The work of W. Huleihel was supported by the MIT - Technion Postdoctoral Fellowship.
its complement will be denoted by $A^c$. Finally, the indicator function of an event $A$ will be denoted by $1\{A\}$.

The set of all probability distributions on an alphabet, say $\mathcal{X}$, is denoted by $P(\mathcal{X})$. Probability mass functions (PMFs) are denoted by capital letters, such as $P$ or $Q$, with a subscript that identifies the random variable and its possible conditioning. For a sequence of random variable $X$, if the entries of $X$ are drawn in an independent and identically distributed (i.i.d.) manner according to $P_X$, then for every $x \in \mathcal{X}^n$ we have $P_X(x) = \prod_{i=1}^n P_X(x_i)$ and we denote $P_X(Y) = P_X^n(Y)$. In a similar fashion, if for every $(x, y) \in \mathcal{X}^n \times \mathcal{Y}^n$ it holds that $P_{Y|X}(y|x) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$, then we denote $P_{Y|X}(Y|X) = P_{Y|X}^n(Y|X) = P_{Y|X}^{n|Y}(Y|X)$.

Information measures induced by the generic joint distribution $Q_{XY}$, will be subscripted by $Q$, for example, $I_Q(X;Y)$ will denote the corresponding mutual information, etc. The relative entropy (or, Kullback-Liebler distance) between two probability measures $P$ and $Q$ is denoted by $D_{KL}(P||Q)$. We also use Sibson's definition of relative divergence [20]

$$I_x(X;Y) = \min_{Q_Y} D_{\alpha}(P_{Y|X}||Q_Y|P_X),$$

where

$$D_{\alpha}(P||Q) = \frac{1}{\alpha-1} \log \left( \sum_{a \in \mathcal{L}} P^\alpha(a)Q^{1-\alpha}(a) \right)$$

is the Rényi divergence of order $\alpha$, and $D_{\alpha}(P_{Y|X}||Q_{Y|X}|P_X) \triangleq D_{\alpha}(P_X \times P_{Y|X}||P_X \times Q_{Y|X})$.

For a given sequence $x \in \mathcal{X}^n$, $P_x \in \mathcal{P}_n(\mathcal{X})$ denote its empirical distribution. Similarly, for a pair of vectors $(x, y)$, the empirical joint distribution will be denoted by $P_{xy}$. All previously defined notation rules for regular distributions will also be used for empirical distributions. We use “$\bigvee$” and “$\ominus$” to denote the Boolean OR operation, and modulo-2 sum, respectively. Finally, we make use of the standard asymptotic notations $O(\cdot), o(\cdot), \Theta(\cdot)$, and $\Omega(\cdot)$.

### III. Problem Formulation

Among a population of $p$ items, $d$ unknowns are of interest. These $d$ items represent the defective set denoted by $D$. The main task in group testing is to design a collection of tests (pool), which is used to detect the defective set, with the goal of reducing the number of required tests $n$. This pool is described by an $n \times p$ binary-valued matrix, whose $(i,j)$ entry is “1” if item $j$ is included in the $i$th designed test, and “0” otherwise. Accordingly, the outcome of the $i$th test is positive if and only if the $j$th item is a member of the test. Upon observing the output for a number of $n$ tests, the goal is to recover the set $D$. In this paper, we take a probabilistic approach and require that the average probability of error will be small. Also, the design of the pool will be generated randomly. We next establish some additional notation.

We assume that $D$ is uniform on $D$, which contain the $\binom{p}{d}$ possible subsets of $\{1, 2, \ldots, p\}$ of cardinality $d$. 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 $\beta_j \triangleq 1\{j \in D\}$. For each item, we associate a “codeword” of length $n$, which is a binary sequence, denoted by $X^n_j$ for $j \in \{1, 2, \ldots, p\}$. The $i$th entry of $X_j$ is 1 if the $j$th item is included in the $i$th test, and 0 otherwise. The pool (or, codebook) is a collection of all the codewords, namely, it is $n \times p$ matrix denoted by $X_{n \times p}$. The vector of $n$ (possibly noisy) observations is denoted $Y \in \{0, 1\}^n$. In the sequel, the $i$-th entry of $Y$ is designated by $Y_i$ and the $i$-th row of $X_{n \times p}$ by $X_{i,:}$. Given a subset $S \subset \{1, 2, \ldots, p\}$ with cardinality $|S|$, the matrix $X_{n \times |S|}$ is an $n \times |S|$ matrix formed from the columns indexed by $S$.

We use the same measurement model described in [18]. Specifically,

$$Y_i | X_{i,:} \sim P_{Y|N(D,X_{i,:})}$$

where $N(D,X_{i,:}) \triangleq \sum_{j=1}^p 1\{j \in D \cap X_{i,j} = 1\}$ denotes the number of defective items in the test. We denote the $j$th column of $X_{n \times p}$ by $X_{:,j}$. It will be convenient to work with random variables that are implicitly conditioned on a fixed value of $D$, say $s = \{1, 2, \ldots, d\}$. Accordingly, we write $P_{Y|X_s}$ in place of $P_{Y|X_D}$ to emphasize that $D = s$, and we define

$$P_{Y|X_s}(y, s) \triangleq P_{Y|X_s}(y|x_s),$$

$$P_{Y|X_s}(x, y) \triangleq P_{Y|X_s}^n(y|x_s).$$

We consider partitions of the defective set $s \in D$ into two sets $s_{\text{diff}} \neq \emptyset$ and $s_{\text{eq}}$. One can think of $s_{\text{eq}}$ as corresponding to an overlap $s \cap \bar{s}$ between the true set $s$ and some incorrect set $\bar{s}$, and $s_{\text{diff}}$ corresponding to the indices in $s \setminus \bar{s}$. There are $2^d - 1$ ways of performing such a partition. For a fixed set $s \in D$ and a corresponding pair $(s_{\text{diff}}, s_{\text{eq}})$, we define

$$P_{Y|X_{s_{\text{diff}}}, X_{s_{\text{eq}}}}(y|x_{s_{\text{diff}}}, x_{s_{\text{eq}}}) \triangleq P_{Y|X_s}(y|x_s).$$

We next describe two special cases which are captured by (3).

- **Noiseless model:** In the noiseless setting, each test takes the form

$$Y = \bigvee_{i \in D} X_i,$$

namely, it is the Boolean sum of the items corresponding to the defective set.

- **Symmetric noise model:** Each test takes the form

$$Y = \left( \bigvee_{i \in D} X_i \right) \oplus W,$$

where $W \sim \text{Bernoulli}(\mu)$, for some $\mu > 0$. Note that contrary to the noiseless model, here, the outcome of a test can be “1” even though none of the defective items are included in the test.

As mentioned before, we take a probabilistic approach (rather than combinatorial), and generate the test matrix randomly. In this paper, we consider i.i.d. Bernoulli testing, where each item is placed in a given test independently with probability $P_X \triangleq \nu/d$, for some constant $\nu > 0$, which can be optimized. For the noiseless and symmetric noise models it is common and convenient to choose (e.g., [15, 18])

$$\nu = \nu_{\text{sym}} \triangleq \left( \nu \cdot \left( 1 - \frac{\nu}{d} \right) = \frac{1}{2} \right) = [1 + o(1)] \cdot \log 2.$$(8)
Given $X_{n \times p}$ and $Y$, a decoder forms an estimate $\hat{D}(X_{n \times p}, Y)$ of $D$. We focus on the average error probability as our performance measure,

$$P_e \triangleq \Pr(\hat{D}(X_{n \times p}, Y) \neq D)$$

where the probability is evaluated w.r.t. the randomness of $D$, $X_{n \times p}$, and $Y$. Our main goal is to derive necessary conditions on $n$ and $d$ (as a function of $p$) such that $P_e$ vanishes as $p \to \infty$. Specifically, we wish to find a sequence of values $n^*\gamma$, indexed by $p$, such that for all $\gamma > 0$, we have $P_e \to 0$ if $n \geq n^*(1 + \gamma)$.

It is a well-known fact that the maximum-likelihood (ML) decoder minimizes the average error probability. Naturally, this decoder requires the knowledge of the codebook distribution, and more importantly, the knowledge of the channel. Furthermore, the computational complexity of the ML decoder is exponential, which precludes its use in practice. The goal of this paper is to devise decoders which are robust in the sense that they are independent of the actual channel, but, nonetheless perform “well”. Such decoders are called universal if they achieve the same performance as the ML decoder.

In this paper, we will propose two such decoders, capturing two sides of the problem. The first is computationally expensive but in some cases provably achieves the performance of the ML decoder, and thus universal, while the second is sub-optimal, but computationally efficient. We next describe these decoders.

1. **Joint decoding**: Fix a set $s \in D$ and a corresponding pair $(s_{diff}, s_{eq})$. Let

$$\hat{i}_n(x_{s_{diff}}, y|x_{s_{eq}}) \triangleq \log \frac{\hat{P}_n(x_{s_{diff}}, x_{s_{eq}}, y)}{\hat{P}_n(x_{s_{eq}}, y) \hat{P}_n(x_{s_{diff}})}$$

where $\hat{P}_n(A)$ designates the Binomial empirical measure, namely, for a given sequences $x, y$ of length $n$,

$$\frac{1}{n} \log \hat{P}_n(x, y) \triangleq \sum_{x, y \in \{0, 1\}} \hat{P}_n(x, y) \log \hat{P}_n(x, y) = -H(\hat{P}_n)$$

Fix some constants $\gamma_1, \gamma_2, \ldots, \gamma_d \in \mathbb{R}$. The first decoder searches for the unique set $s \in D$ such that

$$\hat{i}_n(x_{s_{diff}}, y|x_{s_{eq}}) > \gamma_{|s_{diff}|}$$

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

2. **Separate decoding**: It is evident that similarly to the ML decoder, the above decoder suffers also from an exponential computational complexity, as it requires a search over all $\binom{p}{d}$ possible sets of size $d$. Thus, we next propose a family of decoders, termed separate decoding of items, which have linear computational complexity. Specifically, when the channel law is known, $\beta_j$ (which corresponds to $D$) is a function of $x, y$ and $\epsilon$ only, i.e., (see, e.g., [18]) $\beta_j = \varphi_j(x, y)$, for $j = 1, 2, \ldots, p$ and some set of functions $\{\varphi_j\}$, defined on $\{0, 1\}^n \times \{0, 1\}^n \to \mathbb{R}$. Since we are concerned with robust group testing, the above decoder is inadequate for our task. Accordingly, we define the following decoder, for some $\gamma > 0$,

$$\varphi_j(x, y) \triangleq 1 \left\{ \log \frac{\hat{P}_n(x_{s_{j}}, y)}{\hat{P}_n(x_{s_{j}}, y) \hat{P}_n(y)} > \gamma \right\}$$

(11)

For simplicity of notation we define the empirical information density as follows

$$\hat{i}(x, y) \triangleq \log \frac{\hat{P}_n(x, y)}{\hat{P}_n(x) \hat{P}_n(y)}$$

It should be mentioned that our analysis will apply for any given choice of the defective set $D$, due to the symmetry of the observation model (3) and the i.i.d. test matrix $X$. Therefore, throughout the rest of this paper we focus on the specific set $D = \{1, 2, \ldots, d\}$. Particularly, we assume that item $\#1$ is defective, and we define

$$P_{Y|X_1}(y|x_1) = P_{Y|X_1, \delta_1}(y|x_1, 1).$$

Accordingly, in the following $I(X_1; Y)$ refers to the mutual information evaluated w.r.t. $P_X \times P_{X_1}$. Similarly, note that by symmetry, the mutual information $I(X_{s_{diff}}; Y|X_{s_{eq}})$ depends only on $(s_{diff}, s_{eq})$. In the rest of this paper, asymptotic notation such as $o(\cdot), p. p. o(\cdot)$, etc., are w.r.t. $d, p \to \infty$ with $d = o(p)$. Due to space limitation and for simplicity of demonstration, in this paper, we focus on the “separate decoder” in (11), and briefly discuss the “joint decoder” (10) in Section V.

IV. SEPARATE DECODING

A. Main Results

In this subsection, we present our main results. We start with the following result which provides an upper-bound on the error probability using the proposed separate decoding procedure in (11) for the general noisy group testing model described in Section III.

**Theorem 1.** For the general group testing with Bernoulli($\nu/d$) testing and separate decoding of items (11), we have

$$P_e \leq d \cdot \mathbb{P} \left[ \hat{i}(X_{s_{diff}}, Y) \leq \gamma \right] + C \cdot (p - d) \cdot e^{-\gamma}$$

(12)

where the probability term in (12) is evaluated w.r.t. $P_X \times P_{Y|X_1}$, and $C > 0$.

Using Theorem 1 we would like to get guarantees on the sample complexity $n$ to drive the error probability $P_e$ to zero. The following result is a consequence of Theorem 1.

**Theorem 2.** Consider the general group testing with Bernoulli($\nu/d$) testing and separate decoding of items (11).

Suppose that there exists some function $\psi_n(\delta_1)$ such that

$$\mathbb{P} \left[ \hat{i}(X_{s_{diff}}, Y) \leq (1 - \delta_1) \cdot \mathbb{E}[\hat{i}(X_{s_{diff}}, Y)] \right] \leq \psi_n(\delta_1)$$

(13)

for some $\delta_1 \in (0, 1)$. Also, suppose that

$$n \geq \log \frac{1}{\psi_n(\delta_1)} \frac{(p - d)}{(1 - \delta_1) \cdot I(X_1; Y)}$$

(14a)

1Although the expression in (12) resembles [18, Theorem 1], the proof is different, due to the discrepancy of the empirical measure induced by the decoder and the true distribution.
for some \( \delta_2 \to 0 \). Then, \( P_e \to 0 \), using \( \gamma = \log \frac{p - d}{\delta_2} \).

In order to apply Theorem 2 on specific channels, it is evident that we need to derive an explicit expression for \( \psi_n \).

Corollary 1. For the noisy group testing problem, we can

\[
d \cdot \psi_n(\delta_1) \to 0, \tag{14b}
\]

for some \( \delta_2 \to 0 \). Then, \( P_e \to 0 \), using \( \gamma = \log \frac{p - d}{\delta_2} \).

The above results are general and can be applied for any noisy group testing model in the form of (3), and general scaling of \( d = o(p) \). In the next subsection, we consider a few examples, and compare our results with [18, Th. 2], where the known channel case was considered. As shall be seen, in some special cases, our proposed decoder achieves the same performances as if the channel was known.

B. Comparison and Examples

In the following we provide a few examples which illustrate the results presented in the previous subsection. To this end, we first briefly discuss the results of [18], where general achievability (lower-bounds) and converse (upper-bounds) results for the problem of noisy group testing under separate decoding (the channel law was assumed to be known) were proposed. Unfortunately, the results in [18] are not tight in the sense that the upper and lower bounds on \( n \) do not match. Tight results were shown, nonetheless, for the case where \( d = \Theta(p) \) with \( \theta \to 0 \) (after taking \( p \to \infty \)). Comparing Theorem 2 with [18, Th. 2], we see that conditions (14a)-(14b) are the same. The only difference, though, is that the function \( \psi_n \) is different, which in turn, results in a different constraint on \( n \) through (14b).

1) Fixed Defective Set Size: Consider the case where \( d = \Theta(1) \), namely, the cardinality of the defective set is a constant. When the channel law is known, the information theoretic limits of group testing are well-known [10-17]. Specifically, for the symmetric noise model (7) with \( \mu > 0 \),

\[
n^* = \frac{d \log \frac{p}{\mu}}{\log 2 + H_2(\mu)} + o(1). \tag{20}
\]

Furthermore, it was proved in [19], that when the channel law is known, separate decoding of items achieves exact recovery with vanishing error probability provided that

\[
n \geq \frac{\log p}{I(X_1; Y)} + o(1). \tag{21}
\]

As noted in [15], in the noiseless setting with \( \nu = \log 2 \) we have \( I(X_1; Y) = \frac{\log^2 2}{a} (1 + o(1)) \), and thus in this case (21) matches (20) up to a constant factor of \( \log 2 \approx 0.7 \). Actually, the same is true for the symmetrical noise model. We next show that these results are true also when the channel law is unknown and the decoder in (11) is used.

Corollary 2. For the noisy group testing problem, if \( d = \Theta(1) \), then \( P_e \to 0 \) using (11, provided that

\[
n \geq \frac{\log p}{I(X_1; Y)} + o(1) \tag{22}
\]

for any \( \nu > 0 \).

Proof of Corollary 2: The statement follows immediately by noticing that when \( d = \Theta(1) \), the condition in (14a) dominates. To see this, note that the r.h.s. of (19) is a constant while the r.h.s. of (18a) grows with \( p \). Hence we get the condition in (22) with \( \delta_2 \to 0 \) sufficiently slowly.

Comparing Corollary 2 to (21) and [18, Corollary 6], we can conclude that our decoder is robust in the sense that it achieves exactly the same performance as the decoder that knows the channel.

Lemma 1. The following hold:

1) For any \( \delta_1 \in (0, 1) \),

\[
P\left[ i(X_{1:1}, Y) \leq (1 - \delta_1) \cdot \mathbb{E} i(X_{1:1}, Y) \right] \leq 6 \cdot \exp \left( -n \cdot \frac{\left| \delta_1 \cdot I(X_1; Y) \right|^2}{2 \log^2 n} \right). \tag{15}\]

2) For any \( \rho \in \mathbb{R}_+ \), define

\[
h_\rho(P_{XY}) \triangleq \min_{Q_{XY}} \left[ \rho \cdot I(Q(X); Y) + D_K L(Q_{XY} \| P_{XY}) \right] \geq \min_{Q_X} \left[ D_K L(Q_X \| P_X) + \rho \cdot I_{\mathbb{E}}(Q_X; P_Y | X) \right] \tag{16}\]

where \( I_\alpha(X; Y) \) is defined in (1). Then, for any \( \delta_1 \in (0, 1) \), and \( \rho \in \mathbb{R}_+ \),

\[
P\left[ i(X_{1:1}, Y) \leq (1 - \delta_1) \cdot \mathbb{E} i(X_{1:1}, Y) \right] \leq n^{3/2} e^{-n [h_\rho(P_{XY}) - (1 - \delta_1) \rho \cdot I_{\mathbb{E}}(X_1; Y) - \Theta(1/\delta_2)]}. \tag{17}\]

We can now use Lemma 1 to get an explicit formula for \( \psi_n(\cdot) \), which in turn will imply some condition on the sample complexity \( n \) through condition (14b). Combining Theorem 1, Theorem 2, and Lemma 1, we obtain the following result.

Corollary 1. For the noisy group testing problem, we can achieve \( P_e \to 0 \) using (11) provided that

\[
n \geq \frac{\log \left[ \frac{1}{\delta_2} (p - d) \right]}{(1 - \delta_1) \cdot I(X_1; Y)} (1 + \eta), \tag{18a}\]

\[
n \geq \min_{\rho \geq 0} \frac{\log d}{h_\rho(P_{XY}) - (1 - \delta_1) \cdot \rho \cdot I_{\mathbb{E}}(X_1; Y)} (1 + \eta), \tag{18b}\]

for some \( \delta_1, \eta > 0 \), and \( \delta_2 \to 0 \). Alternatively, (18b) can be replaced with

