Near-Optimal Noisy Group Testing
via Separate Decoding of Items

Jonathan Scarlett and Volkan Cevher

Abstract—**The group testing problem consists of determining a small set of defective items from a larger set of items based on a number of tests, and is relevant in applications such as medical testing, communication protocols, pattern matching, and more. In this paper, we revisit an efficient algorithm for noisy group testing in which each item is decoded separately (Malyutov and Mateev, 1980), and develop novel performance guarantees via an information-theoretic framework for general noise models. For the special cases of no noise and symmetric noise, we find that the asymptotic number of tests required for vanishing error probability is within a factor $\log 2 \approx 0.7$ of the information-theoretic optimum at low sparsity levels, and that with a small fraction of allowed incorrectly decoded items, this guarantee extends to all sublinear sparsity levels. In addition, we provide a converse bound showing that if one tries to move slightly beyond our low-sparsity achievability threshold using separate decoding of items and i.i.d. randomized testing, the average number of items decoded incorrectly approaches that of a trivial decoder.**

I. INTRODUCTION

The group testing problem consists of determining a small subset S of "defective" items within a larger set of items $\{1, \ldots, p\}$ based on a number of tests. This problem has a history in medical testing [1], and has regained significant attention with following applications in areas such as communication protocols [2], pattern matching [3], and database systems [4], and new connections with compressive sensing [5], [6]. In the noiseless setting, each test takes the form

$$Y = \bigvee_{j \in S} X_j, \tag{1}$$

where the test vector $X = (X_1, \ldots, X_p) \in \{0, 1\}^p$ indicates which items are included in the test, and Y is the resulting observation. That is, the output indicates whether at least one defective item was included in the test. One wishes to minimize the total number of tests n while still ensuring the reliable recovery of S. We focus on the non-adaptive setting, in which all tests must be designed in advance. The corresponding test vectors $X^{(1)}, \ldots, X^{(n)}$ are represented by the matrix $\mathbf{X} \in \{0, 1\}^{n \times p}$ with i-th row $X^{(i)}$.

Following both classical works [7]–[10] and recent advances [11]–[13], the information-theoretic performance limits of group testing have become increasingly well-understood, and

several practical near-optimal algorithms for the *noiseless* setting have been developed. In contrast, practical algorithms for *noisy* settings have generally remained less well-understood, with the best known theoretical guarantees usually being far from the information-theoretic limits (though sometimes matching in scaling laws) [14]–[16].

A notable exception to these limitations is the technique of Malyutov and Mateev [8] based on *separate decoding of items*,[1] in which each given item $j \in \{1, \ldots, p\}$ is decoded based only on the j-th column $\mathbf{X}_j$ of $\mathbf{X}$, along with $\mathbf{Y}$ (see Section I-B). This approach is computationally efficient, and was also proved to come with strong theoretical guarantees *in the case that $k := |S| = O(1)$* [8], [17] (see Section I-C).

In this paper, we develop a theoretical framework for understanding separate decoding of items, and move beyond the work of [8] in several important directions: (i) We consider the general case of $k = o(p)$, thus handling much more general scenarios corresponding to "denser" settings, and leading to non-trivial challenges in the theoretical analysis; (ii) We consider not only exact recovery, but also partial recovery, often leading to much milder requirements on the number of tests; (iii) We provide a novel converse bound revealing that under separate decoding of items, our achievability bounds cannot be improved in several cases of interest.

Before discussing the previous work and our contributions in more detail, we formally state the setup.

A. Problem Setup

We let the defective set S be uniform on the $\binom{p}{k}$ subsets of $\{1, \ldots, p\}$ of cardinality k. For convenience, we will sometimes equivalently refer to a vector $\beta \in \{0, 1\}^p$ whose j-th entry indicates whether or not item j is defective:

$$\beta_j = \mathbb{1}\{j \in S\}. \tag{2}$$

We consider i.i.d. Bernoulli testing, where each item is placed in a given test independently with probability $\frac{\nu}{k}$ 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) is denoted by $\mathbf{X} \in \{0, 1\}^{n \times p}$. Denoting the i-th entry of $\mathbf{Y}$ by $Y^{(i)}$ and the i-th row of $\mathbf{X}$ by $X^{(i)} = (X_1^{(i)}, \ldots, X_p^{(i)})$, the measurement model is given by

$$(Y^{(i)}|X^{(i)}) \sim P_{Y|N(S,X^{(i)})}, \tag{3}$$

[1]This is referred to as *separate testing of inputs* in more recent works [17], [18], but there the word "testing" refers to a hypothesis test performed at the decoder, as opposed to testing the sense of designing $\mathbf{X}$.

where $N(S, X^{(i)}) = \sum_{j=1}^{p} \mathbb{1}\{j \in S \cap X_j^{(i)} = 1\}$ denotes the number of defective items in the test. That is, we consider arbitrary noise distributions $P_{Y|N}$ for which $Y^{(i)}$ depends on $X^{(i)}$ only through $N(S, X^{(i)})$, with conditional independence among the tests $i = 1, \ldots, n$. For each item $j = 1, \ldots, p$, the j-th column of $\mathbf{X}$ is written as $\mathbf{X}_j \in \{0,1\}^n$.

While most of our results will be written in terms of general noise models of the form (3), we also pay particular attention to two specific models: The noiseless model in (1), and the *symmetric noise model* with parameter $\rho > 0$:

$$Y = \Big(\bigvee_{j \in S} X_j \Big) \oplus Z, \tag{4}$$

where $Z \sim \text{Bernoulli}(\rho)$, and $\oplus$ denotes modulo-2 addition.

In the general case (i.e., not necessarily using separate decoding of items), given $\mathbf{X}$ and $\mathbf{Y}$, a *decoder* forms an estimate $\hat{S}$ of S, or equivalently, an estimate $\hat{\beta}$ of β. We consider two related performance measures. In the case of *exact* recovery, the error probability is given by

$$P_e := \mathbb{P}[\hat{S} \neq S], \tag{5}$$

and is taken over the realizations of S, $\mathbf{X}$, and $\mathbf{Y}$ (the decoder is assumed to be deterministic). In addition, we consider a less stringent performance criterion in which we allow for up to $d_{\text{pos}} \in \{0, \ldots, p-k-1\}$ false positives and $d_{\text{neg}} \in \{0, \ldots, k-1\}$ false negatives, yielding an error probability of

$$P_e(d_{\text{pos}}, d_{\text{neg}}) := \mathbb{P}\big[|\hat{S}\backslash S| > d_{\text{pos}} \cup |S\backslash\hat{S}| > d_{\text{neg}} \big]. \tag{6}$$

In some cases (particularly for the converse) it will be convenient to consider yet another criterion in which we only seek to bound the *average number of incorrectly-decoded items* N_{err}:

$$\mathbb{E}[N_{\text{err}}] = \sum_{j=1}^{p} \mathbb{P}[\hat{\beta}_j \neq \beta_j]. \tag{7}$$

B. Separate Decoding of Items

We use the terminology *separate decoding of items* to mean any decoding scheme in which $\hat{\beta}_j$ is only a function of $\mathbf{X}_j$ and $\mathbf{Y}$, i.e.,

$$\hat{\beta}_j = \phi_j(\mathbf{X}_j, \mathbf{Y}), \quad j = 1, \ldots, p \tag{8}$$

for some functions $\{\phi_j\}_{j=1}^{p}$. All of our achievability results will choose ϕ_j not depending on j; more specifically, following [8], each decoder is of the following form for some $\gamma > 0$:

$$\phi_j(\mathbf{X}_j, \mathbf{Y}) = \mathbb{1}\Big\{ \sum_{i=1}^{n} \log \frac{P_{Y|X_j,\beta_j}(Y^{(i)}|X_j^{(i)}, 1)}{P_Y(Y^{(i)})} > \gamma \Big\}, \tag{9}$$

where P_Y is the unconditional distribution of a given observation, and $P_{Y|X_j,\beta_j}(\cdot|\cdot, 1)$ is the conditional distribution given $\beta_j = 1$ and the value of X_j. This can be interpreted as the Neyman-Pearson test for binary hypothesis testing with hypotheses $H_0 : \beta_j = 0$ and $H_1 : \beta_j = 1$.

The computational complexity of (9) is $O(n)$ for each $j = 1, \ldots, p$, for a total of $O(np)$. This matches the runtime of typical group testing algorithms [12], though it is slower than recent *sublinear-time* algorithms [15], [16] (see also [19], [20] for earlier works on non-random noise models).

Moreover, considerable speedups are possible via distributed implementations, as shown in [21], [22].

The rule (9) requires knowledge of the crossover probability ρ and the number of defectives k. In practice, one would expect estimates of these values to bear at least minor inaccuracies. Thus, it would be of significant interest to understand separate decoding rules that are mismatched or universal (e.g., see [23]); this is left for future work.

C. Related Work

Figure 1 summarizes the main results known for the noiseless and symmetric noise models (both information-theoretic and practical), along with our novel contributions. We proceed by outlining the relevant related work, and then describe our contributions in more detail and further discuss Figure 1.

Information-theoretic limits. The information-theoretic limits of group testing have long been well-understood for $k = O(1)$ in the Russian literature [7], [8], and have recently become increasingly well-understood for more general $k = o(p)$ [11], [13], [24], [25]. Here we highlight two results from [13], [25], [26] that are particularly relevant to our work:

- For the noiseless model (1) with Bernoulli testing and $k = \Theta(p^\theta)$ for some $\theta \in (0, 1)$, the minimal number of tests n^* ensuring high-probability exact recovery satisfies

$$n^* = \inf_{\nu > 0} \max \left\{ \frac{k \log \frac{p}{k}}{H_2(e^{-\nu})}, \frac{\frac{\theta}{1-\theta} k \log \frac{p}{k}}{e^{-\nu}\nu} \right\}(1 + o(1)), \tag{10}$$

  where $H_2(\cdot)$ is the binary entropy function.[2] In particular, for $\theta \leq \frac{1}{3}$, we have $n^* = \big(k \log_2 \frac{p}{k}\big)(1 + o(1))$, which is optimal even beyond Bernoulli testing.

- For the symmetric noise model (4) with $\rho > 0$, we have *for sufficiently small θ* that

$$n^* = \frac{k \log \frac{p}{k}}{\log 2 - H_2(\rho)}(1 + o(1)). \tag{11}$$

  Moreover, if we move to partial recovery with $d_{\text{pos}} = d_{\text{neg}} = \alpha^* k$ for arbitrarily small $\alpha^* > 0$, then the right-hand side of (11) is achievable for all $\theta \in (0, 1)$ (including in the noiseless case $\rho = 0$).

Practical algorithms. In the noiseless setting, numerous practical group testing algorithms have been proposed with various theoretical guarantees. The best known bounds under Bernoulli testing were given for the *definite defectives* (DD) algorithm in [25], in particular matching (10) for $\theta \geq \frac{1}{2}$. Moreover, in [27], it was shown that the same bound is attained by the linear programming (LP) relaxation techniques of [18]. The DD algorithm was also shown to yield improved bounds under a non-Bernoulli random design in [28], but in this paper we focus on Bernoulli testing.

In noisy settings, less is known. For the symmetric noise model, an algorithm called *noisy combinatorial orthogonal matching pursuit* (NCOMP) [14] (also referred to as *noisy column matching* in [29]) was shown to achieve optimal $O(k \log p)$ scaling, but the constant factors in this result are

---

[2]Here and subsequently, all logarithms have base e, and all information measures are in units of nats.
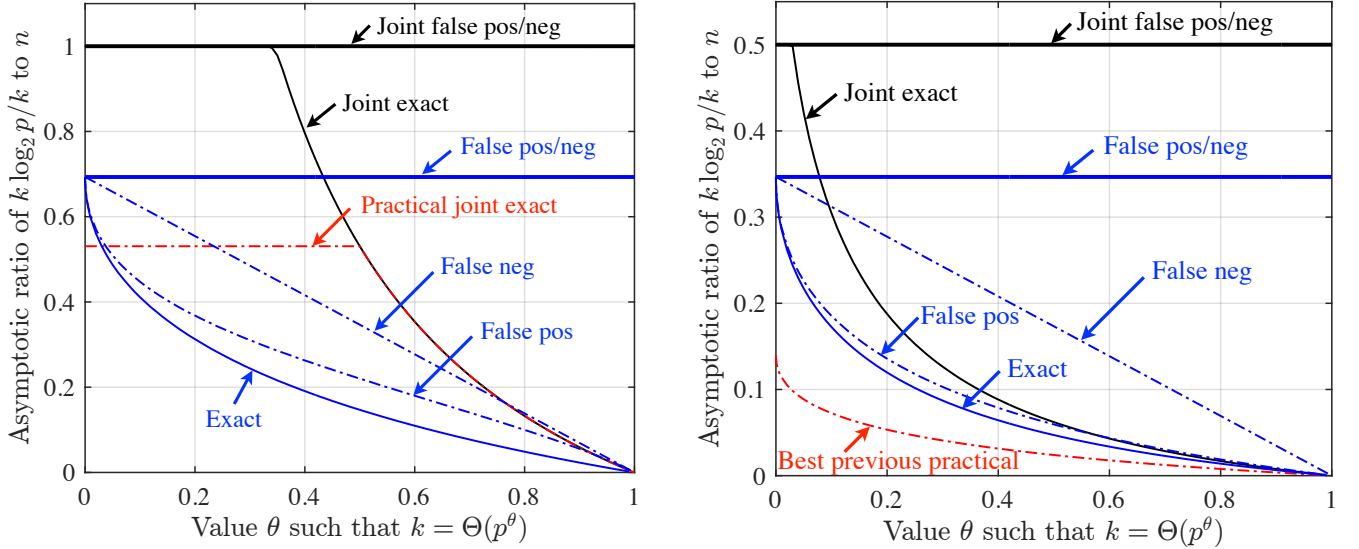
Figure 1: Asymptotic thresholds on the number of tests required for vanishing error probability in the noiseless setting (Left), and the symmetric noise setting with $\rho = 0.11$ (Right). The number of defective items is $k = \Theta(p^\theta)$ for some $\theta > 0$. The vertical axis represents the constant $c(\theta)$ such that the number of tests is $\left(\frac{1}{c(\theta)} k \log_2 \frac{p}{k}\right)(1 + o(1))$. The blue curves correspond to separate decoding of items, with exact recovery (Exact), $\Theta(k)$ false positives only (False pos), $\Theta(k)$ false negatives only (False neg), or both $\Theta(k)$ false positives and $\Theta(k)$ false negatives (False pos/neg). The "practical joint exact" curve is DD [12] or LP [18] in the noiseless case, and the "best previous practical" curve is NCOMP [14] in the noisy case.

quite suboptimal. NCOMP is in fact also an algorithm with separate decoding of items, albeit different than that of [8].

Some heuristic algorithms have been proposed for noisy settings without theoretical guarantees, including belief propagation [30] and a noisy LP relaxation [18]. A different LP relaxation was also given in [29] that only makes use of the negative tests; as a result, we found that it does not perform as well in practice. On the other hand, it was shown to yield optimal scaling laws in the symmetric noise model, with the constant factors again left loose to simplify the analysis.

Another related algorithm is the *column-based algorithm* of [31], which separately computes the number of agreements and disagreements with $\mathbf{Y}$ for each column $\mathbf{X}_j$ (though a final sorting step is also performed, so this is not a "separate decoding" algorithm according to our definition). The main focus in [31] is on the regime that $d_{\text{neg}} = 0$ and $d_{\text{pos}} = \Theta(p)$ (i.e., a very large number of false positives), which we do not consider in this paper.

Recently, algorithms based on *sparse graph codes* (with non-Bernoulli testing) have been proposed with various guarantees [15], [16]. For the noiseless non-adaptive case with exact recovery, however, the scaling laws are $O(k \log k \log p)$, thus failing to match the information-theoretic scaling $\Theta(k \log \frac{p}{k})$. Nevertheless, for partial recovery, a $O(k \log p)$ guarantee was proved in [16] (with loose constant factors). While these algorithms are suboptimal, they have the notable advantage of running in *sublinear time*.

We briefly mention that a rather different line of works has considered group testing with *adversarial* errors [19], [32], [33]. This is a fundamentally different setting to that of random errors, and the corresponding test designs and algorithms are less relevant to the present paper.

Separate decoding of items. The idea of separate decoding of items for sparse recovery problems (including group testing) was initiated by Malyutov and Mateev [8], who showed that when $k = O(1)$ and the decoder (9) is used with suitably-chosen γ, one can achieve exact recovery with vanishing error probability provided that

$$n \geq \frac{\log p}{I_1}(1 + o(1)), \qquad (12)$$

where the *single-item mutual information* I_1 is defined as follows, with implicit conditioning on item 1 being defective:

$$I_1 := I(X_1; Y). \qquad (13)$$

As noted in [17], in the noiseless setting with $\nu = \ln 2$ we have $I_1 = \frac{(\log 2)^2}{k}(1 + o(1))$ as $k \to \infty$, and in this case, (12) matches (10) up to a factor of $\log 2 \approx 0.7$. In fact, the same turns out to be true for the symmetric noise model, matching (11) up to a factor of $\log 2$, or even better if ν is further optimized (see Appendix A). In more recent works [17], [23], similar results were shown when the rule (9) is replaced by a *universal* rule (i.e., not depending on the noise distribution) based on the empirical mutual information. However, we stick to (9) in this paper, as we found it to be more amenable to general scalings of the form $k = o(p)$.

Characterizations of the mutual information I_1 for several specific noise models were given in [34].

D. Contributions

In this paper, we provide an information-theoretic framework for studying separate decoding of items with general

scalings of the form $k = o(p)$, as opposed to the case $k = O(1)$ considered in [8], [23]. As with joint decoding [13], [26], the regime $k = o(p)$ comes with significant challenges, with additional requirements on n arising from *concentration inequalities* and often dominating (12). In addition, we consider the novel aspect of partial recovery, as well as presenting a converse bound that is specific to separate decoding.

As mentioned above, our contributions for the noiseless and symmetric noise models are summarized in Figure 1, where we plot the asymptotic number of tests for achieving $P_e \to 0$ or $P_e(d_{pos}, d_{neg}) \to 0$ under Bernoulli testing with $\nu = \log 2$. Note that in this figure, the number of allowed false positives and/or false negatives (if any) is always assumed to be $\Theta(k)$, though we allow for an arbitrarily small implied constant. Moreover, the horizontal line at the top of each plot also represents a converse for joint decoding with an arbitrarily small fraction of false positives and false negatives.

We make the following observations:

- In the noiseless case, our asymptotic bounds are within a factor $\log 2$ of the optimal threshold for joint decoding as $\theta \to 0$, are reasonable for all $\theta \in (0, 1)$ with improvements when false positives or false negatives are allowed, and are within a factor $\log 2$ of the optimal joint decoding threshold *for all θ* when both are allowed. Moreover, with exact recovery and $\theta \in (0, 0.0398)$, we strictly improve on the best known bound for any efficient algorithm under Bernoulli testing.

- For the symmetric noise model, the general behavior is similar, but we significantly outperform the best known previous bound (NCOMP [14])[3] for all $\theta \in (0, 1)$. Once again, when both false positives and false negatives are allowed, we are within a factor $\log 2$ of the optimal threshold for joint decoding.

- Although it is not shown in Figure 1, we provide a converse bound showing that if one tries to move beyond our achievability threshold for $\theta \to 0$ (or equivalently, the threshold obtained at all θ with both false positives and false negatives), then any separate decoding of items scheme with Bernoulli testing must have $\mathbb{E}[N_{err}] = k(1 - o(1))$, i.e., the average number of errors is close to the trivial value of k that would be obtained by declaring every item as non-defective. In contrast, below the same threshold, we show that $\mathbb{E}[N_{err}] = o(k)$ is achievable.

The exact recovery results are given in Section II, the partial recovery results are given in Section III, and the converse bounds are given in Section IV. In addition to these theoretical developments, we evaluate the performance of separate decoding via numerical experiments in Section V, showing it to perform well despite being outperformed by the LP relaxation technique of [18].

---

[3]The final bound stated in [14, Thm. 6] appears to have a term omitted due an incorrect claim at the end of the proof stating that $\frac{\log k}{\log p}$ tends to zero (which is only true if $k = o(p^\theta)$ for all $\theta > 0$). Upon correcting this, the final bound increases from $n \approx 4.36(1 - 2\rho)^{-2} \log_2 p$ to $n \approx 4.36(1 + \sqrt{\theta})^2 (1 - 2\rho)^{-2} \log_2 p$. This correction was also made in the follow-up paper [29], but that paper only considered the suboptimal choice $\nu = 1$, yielding a bound with worse constants. We also note that the bounds in [29] for an LP relaxation (using negative tests only) are strictly worse than those stated above for NCOMP.

II. ACHIEVABILITY RESULTS WITH EXACT RECOVERY

In this section, we develop the theoretical results leading to the asymptotic bounds for the noiseless and noisy settings in Figure 1. To do this, we first establish non-asymptotic bounds on the error probability, then present the tools for performing an asymptotic analysis, and finally give the details of the applications to specific models.

A. Additional Notation

We define some further notation in addition to that in Section I-A. Our analysis will apply for any given choice of the defective set S, due to the symmetry of the observation model (3) and the i.i.d. test matrix $\mathbf{X}$. Hence, throughout this section, we will focus on the specific set $S = \{1, \ldots, k\}$. In particular, we assume that item 1 is defective, and we define $P_{Y|X_1}$ accordingly:

$$P_{Y|X_1}(y|x_1) = P_{Y|X_1, \beta_1}(y|x_1, 1). \tag{14}$$

Hence, the summation in (9) can be written as

$$\imath_1^n(\mathbf{X}_j, \mathbf{Y}) := \sum_{i=1}^n \imath_1(X_j^{(i)}, Y^{(i)}), \tag{15}$$

where

$$\imath_1(x_1, y) := \log \frac{P_{Y|X_1}(y|x_1)}{P_Y(y)}. \tag{16}$$

Following the terminology of the channel coding literature [35]–[37], we refer to this quantity as the *information density*. Denoting the distribution of a single entry of $\mathbf{X}$ by $P_X \sim \text{Bernoulli}\left(\frac{\nu}{k}\right)$, we find that the average of (16) with respect to $(X_1, Y) \sim P_X \times P_{Y|X_1}$ is the mutual information I_1 in (13). With the above distributions in place, we define $P_X^n(\mathbf{x}_1) = \prod_{i=1}^n P_X(x_1^{(i)})$, $P_Y^n(\mathbf{y}) = \prod_{i=1}^n P_Y(y^{(i)})$, and $P_{Y|X_1}^n(\mathbf{y}|\mathbf{x}_1) = \prod_{i=1}^n P_{Y|X_1}(y^{(i)}|x_1^{(i)})$.

When we specialize our results to the noiseless and symmetric noise models, we will choose

$$\nu = \nu_{symm} := \left\{ \text{unique value such that } \left(1 - \frac{\nu}{k}\right)^k = \frac{1}{2} \right\} \tag{17}$$

$$= (\log 2)(1 + o(1)). \tag{18}$$

For $k \to \infty$ (as we consider), there is essentially no difference between setting $\nu = \nu_{symm}$ or $\nu = \log 2$, but we found the former to be slightly more convenient mathematically. Either choice is known to be asymptotically optimal for maximizing I_1 in the noiseless model [17], [34]. Perhaps surprisingly, this is no longer true in general for the symmetric noise model (see Appendix A); however, it simplifies some of the analysis and does not impact the bounds significantly.

B. Initial Non-Asymptotic Bound

The following theorem provides an initial non-asymptotic upper bound on the error probability for general models. The result is proved using simple change-of-measure techniques that appeared in early studies of channel coding [38], [39], and have also been applied previously in the context of group testing [8], [13], [26].

Theorem 1. (Non-asymptotic, exact recovery) *For a general group testing model with with* $\mathrm{Bernoulli}\left(\frac{\nu}{k}\right)$ *testing and separate decoding of items according to* (9), *we have*

$$P_{\mathrm{e}} \leq k\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \leq \gamma] + (p - k)e^{-\gamma}, \qquad (19)$$

where $(\mathbf{X}_1, \mathbf{Y}) \sim P_X^n(\mathbf{x}_1)P_{Y|X_1}^n(\mathbf{y}|\mathbf{x}_1)$, *and* γ *is given in* (9).

Proof. For the exact recovery criterion, correct decoding requires the k defective items to pass the threshold test in (9), and the $p - k$ non-defective items to fail it. Letting $\overline{\mathbf{X}}_1$ be an independent copy of $\mathbf{X}_1$, i.e., $(\mathbf{X}_1, \mathbf{Y}, \overline{\mathbf{X}}_1) \sim P_X^n(\overline{\mathbf{x}}_1)P_{Y|X_1}^n(\mathbf{y}|\overline{\mathbf{x}}_1)P_X^n(\overline{\mathbf{x}}_1)$, it follows that

$$P_{\mathrm{e}} \leq k\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \leq \gamma] + (p - k)\mathbb{P}[\imath_1^n(\overline{\mathbf{X}}_1, \mathbf{Y}) > \gamma], \quad (20)$$

where the first (respectively, second) term corresponds to the union bound over the defectives (respectively, non-defectives). We bound the second term in (20) by writing

$$\mathbb{P}[\imath_1^n(\overline{\mathbf{X}}_1, \mathbf{Y}) > \gamma]$$
$$= \sum_{\overline{\mathbf{x}}_1, \mathbf{y}} P_X^n(\overline{\mathbf{x}}_1)P_Y^n(\mathbf{y})\mathbb{1}\left\{\log\frac{P_{Y|X_1}^n(\mathbf{y}|\overline{\mathbf{x}}_1)}{P_Y^n(\mathbf{y})} > \gamma\right\} \quad (21)$$
$$\leq \sum_{\overline{\mathbf{x}}_1, \mathbf{y}} P_X^n(\overline{\mathbf{x}}_1)P_{Y|X}^n(\mathbf{y}|\overline{\mathbf{x}}_1)e^{-\gamma} \qquad (22)$$
$$= e^{-\gamma}, \qquad (23)$$

where (22) bounds $P_Y^n(\mathbf{y})$ according to the event in the indicator function, and then bounds the indicator function by one. Combining (20) and (23) completes the proof. $\square$

C. Asymptotic Analysis

In order to apply Theorem 1, we need to characterize the probability appearing in the first term. The idea is to exploit the fact that $\imath_1^n(\mathbf{X}_1, \mathbf{Y})$ is an i.i.d. sum (*cf.*, (15)), and hence concentrates around its mean. While the following corollary is essentially a simple rewriting of Theorem 1, it makes the application of such concentration bounds more transparent. Here and subsequently, asymptotic notation such as $\to$, $o(\cdot)$, and $O(\cdot)$ is with respect to $p \to \infty$, and we assume that $k \to \infty$ with $k = o(p)$.

Theorem 2. (Asymptotic bound, exact recovery) *Under the setup of Theorem 1, suppose that the information density satisfies a concentration inequality of the following form:*

$$\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \leq nI_1(1 - \delta_2)] \leq \psi_n(\delta_2) \qquad (24)$$

for some function $\psi_n(\delta_2)$. *Moreover, suppose that the following conditions hold for some* $\delta_1 \to 0$ *and* $\delta_2 > 0$:

$$n \geq \frac{\log\left(\frac{1}{\delta_1}(p - k)\right)}{I_1(1 - \delta_2)}, \qquad (25)$$
$$k \cdot \psi_n(\delta_2) \to 0. \qquad (26)$$

Then $P_{\mathrm{e}} \to 0$ *under the decoder in* (9) *with* $\gamma = \log\frac{p-k}{\delta_1}$.

Proof. Setting $\gamma = \log\frac{p-k}{\delta_1}$ in Theorem 1, we obtain

$$P_{\mathrm{e}} \leq k\mathbb{P}\left[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \leq \log\frac{p-k}{\delta_1}\right] + \delta_1. \qquad (27)$$

By the condition in (25), the probability in (27) is upper bounded by $\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \leq nI_1(1 - \delta_2)]$, which in turn is upper bounded by $\psi_n(\delta_2)$ by (24). We therefore have from (27) that $P_{\mathrm{e}} \leq k\psi_n(\delta_2) + \delta_1$, and hence the theorem follows from the assumption $\delta_1 \to 0$ along with (26). $\square$

D. Concentration Bounds

In order to apply Theorem 2 to specific models, we need to characterize the concentration of $\imath_1^n(\mathbf{X}_1, \mathbf{Y})$ and attain an explicit expression for $\psi_n(\delta_2)$ in (24). The following lemma brings us one step closer to attaining explicit expressions, giving a general concentration result based on Bernstein's inequality [40, Ch. 2].

Lemma 1. (Concentration via Bernstein's inequality) *Defining*

$$c_{\mathrm{mean}} := k\mathbb{E}[\imath(X_1, Y)] = kI_1, \qquad (28)$$
$$c_{\mathrm{var}} := k\mathrm{Var}[\imath(X_1, Y)], \qquad (29)$$
$$c_{\mathrm{max}} := \max_{x_1, y}|\imath(x_1, y)|, \qquad (30)$$

we have for any $\delta_2 > 0$ *that*

$$\mathbb{P}\Big[\big|\imath_1^n(\mathbf{X}_1, \mathbf{Y}) - nI_1\big| \leq n\delta_2 I_1\Big]$$
$$\leq 2\exp\left(\frac{-\frac{1}{2} \cdot \frac{n}{k} \cdot c_{\mathrm{mean}}^2\delta_2^2}{c_{\mathrm{var}} + \frac{1}{3}c_{\mathrm{mean}}c_{\mathrm{max}}\delta_2}\right) \qquad (31)$$

Proof. Since $\imath_1^n(\mathbf{X}_1, \mathbf{Y})$ is a sum of i.i.d. random variables with mean I_1, this lemma is a direct application of Bernstein's inequality [40, Ch. 2]: For zero-mean i.i.d. random variables $\{W_i\}_{i=1}^n$ with variance at most v and maximum absolute value at most M, it holds that

$$\mathbb{P}\left[\Big|\sum_{i=1}^n W_i\Big| \leq n\delta\right] \leq 2\exp\left(\frac{-\frac{1}{2}n\delta^2}{v + \frac{1}{3}M\delta}\right). \qquad (32)$$

We have defined the values c_{mean}, c_{var} and c_{max} conveniently defined so that they behave as $\Theta(1)$ in typical examples (see Section II-E). $\square$

We will use Lemma 1 to establish the results shown for the symmetric noise model in Figure 1 (Right). While we could also use Lemma 1 for the noiseless model, it turns out that we can in fact do better via the following.

Lemma 2. (Concentration for noiseless model) *Under the noiseless model with* $\nu = \nu_{\mathrm{symm}}$ (*cf.*, (17)), *we have for any* $\delta_2 \in (0, 1)$ *that*

$$\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \leq nI_1(1 - \delta_2)] \leq \exp\bigg(-\frac{n(\log 2)^2}{k}$$
$$\times \Big((1 - \delta_2)\log(1 - \delta_2) + \delta_2\Big)(1 + o(1))\bigg) \qquad (33)$$

as $p \to \infty$ *and* $k \to \infty$ *simultaneously.*

Proof. We begin by characterizing at the various outcomes of (X_1, Y) and their probabilities, as well as the resulting values of $\imath_1(X_1, Y)$. Since $P_Y(0) = P_Y(1) = \frac{1}{2}$ by the definition of ν_{symm}, the information density simplifies to $\imath_1(x_1, y) = \log\big(2P_{Y|X_1}(y|x_1)\big)$, and we have the following:

- $X_1 = 1$ with probability $\frac{\nu}{k}$, and in this case we deterministically have $Y = 1$, yielding $\imath_1 = \log 2$.
- $X_1 = 0$ with probability $1 - \frac{\nu}{k}$, and conditioned on this event we have the following:
  - $Y = 0$ with probability $\left(1 - \frac{\nu}{k}\right)^{k-1} = \frac{1}{2} \cdot \frac{1}{1 - \nu/k} = \frac{1}{2} \cdot \left(1 + \frac{\nu/k}{1 - \nu/k}\right) = \frac{1}{2}(1 + o(1))$, where the first equality follows from the definition of ν_{symm}. Hence, in this case we have $\imath_1 = \log\left(1 + \frac{\nu/k}{1 - \nu/k}\right) = \frac{\nu}{k}(1 + o(1))$.
  - $Y = 1$ with probability $1 - \frac{1}{2} \cdot \left(1 + \frac{\nu/k}{1 - \nu/k}\right) = \frac{1}{2} \cdot \left(1 - \frac{\nu/k}{1 - \nu/k}\right) = \frac{1}{2}(1 + o(1))$, and in this case we have $\imath_1 = \log\left(1 - \frac{\nu/k}{1 - \nu/k}\right) = -\frac{\nu}{k}(1 + o(1))$.

From these calculations and (18), we immediately obtain

$$I_1 = \mathbb{E}[\imath(X_1, Y)] = \frac{(\log 2)^2}{k}(1 + o(1)). \tag{34}$$

To ease the notation, we momentarily omit the arguments to $\imath_1^n$ and $\imath^{(i)}$. By defining $V^{(i)} = \imath_1^{(i)} \cdot \mathbb{1}\{X_1^{(i)} = 1\}$ and $W^{(i)} = \imath_1^{(i)} \cdot \mathbb{1}\{X_1^{(i)} = 0\}$, we deduce that $\imath_1^n = \sum_{i=1}^n V^{(i)} + \sum_{i=1}^n W^{(i)}$, and by the above calculations, the individual distributions of each $V^{(i)}$ and $W^{(i)}$ are as follows:

$$V = \begin{cases} \log 2 & \text{w.p. } \frac{\nu}{k} \\ 0 & \text{w.p. } 1 - \frac{\nu}{k}, \end{cases} \tag{35}$$

$$W = \begin{cases} 0 & \text{w.p. } \frac{\nu}{k} \\ \log\left(1 + \frac{\nu/k}{1 - \nu/k}\right) & \text{w.p. } \frac{1}{2}\left(1 + \frac{\nu/k}{1 - \nu/k}\right) \\ \log\left(1 - \frac{\nu/k}{1 - \nu/k}\right) & \text{w.p. } \frac{1}{2}\left(1 - \frac{\nu/k}{1 - \nu/k}\right). \end{cases} \tag{36}$$

We proceed by fixing $\delta' > 0$ (later to be equated with $\delta_2(1 + o(1))$) and $\epsilon \in (0, \delta')$ (later to be taken to zero), and writing

$$\mathbb{P}\left[\imath_1^n \leq \frac{n\nu \log 2}{k}(1 - \delta')\right]$$
$$\leq \mathbb{P}\left[\sum_{i=1}^n V^{(i)} \leq \frac{n\nu \log 2}{k}(1 - \delta' + \epsilon)\right.$$
$$\left. \cup \sum_{i=1}^n W^{(i)} \leq \frac{-n\nu\epsilon \log 2}{k}\right] \tag{37}$$
$$\leq \mathbb{P}\left[\sum_{i=1}^n V^{(i)} \leq \frac{n\nu \log 2}{k}(1 - \delta' + \epsilon)\right]$$
$$+ \mathbb{P}\left[\sum_{i=1}^n W^{(i)} \leq \frac{-n\nu\epsilon \log 2}{k}\right] \tag{38}$$
$$=: T_1 + T_2, \tag{39}$$

where (37) follows since if both of the events on the right-hand side are violated then so is the event on the left-hand side, and (38) follows from the union bound.

To bound the term T_1, we simply apply the multiplicative form of the Chernoff bound for Binomial random variables [41, Sec. 4.1] to obtain

$$T_1 \leq \exp\left(-\frac{n\nu(\log 2)}{k}\left((1 - \delta' + \epsilon)\log(1 - \delta' + \epsilon) + \delta' + \epsilon\right)\right), \tag{40}$$

since we have $\mathbb{E}[V] = \frac{\nu \log 2}{k}$.

As for T_2, a direct calculation yields $\mathbb{E}[W] = o\left(\frac{1}{k^2}\right)$ and $\max[|W|] = O\left(\frac{1}{k}\right)$. Hence, by Hoeffding's inequality [40, Ch. 2], we have for any fixed $\epsilon > 0$ (not depending on p) that

$$T_2 \leq \exp\left(-nC\epsilon^2\right) \tag{41}$$

for some constant $C > 0$ and sufficiently large p. In particular, since $k \to \infty$, we have $T_2 = o(T_1)$.

Substituting the preceding bounds into (39), we obtain (33) with $1 - \delta' - \epsilon$ in place of δ_2, and $\frac{\nu \log 2}{k}$ in place of I_1. The lemma is concluded by noting that ϵ may be arbitrarily small, and noting from (18) and (34) that $\frac{\nu \log 2}{k} = I_1(1 + o(1))$. $\square$

E. Applications to Specific Models

Noiseless model: For the noiseless group testing model (*cf.*, (1)), we immediately obtain the following from Theorem 2 and Lemma 2.

Corollary 1. (Noiseless, exact recovery) *For the noiseless group testing problem with $\nu = \nu_{\text{symm}}$ (cf., (17)) and $k = \Theta(p^\theta)$ for some $\theta \in (0, 1)$, we can achieve $P_e \to 0$ with separate decoding of items provided that*

$$n \geq \min_{\delta_2 > 0} \max\left\{\frac{k \log p}{(\log 2)^2(1 - \delta_2)},\right.$$
$$\left.\frac{k \log k}{(\log 2)^2\left((1 - \delta_2)\log(1 - \delta_2) + \delta_2\right)}\right\}(1 + \eta) \tag{42}$$

for some $\eta > 0$.

Proof. We know from (34) that $I_1 = \frac{(\log 2)^2}{k}(1 + o(1))$, and hence, the first term in (42) follows from (25) with $\delta_1 \to 0$ sufficiently slowly. Moreover, by equating $\psi_n(\delta_2)$ with the right-hand side of (33) and performing simple rearranging in (26), we obtain the second term in (42). $\square$

Symmetric noise model: For the symmetric noisy model (*cf.*, (1)), we make use of Lemma 1, with the constants c_{mean}, c_{var} and c_{max} therein characterized in the following. Recall that H_2 is the binary entropy function in nats.

Lemma 3. (Bernstein parameters for symmetric noise) *Under the symmetric noise model with a fixed parameter $\rho \in \left(0, \frac{1}{2}\right)$ (not depending on p) and $\nu = \nu_{\text{symm}}$ (cf., (17)), we have*

$$k\mathbb{E}[\imath(X_1, Y)] = (\log 2)\left(\log 2 - H_2(\rho)\right)(1 + o(1)), \tag{43}$$

$$k\text{Var}[\imath(X_1, Y)] \leq (\log 2)\left((1 - \rho)\log^2\left(2(1 - \rho)\right)\right.$$
$$\left. + \rho \log^2(2\rho)\right)(1 + o(1)), \tag{44}$$

$$\max_{x_1, y}\left|\imath(x_1, y)\right| = \log\frac{1}{2\rho} \tag{45}$$

as $p \to \infty$ and $k \to \infty$ simultaneously.

Proof. We begin by looking at the various possible outcomes of (X_1, Y) and their probabilities, as well as the resulting values of $\imath_1(X_1, Y)$. Since $P_Y(0) = P_Y(1) = \frac{1}{2}$ by the definition of ν_{symm}, the information density simplifies to $\imath_1(x_1, y) = \log\left(2P_{Y|X_1}(y|x_1)\right)$, and we have the following:

- $X_1 = 1$ with probability $\frac{\nu}{k}$, and conditioned on this event, we have the following:
  - $Y = 1$ with probability $1 - \rho$, and in this case we have $\imath_1 = \log\big(2(1-\rho)\big)$.
  - $Y = 0$ with probability ρ, and in this case we have $\imath_1 = \log(2\rho)$.
- $X_1 = 0$ with probability $1 - \frac{\nu}{k}$, and conditioned on this event, we have the following:
  - $Y = 0$ with probability $(1 - \rho) \cdot \frac{1}{2}(1 + \xi) + \rho \cdot \frac{1}{2}(1 - \xi) = \frac{1}{2}\big(1 + (1 - 2\rho)\xi\big)$, where $\xi = \frac{\nu/k}{1 - \nu/k}$ as derived following (33). Hence, in this case, we have $\imath_1 = \log\big(1 + (1 - 2\rho)\xi\big) = \frac{(1-2\rho)\nu}{k}(1 + o(1))$.
  - $Y = 1$ with probability $(1 - \rho) \cdot \frac{1}{2}(1 - \xi) + \rho \cdot \frac{1}{2}(1 + \xi) = \frac{1}{2}\big(1 - (1 - 2\rho)\xi\big)$, and in this case, we have $\imath_1 = \log\big(1 - (1 - 2\rho)\xi\big) = -\frac{(1-2\rho)\nu}{k}(1 + o(1))$.

With these computations in place, the lemma follows easily by evaluating the expectation $\mathbb{E}[\imath]$, variance $\mathbb{E}[(\imath - I_1)^2]$, and maximum $\max[\imath]$ directly, and substituting $\nu = \nu_{\text{symm}} = (\log 2)(1 + o(1))$. We briefly outline the details:

- For the mean, the contributions corresponding to $X_1 = 1$ already give the right-hand side of (43), whereas for $X_1 = 0$, the terms corresponding to $Y = 0$ and $Y = 1$ effectively cancel, i.e., their sum is $o\big(\frac{1}{k}\big)$.
- For the variance, the contributions corresponding to $X_1 = 1$ already give the right-hand side of (44), whereas the contributions from the sub-cases of $X_1 = 0$ are $O\big(\frac{1}{k^2}\big)$.
- For the maximum, we use the fact that $\log\frac{1}{2\rho} \geq \log\big(2(1-\rho)\big)$ for all $\rho \in \big(0, \frac{1}{2}\big)$.

□

From this lemma, we immediately obtain the following.

Corollary 2. (Symmetric noise, exact recovery) *For noisy group testing with $\rho \in \big(0, \frac{1}{2}\big)$ (not depending on p), $\nu = \nu_{\text{symm}}$, and $k = \Theta(p^\theta)$ for some $\theta \in (0, 1)$, we can achieve $P_e \to 0$ with separate decoding of items provided that*

$$n \geq \min_{\delta_2 > 0} \max \left\{ \frac{k \log p}{(\log 2)(\log 2 - H_2(\rho))(1 - \delta_2)}, \right.$$
$$\left. \frac{(k \log k) \cdot \big(c_{\text{var}} + \frac{1}{3} c_{\text{mean}} c_{\text{max}} \delta_2\big)}{\frac{1}{2} c_{\text{mean}}^2 \delta_2^2} \right\}(1 + \eta) \quad (46)$$

for some $\eta > 0$, where c_{mean}, c_{var}, and c_{max} are respectively given by the right-hand sides of (43)–(45).

We have focused on the case $\nu = \nu_{\text{symm}}$ to simplify the analysis and establish an explicit $\log 2$ gap to joint decoding as $\theta \to 0$ (in which case, the first term of (46) dominates for arbitrarily small δ_2). However, we show in Appendix A that this choice can be suboptimal even as $\theta \to 0$, and that more generally we obtain the sufficient condition $n \geq \frac{k \log p}{\nu D_2(\rho \| \rho \star e^{-\nu})}(1 + \eta)$ in this limit, where $a \star b = a(1 - b) + b(1 - a)$, and $D_2(a \| b) = a \log \frac{a}{b} + (1 - a) \log \frac{1-a}{1-b}$ is the binary KL divergence function. This turns out to marginally improve on the condition $n \geq \frac{k \log p}{(\log 2)(\log 2 - H_2(\rho))}(1 + \eta)$ obtained via the specific choice $\nu = \nu_{\text{symm}} = (\log 2)(1 + o(1))$.

Other noise models: We showed above how to apply Lemma 1 to the symmetric noise model. However, it can also be

applied more generally, yielding an analogous result for any model in which the quantities c_{mean}, c_{var}, and c_{max} in (28)–(30) behave as $\Theta(1)$. In particular, for any such model and any fixed $\nu > 0$, in the limit as $\theta \to 0$, it suffices to have

$$n \geq \frac{\log p}{I_1}(1 + \eta) = \frac{k \log p}{c_{\text{mean}}}(1 + \eta), \quad (47)$$

for arbitrarily small $\eta > 0$. In contrast, for θ strictly greater than zero, the conditions on n resulting from Bernstein's inequality may dominate (47), similarly to Corollary 2.

In the following section, we show that when we move to *partial recovery*, it is possible to circumvent the difficulties of the concentration bounds for $\theta > 0$, and to derive sufficient conditions of the form (47) valid *for all* $\theta \in (0, 1)$, even when the term $k \log p$ is improved to $k \log \frac{p}{k}$.

III. ACHIEVABILITY RESULTS WITH PARTIAL RECOVERY

In this section, we show that the analysis of the previous section can easily be adapted to provide achievability results when a certain number false positives d_{pos} and/or false negatives d_{neg} are allowed. We make use of the notation from Sections I-A and II-A.

The main tool we need is the following, whose proof is in fact implicit in our analysis for the exact recovery criterion.

Lemma 4. (Auxiliary result for partial recovery) *For any group testing model of the form* (3)*, under the decoder in* (9) *with threshold $\gamma > 0$, we have the following:*

(i) For any $j \notin S$, the probability of passing the threshold test is upper bounded by $e^{-\gamma}$.

(ii) Suppose that the information density satisfies a concentration inequality of the form (24) *for some function $\psi_n(\delta_2)$, and that the number of tests satisfies $n \geq \frac{\gamma}{I_1(1 - \delta_2)}$. Then for any $j \in S$, the probability of failing the threshold test is upper bounded by $\psi_n(\delta_2)$.*

Proof. The first part was shown in (23), and the second part is implicit in the proof of Theorem 2. □

A. General Partial Recovery Achievability Results

We proceed by giving three variations of Theorem 2 for the case that way may tolerate false positives, false negatives, or both. We focus on the case that the number of false negatives and/or false positives is $\Theta(k)$, but we note that the implied constant in the $\Theta(\cdot)$ notation may be arbitrarily small.

We begin with the case that only false positives are allowed. This setting is closely related to that of *list decoding*, which was studied in [9], [31], [42], [43]. Recall that asymptotic notation such as $\to, o(\cdot), O(\cdot)$ is with respect to $p \to \infty$, and we assume that $k \to \infty$ with $k = o(p)$.

Theorem 3. (Asymptotic bound, false positives only) *Consider the group testing problem with $d_{\text{pos}} = \alpha_{\text{pos}} k$ for some $\alpha_{\text{pos}} > 0$ (not depending on p) and $d_{\text{neg}} = 0$, and suppose that the information density satisfies a concentration inequality of*

the form (24) *for some* $\psi_n(\delta_2)$. *Moreover, suppose that the following conditions hold for some* $\delta_1 \to 0$ *and* $\delta_2 > 0$:

$$n \geq \frac{\log\left(\frac{1}{\delta_1} \cdot \frac{p-k}{k}\right)}{I_1(1-\delta_2)}, \tag{48}$$

$$k \cdot \psi_n(\delta_2) \to 0. \tag{49}$$

Then under the decoder in (9) *with* $\gamma = \log\left(\frac{1}{\delta_1} \cdot \frac{p-k}{k}\right)$, *we have* $P_{\mathrm{e}}(d_{\mathrm{pos}}, 0) \to 0$.

Proof. By applying the second part of Lemma 4, and following the proof of (26) in Theorem 2, we find that the probability of one or more false negatives tends to zero when (49) holds.

Let N_{pos} denote the (random) number of false positives. Setting $\gamma = \log\left(\frac{1}{\delta_1} \cdot \frac{p-k}{k}\right)$ in the first part of Lemma 4, we obtain $\mathbb{E}[N_{\mathrm{pos}}] \leq (p-k)e^{-\gamma} = k\delta_1$. By the assumption $\delta_1 \to 0$, it follows that $\mathbb{E}[N_{\mathrm{pos}}] = o(k)$. Hence, by Markov's inequality, the probability of $N_{\mathrm{pos}} > \alpha_{\mathrm{pos}}k$ must vanish for any $\alpha_{\mathrm{pos}} = \Theta(1)$. □

The main difference in Theorem 3 compared to Theorem 1 is that $\log(p-k)$ is replaced by $\log\frac{p-k}{k}$ in the numerator. While this may not appear to be a drastic change, it can lead to visible improvements (*cf.*, Figure 1), particularly for moderate to large values of θ.

Next, we consider the case that only false negatives are allowed, i.e., $d_{\mathrm{pos}} = 0$ and $d_{\mathrm{neg}} > 0$.

Theorem 4. (Asymptotic bound, false negatives only) *Consider the group testing problem with* $d_{\mathrm{neg}} = \alpha_{\mathrm{neg}}k$ *for some* $\alpha_{\mathrm{neg}} \in (0,1)$ *(not depending on* p) *and* $d_{\mathrm{pos}} = 0$, *and suppose that the information density satisfies a concentration inequality of the form* (24) *for some* $\psi_n(\delta_2)$. *Moreover, suppose that the following conditions hold for some* $\delta_1 \to 0$ *and* $\delta_2 > 0$:

$$n \geq \frac{\log\left(\frac{1}{\delta_1} \cdot (p-k)\right)}{I_1(1-\delta_2)}, \tag{50}$$

$$\psi_n(\delta_2) \to 0. \tag{51}$$

Then under the decoder in (9) *with* $\gamma = \log\frac{p-k}{\delta_1}$, *we have* $P_{\mathrm{e}}(0, d_{\mathrm{neg}}) \to 0$.

Proof. Applying the first part of Lemma 4 along with the union bound over the $p-k$ non-defectives and the choice $\gamma = \log\frac{p-k}{\delta_1}$, we find that the probability of one or more false negatives is at most δ_1, which vanishes by assumption.

Let N_{neg} denote the (random) number of false negatives. Setting $\gamma = \log\left(\frac{1}{\delta_1} \cdot \frac{p-k}{k}\right)$ in the second part of Lemma 4, we obtain $\mathbb{E}[N_{\mathrm{neg}}] \leq k\psi_n(\delta_2)$. By the assumption $\psi_n(\delta_2) \to 0$, it follows that $\mathbb{E}[N_{\mathrm{neg}}] = o(k)$. Hence, by Markov's inequality, $\mathbb{P}[N_{\mathrm{neg}} > \alpha_{\mathrm{neg}}k]$ must vanish for any $\alpha_{\mathrm{neg}} = \Theta(1)$. □

By allowing false negatives, we obtain a significantly milder condition on ψ_n in (51), corresponding to the concentration of $\imath_1^n(\mathbf{X}_1, \mathbf{Y})$. Specifically, all we need is concentration about the mean at an arbitrarily slow rate, whereas in Theorem 2 we needed a rate of $O\left(\frac{1}{k}\right)$. This turns out to significantly reduce the required number of tests for moderate to large values of θ; see Corollary 3 below, as well as Figure 1.

Finally, we consider the case that both false positives and false negatives are allowed.

Theorem 5. (Asymptotic bound, false positives and false negatives) *Consider the group testing problem with* $d_{\mathrm{pos}} = \alpha_{\mathrm{pos}}k$ *and* $d_{\mathrm{neg}} = \alpha_{\mathrm{neg}}k$ *for some* $\alpha_{\mathrm{pos}} > 0$ *and* $\alpha_{\mathrm{neg}} \in (0,1)$ *(not depending on* p), *and suppose that the information density satisfies a concentration inequality of the form* (24) *for some function* $\psi_n(\delta_2)$. *Moreover, suppose that the following conditions hold for some* $\delta_1 \to 0$ *and* $\delta_2 > 0$:

$$n \geq \frac{\log\left(\frac{1}{\delta_1} \cdot \frac{p-k}{k}\right)}{I_1(1-\delta_2)}, \tag{52}$$

$$\psi_n(\delta_2) \to 0. \tag{53}$$

Then under the decoder in (9) *with* $\gamma = \log\frac{p-k}{\delta_1}$, *we have* $P_{\mathrm{e}}(d_{\mathrm{pos}}, d_{\mathrm{neg}}) \to 0$.

Proof. This result is directly deduced from the proofs of Theorems 3 and 4. □

As we show in Corollary 5 below, this result leads to a broad class of noise models where the threshold $n^* = \frac{\log\frac{p}{k}}{I_1}(1+o(1))$ can be achieved *for all* $\theta \in (0,1)$ with partial recovery.

B. Applications to Specific Models

Noiseless model: The following corollary gives three variations of the result in Corollary 1 corresponding to the three partial recovery settings considered in the previous subsection.

Corollary 3. (Noiseless model, partial recovery) *For the noiseless group testing problem with* $\nu = \nu_{\mathrm{symm}}$ *(cf.,* (17)*) and* $k = \Theta(p^{\theta})$ *for some* $\theta \in (0,1)$, *we can achieve* $P_{\mathrm{e}}(d_{\mathrm{pos}}, d_{\mathrm{neg}}) \to 0$ *with separate decoding of items under any of the following conditions:*

(i) $d_{\mathrm{pos}} = \Theta(k)$, $d_{\mathrm{neg}} = 0$, *and*

$$n \geq \min_{\delta_2 > 0} \max\left\{\frac{k\log\frac{p}{k}}{(\log 2)^2(1-\delta_2)}, \frac{k\log k}{(\log 2)^2\left((1-\delta_2)\log(1-\delta_2) + \delta_2\right)}\right\}(1+\eta) \tag{54}$$

for some $\eta > 0$.

(ii) $d_{\mathrm{pos}} = 0$, $d_{\mathrm{neg}} = \Theta(k)$, *and*

$$n \geq \frac{k\log p}{(\log 2)^2}(1+\eta) \tag{55}$$

for some $\eta > 0$.

(iii) $d_{\mathrm{pos}} = \Theta(k)$, $d_{\mathrm{neg}} = \Theta(k)$, *and*

$$n \geq \frac{k\log\frac{p}{k}}{(\log 2)^2}(1+\eta) \tag{56}$$

for some $\eta > 0$.

Proof. These results follow in the same way as Corollary 2, but with Theorems 3–5 in place of Theorem 2. When false negatives are allowed, δ_2 can be arbitrarily small, so it is factored into the $1 + \eta$ remainder term. □

The above bounds are depicted visually and compared to existing bounds in Figure 1 (Left), and were discussed in Section I-D.

Symmetric noise model: Next, we provide an analog of Corollary 2 for partial recovery.

Corollary 4. (Symmetric noise model, partial recovery) *Under the symmetric noise model with a fixed parameter $\rho \in \left(0, \frac{1}{2}\right)$ (not depending on p) and $\nu = \nu_{\mathrm{symm}}$ (cf., (17)), we can achieve we can achieve $P_e(d_{\mathrm{pos}}, d_{\mathrm{neg}}) \to 0$ with separate decoding of items under any of the following conditions:*

(i) $d_{\mathrm{pos}} = \Theta(k)$, $d_{\mathrm{neg}} = 0$, and

$$n \geq \min_{\delta_2 > 0} \max\left\{ \frac{k \log \frac{p}{k}}{(\log 2)(\log 2 - H_2(\rho))(1 - \delta_2)}, \right.$$
$$\left. \frac{(k \log k) \cdot \left(c_{\mathrm{var}} + \frac{1}{3} c_{\mathrm{mean}} c_{\mathrm{max}} \delta_2\right)}{\frac{1}{2} c_{\mathrm{mean}}^2 \delta_2^2} \right\}(1 + \eta) \quad (57)$$

for some $\eta > 0$, where c_{mean}, c_{var}, and c_{max} are respectively given by the right-hand sides of (43)–(45).

(ii) $d_{\mathrm{pos}} = 0$, $d_{\mathrm{neg}} = \Theta(k)$, and

$$n \geq \frac{k \log p}{(\log 2)(\log 2 - H_2(\rho))}(1 + \eta) \quad (58)$$

for some $\eta > 0$.

(iii) $d_{\mathrm{pos}} = \Theta(k)$, $d_{\mathrm{neg}} = \Theta(k)$, and

$$n \geq \frac{k \log \frac{p}{k}}{(\log 2)(\log 2 - H_2(\rho))}(1 + \eta) \quad (59)$$

for some $\eta > 0$.

Proof. These bounds follow in the same way as Corollary 2, but with Theorems 3–5 in place of Theorem 2. $\square$

The above bounds are depicted visually and compared to existing bounds in Figure 1 (Right), and were discussed in Section I-D. Similarly to the discussion following Corollary 2, slightly improved bounds can be obtained by considering $\nu \neq \nu_{\mathrm{symm}}$; for instance, (59) can be generalized to $n \geq \frac{k \log \frac{p}{k}}{\nu D_2(\rho \| \rho \star e^{-\nu})}(1 + \eta)$, where $a \star b = a(1 - b) + b(1 - a)$, and $D_2(a\|b) = a \log \frac{a}{b} + (1 - a) \log \frac{1-a}{1-b}$ (see Appendix A).

Other noise models: Analogous bounds to those in Corollary 4 can be obtained in an identical manner for any noise model such that the quantities c_{mean}, c_{var}, and c_{max} in (28)–(30) behave as $\Theta(1)$. To avoid repetition, we state this formally only for the case of both false positives and false negatives.

Corollary 5. (General noise models, partial recovery) *For any group testing model such that the quantities c_{mean}, c_{var}, and c_{max} in (28)–(30) behave as $\Theta(1)$, we can achieve $P_e(d_{\mathrm{pos}}, d_{\mathrm{neg}}) \to 0$ with separate decoding provided that $d_{\mathrm{pos}} = \Theta(k)$, $d_{\mathrm{neg}} = \Theta(k)$, and*

$$n \geq \frac{\log \frac{p}{k}}{I_1}(1 + \eta) = \frac{k \log \frac{p}{k}}{c_{\mathrm{mean}}}(1 + \eta) \quad (60)$$

for some $\eta > 0$.

As mentioned in Section I-C, characterizations of c_{mean} (or equivalently, I_1) were given for several noise models in [34].

IV. CONVERSE RESULTS

In this section, we present lower bounds on the required number of tests to meet certain recovery criteria with separate decoding of items and Bernoulli testing. Specifically, these bounds apply to any decoders of the form (8), with functions ϕ_j that may differ from (9). Throughout the section, we again make use of the notation from Sections I-A and II-A.

It will prove convenient to consider converse bounds with respect to the *average* number of errors $\mathbb{E}[N_{\mathrm{err}}] = \sum_{j=1}^p P_{\mathrm{e},j}$ (*cf.*, (7)), where we define

$$P_{\mathrm{e},j} := \mathbb{P}[\hat{\beta}_j \neq \beta_j], \quad (61)$$

and where the probability is with respect to β, $\mathbf{X}$, and $\mathbf{Y}$. While this is different to the criteria in (5) and (6), our analysis in Section II for exact recovery can be interpreted as proving $\mathbb{E}[N_{\mathrm{err}}] = o(1)$ and then using $\mathbb{P}[N_{\mathrm{err}} \neq 0] \leq \mathbb{E}[N_{\mathrm{err}}]$ (i.e., the first moment method). Similarly, Theorem 5 in Section III is implicitly based on showing that $\mathbb{E}[N_{\mathrm{err}}] = o(k)$. Our main converse result, Theorem 7, shows (under mild concentration conditions) that if n is slightly below the threshold (52) in Theorem 5, then we must have $\mathbb{E}[N_{\mathrm{err}}] \geq k(1 - o(1))$ (i.e., the number of errors is close to the trivial value of k that would be attained by declaring every item to be non-defective). Hence, these results collectively indicate a *phase transition*: For n slightly below the threshold we have $\frac{\mathbb{E}[N_{\mathrm{err}}]}{k} \to 1$, but for n slightly above that threshold we have $\frac{\mathbb{E}[N_{\mathrm{err}}]}{k} \to 0$.

We emphasize that the converse bounds in this section are proved for Bernoulli testing, as opposed to arbitrary test matrices (e.g., see [44]). Converse bounds for arbitrary matrices appear to be difficult, and are left for future work.

A. Initial Non-Asymptotic Bound

We begin with a non-asymptotic lower bound on $P_{\mathrm{e},j}$, which holds for any given $j \in \{1, \ldots, p\}$. The proof is based on change-of-measure techniques that have been used extensively in channel coding [35], [36], though since the distribution of β_j is non-uniform here, the analysis is in fact more akin to *joint source-channel coding* [45].

Theorem 6. (Non-asymptotic converse for single item) *For a general group testing model with Bernoulli testing and any separate decoding of items rule of the form* (8)*, we have*

$$P_{\mathrm{e},j} \geq \frac{k}{p} \mathbb{P}\left[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \leq \log \frac{p - k}{k}\right], \quad (62)$$

where $(\mathbf{X}_1, \mathbf{Y}) \sim P_X^n(\mathbf{x}_1) P_{Y|X_1}^n(\mathbf{y}|\mathbf{x}_1)$.

Proof. Any given item is defective with probability $\frac{k}{p}$. Hence, we let $P_{\beta_j}(b_j)$ be the Bernoulli$\left(\frac{k}{p}\right)$ probability mass function. Moreover, we let $P_{Y|X_j, \beta_j}^n(\cdot|\cdot, b_j)$ be the n-th product of $P_{Y|X_j, \beta_j}(\cdot|\cdot, b_j)$, with the latter defined following (9).

By the law of total probability, under any event $\mathcal{A}_j$, we have

$$P_{\mathrm{e},j} \geq \mathbb{P}[\mathcal{A}_j] - \mathbb{P}[\mathcal{A}_j \cap \mathcal{C}_j], \quad (63)$$

where $\mathcal{C}_j = \mathbb{1}\{\hat{\beta}_j = \beta_j\}$ is the correct decoding event for item j. In analogy with change-of-measure techniques from joint source-channel coding [45], we take $\mathcal{A}_j = \left\{\log \frac{P_{\beta_j}(\beta_j) P_{Y|X_j, \beta_j}^n(\mathbf{Y}|\mathbf{X}_j, \beta_j)}{P_Y^n(\mathbf{Y})} \leq \gamma\right\}$ for some $\gamma > 0$.

To proceed, we make use of the fact that under separate decoding of items (*cf.*, (8)), the correct decoding event $\mathcal{C}_j$ depends on the tests $(\mathbf{X}, \mathbf{Y})$ only through $(\mathbf{X}_j, \mathbf{Y})$; the other columns $\mathbf{X}_{j'}$ ($j' \neq j$) of the test matrix have no impact. In

particular, for each possible outcome $b_j \in \{0, 1\}$ of β_j, we can define a set $\mathcal{D}(b_j)$ of $(\mathbf{x}_j, \mathbf{y})$ pairs that lead to the decision $\hat{\beta}_j = b_j$. Under this definition, we can bound the second term in (63) as follows:

$$\mathbb{P}[\mathcal{A}_j \cap \mathcal{C}_j]$$
$$= \sum_{b_j} \sum_{(\mathbf{x}_j, \mathbf{y}) \in \mathcal{D}(b_j)} P_{\beta_j}(b_j) P_X^n(\mathbf{x}_j) P_{Y|X_j, \beta_j}^n(\mathbf{y}|\mathbf{x}_j, b_j)$$
$$\times \mathbb{1}\left\{ \log \frac{P_{\beta_j}(b_j) P_{Y|X_j, \beta_j}^n(\mathbf{y}|\mathbf{x}_j, b_j)}{P_Y^n(\mathbf{y})} \le \gamma \right\} \quad (64)$$
$$\le \sum_{b_j} \sum_{(\mathbf{x}_j, \mathbf{y}) \in \mathcal{D}(b_j)} P_X^n(\mathbf{x}_j) P_Y^n(\mathbf{y}) e^\gamma \quad (65)$$
$$= e^\gamma, \quad (66)$$

where

- (64) follows directly from the preceding definition of $\mathcal{D}(b_j)$, the choice of $\mathcal{A}_j$, and the joint distribution $(\beta_j, \mathbf{X}_j, \mathbf{Y}) \sim P_{\beta_j} \times P_X^n \times P_{Y|X_j, \beta_j}^n$;
- (65) follows by upper bounding $P_{\beta_j}(b_j) P_{Y|X_j, \beta_j}^n(\mathbf{y}|\mathbf{x}_j, b_j)$ according to the event in the indicator function, and then upper bounding the indicator function by one;
- (66) follows since by definition, the union of the sets $\mathcal{D}(b_j)$ over all b_j is precisely the set of all $(\mathbf{x}_j, \mathbf{y})$ pairs.

Substituting (66) and the choice of $\mathcal{A}_j$ into (63), we obtain

$$P_{\mathrm{e},j} \ge \mathbb{P}\left[\log \frac{P_{\beta_j}(\beta_j) P_{Y|X_j, \beta_j}^n(\mathbf{Y}|\mathbf{X}_j, \beta_j)}{P_Y^n(\mathbf{Y})} \le \gamma \right] - e^\gamma \quad (67)$$
$$= \frac{k}{p} \mathbb{P}\left[\log \frac{P_{Y|X_j, \beta_j}^n(\mathbf{Y}|\mathbf{X}_j, 1)}{P_Y^n(\mathbf{Y})} \le \log \frac{p}{k} + \gamma \;\Big|\; \beta_j = 1 \right]$$
$$+ \left(1 - \frac{k}{p}\right) \mathbb{P}\left[\log\left(1 - \frac{k}{p}\right) \le \gamma \right] - e^\gamma, \quad (68)$$

where we have applied $\beta_j \sim \mathrm{Bernoulli}\left(\frac{k}{p}\right)$ and used $\log \frac{P_{Y|X_j, \beta_j}^n(\mathbf{Y}|\mathbf{X}_j, 0)}{P_Y^n(\mathbf{Y})} = 0$. The latter claim follows since conditioned on $\beta_j = 0$, the output is independent of $\mathbf{X}_j$ and is randomly generated according to k different columns of $\mathbf{X}$, meaning $P_{Y|X_j, \beta_j}^n(\mathbf{Y}|\mathbf{X}_j, 0) = P_Y^n(\mathbf{Y})$.

Setting $\gamma = \log\left(1 - \frac{k}{p}\right)$ in (68), the second term becomes $1 - \frac{k}{p}$, and the third term exactly cancels out the second term. Finally, the first term equals the right-hand side of (62), since by definition $(\mathbf{X}_1, \mathbf{Y})$ in (62) has the same distribution as $(\mathbf{X}_j, \mathbf{Y})$ conditioned on $\beta_j = 1$ in (68). $\qquad \square$

B. Asymptotic Analysis

In order to apply Theorem 6, we need to characterize the probability appearing in the first term. Similarly to the achievability analysis in Section II, the idea is to exploit the fact that $\imath_1^n(\mathbf{X}_1, \mathbf{Y})$ is an i.i.d. sum, and therefore concentrates around its mean.

Theorem 7. (Converse for average number of errors) *Under the setup of Theorem 6, suppose that the information density satisfies a concentration inequality of the following form:*

$$\mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \ge nI_1(1 + \delta_2)] \le \psi_n'(\delta_2) \quad (69)$$

for some function $\psi_n'(\delta_2)$. Moreover, suppose that the following conditions hold for some $\delta_2 > 0$:

$$n \le \frac{\log \frac{p-k}{k}}{I_1(1 + \delta_2)}, \quad (70)$$
$$\psi_n'(\delta_2) \to 0. \quad (71)$$

Then any decoder based on separate decoding of items must have a number of errors N_{err} satisfying $\mathbb{E}[N_{\mathrm{err}}] \ge k(1 - o(1))$.

Proof. Substituting (70) into (62) gives $P_{\mathrm{e},j} \ge \frac{k}{p} \mathbb{P}[\imath_1^n(\mathbf{X}_1, \mathbf{Y}) \le nI_1(1 + \delta_2)]$, which is lower bounded by $\frac{k}{p}(1 - \psi_n'(\delta_2))$ by the definition of ψ_n'. Hence, we have from (71) that $P_{\mathrm{e},j} \ge \frac{k}{p}(1 - o(1))$, and the theorem follows since $\mathbb{E}[N_{\mathrm{err}}] = \sum_{j=1}^p P_{\mathrm{e},j}$ by definition. $\qquad \square$

C. Application to Specific Models

The following corollary applies Theorem 7 to the noiseless setting, the symmetric noise model, and more general noise models where Bernstein's inequality can be applied.

Corollary 6. (Converses for specific models) *For the group testing problem under Bernoulli testing with parameter $\nu > 0$, and $k = \Theta(p^\theta)$ for some $\theta \in (0, 1)$, any decoding rule based on separate decoding of items must satisfy $\mathbb{E}[N_{\mathrm{err}}] \ge k(1 - o(1))$ in any of the following settings:*

(i) The tests are noiseless, and the number of tests satisfies

$$n \le \frac{k \log \frac{p}{k}}{(\log 2)^2}(1 - \eta) \quad (72)$$

for arbitrarily small $\eta > 0$.

(ii) The noise is symmetric with parameter $\rho \in \left(0, \frac{1}{2}\right)$ (not depending on p), and the number of tests satisfies

$$n \le \frac{k \log \frac{p}{k}}{\nu D_2(\rho \| \rho \star e^{-\nu})}(1 - \eta) \quad (73)$$

for arbitrarily small $\eta > 0$, where $a \star b = a(1-b) + b(1-a)$, and $D_2(a\|b) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}$.

(iii) The noise follows an arbitrary distribution such that c_{mean}, c_{var}, and c_{max} in (28)–(30) all behave as $\Theta(1)$, and the number of tests satisfies

$$n \le \frac{\log \frac{p}{k}}{I_1}(1 - \eta) = \frac{k \log \frac{p}{k}}{c_{\mathrm{mean}}}(1 - \eta) \quad (74)$$

for arbitrarily small $\eta > 0$.

Proof. Without loss of generality, we assume that the upper bounds on n in the corollary statement hold with equality, since the decoders could always choose to ignore some tests. Once n is set in this way, we simply apply Theorem 7, using Lemma 1 with $c_{\mathrm{mean}} = \Theta(1)$, $c_{\mathrm{var}} = \Theta(1)$, and $c_{\mathrm{max}} = \Theta(1)$ to establish that $\psi_n'(\delta_2) \to 0$ for arbitrarily small δ_2. For part (i), we use the fact that $\nu = \nu_{\mathrm{symm}}$ (cf., (17)) is asymptotically optimal in the noiseless setting [17], [34], and use the corresponding characterization of I_1 in (34). For part (ii), we use the fact that $I_1 = \left(\frac{\nu}{k} D_2(\rho \| \rho \star e^{-\nu})\right)(1 + o(1))$, as shown in Appendix A. $\qquad \square$

Note that when $\nu = \log 2$ or $\nu = \nu_{\mathrm{symm}}$ in (73), the denominator simplifies to $(\log 2)(\log 2 - H_2(\rho))(1 + o(1))$,

thus matching the achievability threshold in (59). Conversely, as discussed in Appendix A, the achievability arguments can be adapted to match (73) more generally.

V. NUMERICAL EXPERIMENTS

In this section, we complement our theoretical findings with numerical experiments, comparing separate decoding of items against alternative algorithms in both the noiseless and noisy settings. We will see that although the linear programming (LP) methods of [18] tend to be superior, the experiments are consistent with our theoretical findings.

In all experiments below, we let $\mathbf{X}$ be an i.i.d. Bernoulli matrix with parameter $\nu = \log 2$, and we estimate the success probabilities by averaging over 500 independent trials. All LP based methods are solved using Gurobi [46]. For separate decoding of items as in (9), in accordance with our theoretical analysis, we choose

$$\gamma = nI_1(1-\delta) \tag{75}$$

for some $\delta \in (0,1)$. Specifically, we let $\delta = 0.5$, as we found this to work well in the examples considered below.

A. Noiseless Setting

In the noiseless setting, we compare against the following existing algorithms:

- *Combinatorial Orthogonal Matching Pursuit* (COMP) [14], which declares an item as negative if and only if it appears in some negative test;
- *Definite Defectives* (DD) [12], which uses COMP to construct a set of *possible defectives*, and declares an item to be defective if and only if there exists a positive test in which it is the unique possibly defective item.
- The *linear programming* (LP) relaxation of [18], with non-integer returned values of the defective status rounded to $\{0, 1\}$.

Figure 2 plots the success probability as a function of the number of tests with $p = 3000$, $k = 30$, and $d_{\mathrm{pos}} = d_{\mathrm{neg}} = 3$. We observe that separate decoding slightly outperforms COMP, whereas both DD and LP provide better performance, with LP performing best. This is consistent with the fact that DD and LP have better theoretical guarantees in the noiseless setting for most $\theta < \frac{1}{2}$ (cf., Figure 1). Despite this, separate decoding does provide good performance, with the gaps to LP and DD being relatively small.

B. Symmetric Noise Setting

In the symmetric noise setting, we compare against the following existing algorithms:

- The The *Noisy Combinatorial Orthogonal Matching Pursuit* (NCOMP) algorithm declares the j-th item to be defective if and only if $\frac{\sum_{i=1}^{n} \mathbb{1}\{X_j^{(i)}=1 \cap Y^{(i)}=1\}}{\sum_{i=1}^{n} \mathbb{1}\{X_j^{(i)}=1\}} \geq 1 - \rho(1+\Delta)$, for some threshold $\Delta > 0$. We set $\Delta = 1.5$ based on manual tuning.
- The (noisy) *Linear Programming* (LP) of [18], which introduces slack variables indicating tests where errors
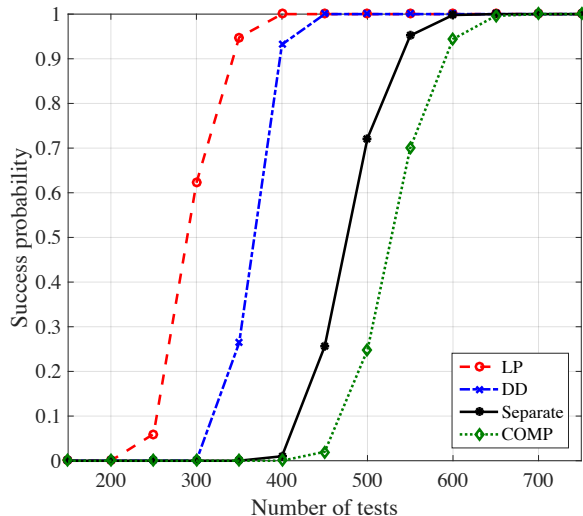

Figure 2: Experimental evaluation of practical algorithms for noiseless group testing, with $p = 3000$ items and $k = 30$ defectives, allowing up to $d_{\mathrm{pos}} = 3$ false positives and $d_{\mathrm{neg}} = 3$ false negatives.
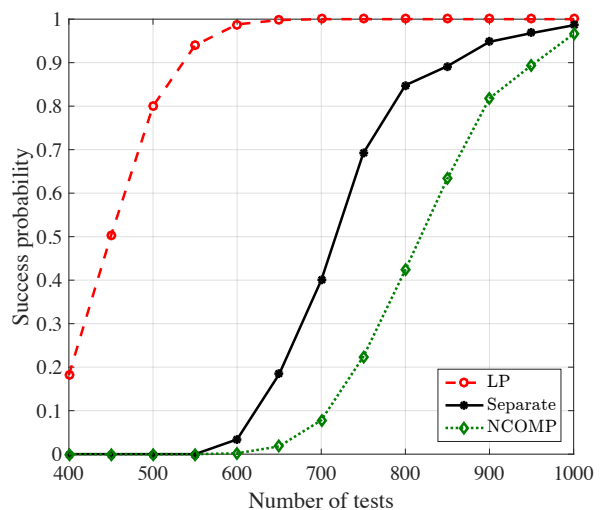

Figure 3: Experimental evaluation of practical algorithms for symmetric noisy group testing, with $p = 3000$ items, $k = 30$ defectives, and crossover probability $\rho = 0.05$, allowing up to $d_{\mathrm{pos}} = 3$ false positives and $d_{\mathrm{neg}} = 3$ false negatives.

occurred. This algorithm depends on a regularization parameter λ weighting the slack variables, and we set this to be $\lambda = 0.5$ based on manual tuning.

Figure 3 plots the success probability as a function of the number of tests with $p = 3000$, $k = 30$, $d_{\mathrm{pos}} = d_{\mathrm{neg}} = 3$, and $\rho = 0.05$. Similarly to the noiseless example, we observe that separate decoding slightly outperforms NCOMP, with LP again providing the best performance.

C. Discussion

In the above examples, we see that the LP method of [18] provides the best performance. However, we recall from Section I-C that LP methods currently only have theoretical

guarantees in the noiseless setting. Moreover, it is non-trivial to develop large-scale distributed algorithms, as was done for separate decoding of items in [21], [22]. These limitations present important directions for future research.

It may appear to be slightly surprising that the performance of NCOMP is close to the separate decoding rule of (9), since the former only depends on a small fraction of the tests (namely, those with $X_j^{(i)} = 1$). The intuition is that although most tests do not contain item j (i.e., $X_j^{(i)} = 0$), those that do contain item j are considerably more informative for determining whether j is defective. Hence, focusing only on these tests does not degrade the performance too much.

VI. Conclusion

We have provided an information-theoretic framework for studying group testing under the separate decoding of items method proposed in [8]. Our bounds are the best known for any practical group testing algorithm in several cases of interest, including (i) the noiseless model with partial recovery, and (ii) the symmetric noise model with exact recovery *or* partial recovery. Overall, the results of this paper establish separate decoding of items as a technique that is not only computationally efficient, but also comes with near-optimal theoretical guarantees.

An interesting direction for future work is to extend the achievability bounds to the case that the universal empirical mutual information based decoder [23] is used. Moreover, two key open challenges regarding the converse bounds of Section IV include (i) considering the error probabilities P_e or $P_e(d_{pos}, d_{neg})$ in place of $\mathbb{E}[N_{err}]$, and (ii) moving beyond i.i.d. testing towards arbitrary test matrices.

Appendix

A. Optimizing ν for the Symmetric Noise Model

In this section, we characterize the mutual information I_1 in (13) for the symmetric noise model described in (4), with a given crossover probability $\rho \in \left(0, \frac{1}{2}\right)$. In contrast to the noiseless model, we will see that the choices $\nu = \log 2$ and $\nu = \nu_{symm}$ (*cf.*, (17)) do not always maximize I_1.

Throughout this section, asymptotic notation is with respect to the limit as $k \to \infty$. We consider the case that the parameter ν (such that $\mathbf{X}$ is i.i.d. on $\text{Bernoulli}\left(\frac{\nu}{k}\right)$) behaves as $\Theta(1)$, since the regimes $\nu \to 0$ and $\nu \to \infty$ are easily shown to be strictly suboptimal regardless of the decoding rule.

