Near-Optimal Stochastic Threshold Group Testing

C. L. Chan∗, S. Cai∗, M. Bakshi∗, S. Jaggi∗, and V. Saligrama†
The Chinese University of Hong Kong∗, Boston University†

Abstract—We formulate and analyze a stochastic threshold group testing problem motivated by biological applications. Here a set of \( n \) items contains a subset of \( d \ll n \) defective items. Subsets (pools) of the \( n \) items are tested – the test outcomes are negative, positive, or stochastic (negative or positive with certain probabilities that might depend on the number of defectives being tested in the pool), depending on whether the number of defective items in the pool being tested are fewer than the lower threshold \( l \), greater than the upper threshold \( u \), or in between. The goal of a stochastic threshold group testing scheme is to identify the set of \( d \) defective items via a "small" number of such tests. In the regime that \( l = o(d) \) we present schemes that are computationally feasible to design and implement, and require near-optimal number of tests (significantly improving on existing schemes). Our schemes are robust to a variety of models for probabilistic threshold group testing.

I. INTRODUCTION

Classical Group Testing: The set \( N \) of \( n \) items contains a set \( D \) of \( d \) "defectives" – here \( d \) is assumed to be \( o(n) \). The classical version of the group-testing problem was first considered by Dorfman in 1943 [1] as a means of identifying a small number of diseased individuals from a large population via as few “pooled tests” as possible. In this scenario, blood from a subset of individuals is pooled together and tested – if none of the individuals being tested in a pool have the disease the test outcome is “negative”, else it is “positive”. In the non-adaptive group testing problem, each test is designed independently of the outcome of any other test, whereas for adaptive group-testing problems, the testing procedure may be conducted sequentially. For both problems, \( O(d \log(n)) \) tests are known to be necessary and sufficient – a good survey of some of the algorithms and bounds can be found in the books by Du and Hwang [2, 3] and the paper by Chen and Hwang [4].

Threshold Group Testing: In this work we focus on a generalization of the classical group testing problem called threshold group testing, first considered by Damaschke [5]. The difference is that the outcome of each pooled test is “positive” if the number of defectives in the test is no smaller than the upper threshold (denoted \( u \)), is “negative” if no larger than the lower threshold (denoted \( l \)) defectives were contained in the test, and otherwise it is arbitrary (“worst-case”). Clearly, when \( u = 1 \) and \( l = 0 \), this reduces to the classical group testing problem. There are other generalizations of classical group testing [6, 7, 8]. Applications of the threshold group testing model include the problem of reconstructing a hidden hypergraph [9, 10, 11, 12], and a searching problem called “guessing secrets” [5, 13].

The first adaptive algorithm for threshold group testing was proposed in [5]. When the gap \( g \) (defined as \( u - l - 1 \), the difference between the upper and lower thresholds) equals 0, the number of tests in [5] for identification of the set of defectives is \( O((d + u^2) \log n) \). When the gap \( g \neq 0 \), the number of tests required by [5] scales as \( O(d n^b + d^e) \), if \( g + (u - 1)/b \) misclassifications are allowed (here \( b > 0 \) is an arbitrary constant), with polynomial-time decoding complexity. The work of [12] showed that \( O(ed^{u+1} \log(n/d)) \) non-adaptive threshold tests suffice to identify the set of defectives with up to \( g \) misclassifications and \( e \) erroneous tests allowed. The computational complexity of decoding is \( O(n \log n) \) for fixed \((d, e)\). In [9], instead of the strongly disjoint matrices used in [12], a probabilistic construction of a weaker version of disjoint matrices is used to reduce the number of tests from \( O(d^{a+1} \log(n/d)) \) to \( O(d^{a+2} \log d \log(n/d)) \). Also, two explicit constructions with number of tests equaling \( O(d^{b+3} \log(d) \quad \text{quasipoly}(\log(n))) \) and \( O(d^{b+3+\beta} \quad \text{poly}(\log(n))) \) (for arbitrary \( \beta > 0 \)) are proposed. However, the computational complexity of decoding is not addressed. Also, [14] draws a connection between “threshold codes”, non-adaptive threshold group testing, and a model called “majority group testing”.

(A) Worst-case Model: If the number of defective items in a pool is between the upper and lower thresholds (“in the gap”), then the test outcome is assumed to be arbitrary. Algorithms must therefore be designed to account for a malicious adversary that can set test outcomes to maximally confuse the threshold group testing scheme.

(B) Zero-error (with misclassifications): The algorithm is required to guarantee (with probability 1), that the output is “correct” (it contains the set of defective items, up to a certain number of misclassifications). (A fundamental consequence of these two models assumptions is that if the gap \( g = u - l - 1 > 0 \), the set of defectives cannot be exactly identified – regardless of what algorithm is used, one can only reconstruct the set of defective items up to a certain number of misclassifications [5]).