\[
n \log \frac{n}{\log n} \geq \frac{\log d}{1/2 \cdot \left| \delta_1 \cdot I(X_1; Y) \right|^2} (1 + \eta). \tag{19}\]
2) The $\theta \in (0,1)$ case: For the noiseless and symmetric noise models it was shown in [18] that it is sufficient and necessary to have

$$n \geq \frac{\log p}{I(X_1;Y)} [1 + o(1)]$$

in order to get $P_e \to 0$ when $d = \Theta(p^\theta)$ with $\theta$ sufficiently small. We next show that this is true also in our setting, namely, that in this case, (14a) dominates and thus we achieve exactly the same condition as in the case where the channel law is known. Accordingly, similarly as in the previous subsection also when $\theta$ is sufficiently small our separate decoder matched the performance of the optimal ML decoder up to a log 2 factor.

**Corollary 3.** For the noiseless and symmetric noise group testing models, with $\nu = \nu_{sym}$, if $d = \Theta(p^\theta)$ and $\theta \to 0$, then $P_e \to 0$ using (11), provided that

$$n \geq \frac{\log p}{I(X_1;Y)} [1 + o(1)]. \quad (23)$$

Finally, for any $\theta \in (0,1)$, we can show that (11) achieves order-optimal performance.

**Corollary 4.** For the noiseless and symmetric noise group testing models, with $\nu = \nu_{sym}$, if $d = \Theta(p^\theta)$ with $\theta \in (0,1)$, then $P_e \to 0$ using (11), provided that

$$n \geq \Omega(d \log p). \quad (24)$$

**V. JOINT DECODING**

We start with the following result which is parallel to Theorem 2.

**Theorem 3.** Consider the a general group testing with Bernoulli($\nu/d$) testing and the decoder in (10), and suppose that there exist some functions $\psi_n(\delta_{1,\ell})$, $\ell \in \{1, \ldots, d\}$ such that

$$F \left[ i_n(X_{s_{dif}}; Y|X_{s_{eq}}) \leq (1 - \delta_{1,|s_{dif}|}) \cdot \mathbb{E}i_n(X_{s_{dif}}; Y|X_{s_{eq}}) \right] \leq \psi_n(\delta_{1,|s_{dif}|}), \quad (25)$$

for some constants $\{\delta_{1,\ell}\}_{\ell=1}^d \in (0,1)^d$. Also, suppose that

$$n \geq \max_{|s_{dif}|,|s_{eq}|} \max_{|s_{dif}|,|s_{eq}| \neq \emptyset } (1 - \delta_{1,|s_{dif}|}) \cdot I(X_{s_{dif}}; Y|X_{s_{eq}}), \quad (26a)$$

$$\sum_{\ell=1}^d \delta_{1,\ell} \cdot \psi_n(\delta_{1,\ell}) \to 0, \quad (26b)$$

for some $\delta_{1,\ell} \to 0$. Then, $P_e \to 0$, using $\gamma_{|s_{dif}|} = \log \left[ \frac{d}{\delta_2} \left( \frac{p-d}{|s_{dif}|} \right)^\left( \frac{d}{|s_{dif}|} \right) \right]$ in (10).

As in the previous section, when $d = \Theta(1)$, or, $d = \Theta(\theta^p)$ with $\theta$ sufficiently small, the condition in (26a) dominates. In fact, it was also shown in [16, 17] that (26a) is also a necessary condition when the channel is known, which implies that our decoder is universal.

**Theorem 4.** For the noisy group testing problem, if $d = \Theta(1)$, or, $d = \Theta(\theta^p)$ with $\theta \to 0$, then $P_e \to 0$ using (10), provided that

$$n \geq \max_{|s_{dif}|,|s_{eq}|} \max_{|s_{dif}|,|s_{eq}| \neq \emptyset } (1 - \delta_{1,|s_{dif}|}) \cdot I(X_{s_{dif}}; Y|X_{s_{eq}}). \quad (27)$$

for any $\nu > 0$.

**REFERENCES**