Preliminary calculations. We have the following under the symmetric noise model (4):

$$\mathbb{P}[Y = 0] = (1 - \rho)\left(1 - \frac{\nu}{k}\right)^k + \rho\left(1 - \left(1 - \frac{\nu}{k}\right)^k\right) \quad (76)$$

$$= \left((1 - \rho)e^{-\nu} + \rho(1 - e^{-\nu})\right)(1 + o(1)). \quad (77)$$

In addition, when item 1 is defective, we have

$$\mathbb{P}[Y = 0|X_1 = 0]$$

$$= (1 - \rho)\left(1 - \frac{\nu}{k}\right)^{k-1} + \rho\left(1 - \left(1 - \frac{\nu}{k}\right)^{k-1}\right) \quad (78)$$

$$= (1 - \rho)\left(1 - \frac{\nu}{k}\right)^k \cdot \frac{1}{1 - \frac{\nu}{k}}$$
$$+ \rho\left(1 - \left(1 - \frac{\nu}{k}\right)^k \cdot \frac{1}{1 - \frac{\nu}{k}}\right) \quad (79)$$

$$= (1 - \rho)\left(1 - \frac{\nu}{k}\right)^k \cdot \left(1 + \frac{\nu}{k} + O\left(\frac{1}{k^2}\right)\right)$$
$$+ \rho\left(1 - \left(1 - \frac{\nu}{k}\right)^k \cdot \left(1 + \frac{\nu}{k} + O\left(\frac{1}{k^2}\right)\right)\right) \quad (80)$$

$$= \underbrace{\mathbb{P}[Y = 0]}_{=:\zeta} + \underbrace{\frac{\nu}{k}\left((1 - 2\rho)\left(1 - \frac{\nu}{k}\right)^k\right) + O\left(\frac{1}{k^2}\right)}_{=:\Delta}, \quad (81)$$

where (80) follows from a Taylor expansion of $\frac{1}{1 - \frac{\nu}{k}}$, and (81) follows by substituting (76).

Characterizing I_1. Using the above definitions of ζ and Δ, we simplify the mutual information I_1 in (13) as follows:

$$I_1 = H(Y) - H(Y|X_1) \quad (82)$$

$$= H_2(\zeta) - \frac{\nu}{k}H(Y|X_1 = 1) - \left(1 - \frac{\nu}{k}\right)H(Y|X_1 = 0) \quad (83)$$

$$= H_2(\zeta) - \frac{\nu}{k}H_2(\rho) - \left(1 - \frac{\nu}{k}\right)H_2(\zeta + \Delta), \quad (84)$$

where (84) follows since given $X_1 = 1$, the only uncertainty in Y is the additive noise, whereas given $X_1 = 0$, the probability that $Y = 0$ is $\zeta + \Delta$.

The derivative of $H_2(\alpha)$ is $\log \frac{1-\alpha}{\alpha}$, and hence, applying a first-order Taylor expansion in (84) yields

$$I_1 = \frac{\nu}{k}\left(H_2(\zeta) - H_2(\rho)\right) - \left(1 - \frac{\nu}{k}\right)\Delta \log \frac{1-\zeta}{\zeta} + O(\Delta^2). \quad (85)$$

Special cases $\nu = \log 2$ and $\nu = \nu_{symm}$. Setting $\nu = \log 2$, we obtain from (77) that $\zeta \to \frac{1}{2}$. Combined with the fact that $\Delta = O\left(\frac{1}{k}\right)$, we deduce from (85) that

$$I_1 = \frac{\log 2}{k}\left(\log 2 - H_2(\rho)\right)(1 + o(1)). \quad (86)$$

Here we have used $H_2\left(\frac{1}{2}\right) = \log 2$ and the continuity of entropy. Since $\nu_{symm} \to \log 2$ (*cf.*, (18)), we deduce that (86) also holds when $\nu = \nu_{symm}$.

While (86) can be used to establish the optimality of separate decoding of inputs to within a factor of $\log 2$ in certain cases, it turns out that we can in fact do better via different choices of ν.

General ν. Observing from (76) and (81) that $\Delta = \frac{\nu}{k}(\zeta - \rho) + o\left(\frac{1}{k}\right)$, we deduce from (85) that

$$I_1 = \frac{\nu}{k}\big(H_2(\zeta) - H_2(\rho)\big) - \frac{\nu}{k}(\zeta - \rho)\log\frac{1-\zeta}{\zeta} + o\left(\frac{1}{k}\right) \tag{87}$$

$$= \underbrace{\frac{\nu}{k}\left(H_2(\zeta) - (\zeta - \rho)\log\frac{1-\zeta}{\zeta} - H_2(\rho)\right)}_{=:I_1'} + o\left(\frac{1}{k}\right). \tag{88}$$

We proceed by simplifying the leading term I_1' by substituting the definition of the binary entropy function:

$$I_1' = \frac{\nu}{k}\bigg(-\zeta\log\zeta - (1-\zeta)\log(1-\zeta)$$
$$- (\zeta - \rho)\log(1-\zeta) + (\zeta - \rho)\log\zeta - H_2(\rho)\bigg) \tag{89}$$

$$= \frac{\nu}{k}\bigg(-(1-\rho)\log(1-\zeta) - \rho\log\zeta$$
$$+ (1-\rho)\log(1-\rho) + \rho\log\rho\bigg) \tag{90}$$

$$= \frac{\nu}{k}D_2(\rho\|\zeta), \tag{91}$$

where $D_2(a\|b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}$ is the binary KL divergence function.

Combining (91) with (77) and (88), we obtain

$$I_1 = \left(\frac{\nu}{k}D_2(\rho\|\rho \star e^{-\nu})\right)(1 + o(1)), \tag{92}$$

where $a \star b = a(1-b) + b(1-a)$. A simple numerical computation reveals that when $\rho > 0$ (e.g., $\rho = 0.11$), (92) may not be optimized by $\nu = \log 2$; however, we found the optimality gap to be small, and we therefore focused primarily on the latter choice.

REFERENCES

[1] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Stats.*, vol. 14, no. 4, pp. 436–440, 1943.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, "One sketch for all: Fast algorithms for compressed sensing," in *Proc. ACM-SIAM Symp. Disc. Alg. (SODA)*, New York, 2007, pp. 237–246.

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

[8] M. B. Malyutov and P. S. Mateev, "Screening designs for non-symmetric response function," *Mat. Zametki*, vol. 29, pp. 109–127, 1980.

[9] A. G. D'yachkov, "Error probability bounds for the symmetrical model of the design of screening experiments," *Prob. Inf. Transm.*, 1982.

[10] 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.

[11] 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.

[12] 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.

[13] J. Scarlett and V. Cevher, "Phase transitions in group testing," in *Proc. ACM-SIAM Symp. Disc. Alg. (SODA)*, 2016.

[14] 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," in *Allerton Conf. Comm., Ctrl., Comp.*, Sep. 2011, pp. 1832–1839.

[15] S. Cai, M. Jahangoshahi, M. Bakshi, and S. Jaggi, "Efficient algorithms for noisy group testing," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2113–2136, Apr. 2017.

[16] K. Lee, R. Pedarsani, and K. Ramchandran, "SAFFRON: A fast, efficient, and robust framework for group testing based on sparse-graph codes," 2015, https://arxiv.org/abs/1508.04485.

[17] M. B. Malyutov, "Search for sparse active inputs: A review," in *Inf. Theory, Comb. and Search Theory*, 2013, pp. 609–647.

[18] 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.

[19] H. Q. Ngo, E. Porat, and A. Rudra, "Efficiently decodable error-correcting list disjunct matrices and applications," in *Int. Colloq. Automata, Lang., and Prog.*, 2011.

[20] P. Indyk, H. Q. Ngo, and A. Rudra, "Efficiently decodable non-adaptive group testing," in *ACM-SIAM Symp. Disc. Alg. (SODA)*, 2010.

[21] M. Malyutov, P. Grosu, and H. Sadaka, "Separate testing inputs vs. linear programming relaxation," *Information Processes*, vol. 15, no. 3, pp. 351–376, 2015.

[22] P. Grosu, "Parallel implementation of information-theoretic separate testing of inputs," Ph.D. dissertation, Northeastern University, 2017.

[23] M. Malyutov and H. Sadaka, "Maximization of ESI. Jaynes principle for testing significant inputs of linear model," *Rand. Opt. Stoch. Eq.*, vol. 6, no. 4, pp. 339–358, 1998.

[24] M. Aldridge, L. Baldassini, and K. Gunderson, "Almost separable matrices," *J. Comb. Opt.*, pp. 1–22, 2015.

[25] M. Aldridge, "The capacity of Bernoulli nonadaptive group testing," 2015, http://arxiv.org/abs/1511.05201.

[26] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: An information-theoretic framework," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 593–620, 2017.

[27] M. Aldridge, "On the optimality of some group testing algorithms," in *IEEE Int. Symp. Inf. Theory*, 2017.

[28] O. Johnson, M. Aldridge, and J. Scarlett, "Performance of group testing algorithms with constant tests-per-item," 2016, submitted to *IEEE Trans. Sig. Proc.*

[29] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3019–3035, May 2014.

[30] D. Sejdinovic and O. Johnson, "Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction," in *Allerton Conf. Comm., Control and Comp.*, 2010.

[31] A. Sharma and C. R. Murthy, "Computationally tractable algorithms for finding a subset of non-defective items from a large population," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7149–7165, Nov. 2017.

[32] A. J. Macula, "Error-correcting nonadaptive group testing with de-disjunct matrices," *Disc. App. Math.*, vol. 80, no. 2-3, pp. 217–222, 1997.

[33] M. Cheraghchi, "Noise-resilient group testing: Limitations and constructions," *Disc. App. Math.*, vol. 161, no. 1, pp. 81–95, 2013.

[34] T. Laarhoven, "Asymptotics of fingerprinting and group testing: Tight bounds from channel capacities," *IEEE Trans. Inf. Forens. Sec.*, vol. 10, no. 9, pp. 1967–1980, 2015.

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

[36] 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.

[37] 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.

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

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

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

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

[42] A. G. D'yachkov, I. V. Vorobev, N. Polyansky, and V. Y. Shchukin, "Almost disjunctive list-decoding codes," *Prob. Inf. Transm.*, vol. 51, no. 2, pp. 110–131, 2015.

[43] J. Scarlett and V. Cevher, "How little does non-exact recovery help in group tesitng?" in *IEEE Int. Conf. Acoust. Sp. Sig. Proc. (ICASSP)*, New Orleans, 2017.

[44] ——, "Converse bounds for noisy group testing with arbitrary measurement matrices," in *IEEE Int. Symp. Inf. Theory*, Barcelona, 2016.

[45] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fabregas, and A. Martinez, "Random-coding joint source-channel bounds," in *IEEE Int. Symp. Inf. Theory (ISIT)*. IEEE, 2011, pp. 899–902.

[46] "Gurobi optimization," https://www.gurobi.com/.

Jonathan Scarlett (S'14 – M'15) received the B.Eng. degree in electrical engineering and the B.Sci. degree in computer science from the University of Melbourne, Australia, both in 2010. From October 2011 to August 2014, he was a Ph.D. student in the Signal Processing and Communications Group at the University of Cambridge, United Kingdom. From September 2014 to September 2017, he was a post-doctoral researcher with the Laboratory for Information and Inference Systems at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is now an assistant professor in the Department of Computer Science and Department of Mathematics, National University of Singapore. His research interests are in the areas of information theory, machine learning, signal processing, and high-dimensional statistics.

Volkan Cevher (SM'10) received the B.Sc. (valedictorian) in electrical engineering from Bilkent University in Ankara, Turkey, in 1999 and the Ph.D. in electrical and computer engineering from the Georgia Institute of Technology in Atlanta, GA in 2005. He was a Research Scientist with the University of Maryland, College Park from 2006-2007 and also with Rice University in Houston, TX, from 2008-2009. Currently, he is an Associate Professor at the Swiss Federal Institute of Technology Lausanne and a Faculty Fellow in the Electrical and Computer Engineering Department at Rice University. His research interests include signal processing theory, machine learning, convex optimization, and information theory. Dr. Cevher was the recipient of a Best Paper Award at SPARS in 2009, a Best Paper Award at CAMSAP in 2015, and an ERC StG in 2011.