Stochastic Threshold Group Testing: We relax these aspects of the conventional setting. In particular, we relax the worst-case model to a stochastic model. We seek probabilistic guarantees instead of the absolute zero error guarantees.

(A) Stochastic Model: This setup is motivated by a class of biological applications [15] where the test outcomes are observed to be random whenever the number of defectives in a pool falls within a given range. We consider two models. For the first model, we assume that the outcome of a test is equally likely to be positive or negative whenever the number of
defectives in a pool is in the range \((l, u)\). In our second model, the probability of a test outcome being positive depends on the number of defective items in the test, and for concreteness, we assume that this dependence scales linearly from \(l\) to \(u\) (though our results hold for more general models as well\(^1\)). These two models are represented in Figures 1 and 2.

**B) Probabilistic Guarantee:** We allow for a “small” probability of error for our algorithm, where this probability is both with respect to the randomness of the measurements within the gap, and the test design.

These “natural” information-theoretic relaxations in the model result in schemes that have significantly improved performance, compared to prior work. In particular, our schemes require far fewer tests than prior algorithms, and also admit computationally efficient decoding schemes. They also directly lend themselves to scenarios with zero gaps, and also to other models similar to group-testing, such as the Semi-Quantitative Group Testing [3].

For the stochastic threshold group-testing problem we present three algorithms (TGT-BERN-NONA, TGT-BERN-ADA, and TGT-LIN-NONA, respectively for the non-adaptive problem with Bernoulli gap stochasticity, adaptive problem with Bernoulli gap stochasticity, and non-adaptive problem with linear gap stochasticity). Our results are summarized as follows.

**Theorem 1:** (Non-adaptive algorithm with Bernoulli gap model) For \(l = o(d)\), TGT-BERN-NONA with error probability at most \(\epsilon\) requires \((4e^\epsilon \ln(2)/\pi^2)\ln(1/\epsilon)\sqrt{d} \ln(n) + O(\ln(1/\epsilon) d \sqrt{l})\) tests and computational complexity of decoding \(O(n \ln(n) + n \ln(1/\epsilon))\).

**Theorem 2:** (Two-stage Adaptive algorithm) For \(l = o(d)\), TGT-BERN-ADA with error probability at most \(\epsilon\) requires \(16e^2 d \ln(n) + O(\ln(1/\epsilon) d)\) tests and computational complexity of decoding \(O(n \ln(n) + n \ln(1/\epsilon))\).

**Theorem 3:** (Non-adaptive algorithm with linear gap model) TGT-LIN-NONA with error probability at most \(\epsilon\) requires \(O(g^2d \ln(n)) + O(\ln(1/\epsilon) d)\) tests and computational complexity of decoding \(O(g^2n \ln(n) + n \ln(1/\epsilon))\).

**Remark:** Note that the number of tests required by our algorithms are, in general, much smaller than those required by prior works – this demonstrates the power of using the stochasticity that may naturally be inherent in the measurement model.

### II. INTUITION

To build intuition into our proof techniques consider the Bernoulli Stochastic Model described in Sec. 1 and Fig. 1. We note the discrete transition in terms of distribution of test outcomes for pools consisting of \(l + 1\) defectives relative to those that contain \(l\) defectives. If negative outcomes are labeled zero and positive outcomes labeled one, the test outcomes are identically zero for pools containing exactly \(l\) defectives. So the distribution of test outcomes is concentrated at zero. For pools containing \(l + 1\) defectives the distribution of test outcomes is split equally at zero and one. We can exploit this aspect of the model in the following way. Suppose we had a pool, \(R^+\) consisting of exactly \(l\) defectives then one could test whether or not an item, \(x_j \notin R^+\) is defective by augmenting \(R^+\) with \(x_j\) and testing the new pool \(R^+ \cup \{x_j\}\).

To exploit this idea we have to account for several issues. First, we do not really have a candidate pool \(R^+\). Second, with this naive strategy, the number of tests would grow with the number of items even when we have a candidate pool \(R^+\) consisting of \(l\) defectives.

To address these requirements we construct two distinctive collections of pools based on random designs. The reference group collection, \(R\), is a collection of \(R\) pools, each with \(nl/d\) items, such that at least one among the \(R\) pools has exactly \(l\) defective items in it. The idea is that with high probability one among the \(R\) pools contains the critical candidate \(R^+\). The second collection, the transversal design, is a family \(\mathcal{I}\) of sub-collections of size \(I\). