters*, vol. 21, no. 3, pp. 260–264, March 2014.

[19] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: An information-theoretic framework," 2015, http://infoscience.epfl.ch/record/204670.

[20] M. Mézard, M. Tarzia, and C. Toninelli, "Group testing with random pools: Phase transitions and optimal strategy," *J. Stat. Phys.*, vol. 131, no. 5, pp. 783–801, 2008.

[21] G. Reeves and M. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3451–3465, June 2013.

[22] C. Aksoylar, G. Atia, and V. Saligrama, "Sparse signal processing with linear and non-linear observations: A unified Shannon theoretic approach," April 2013, http://arxiv.org/abs/1304.0682.

[23] Y. Polyanskiy, V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[24] T. S. Han, *Information-Spectrum Methods in Information Theory.* Springer, 2003.

[25] C. E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, vol. 1, no. 1, pp. 6–25, 1957.

[26] A. Feinstein, "A new basic theorem of information theory," *IRE Prof. Group. on Inf. Theory*, vol. 4, no. 4, pp. 2–22, Sept. 1954.

[27] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[28] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[29] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence.* OUP Oxford, 2013.

[30] V. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, Feb. 2014.

[31] R. Motwani and P. Raghavan, *Randomized Algorithms.* Chapman & Hall/CRC, 2010.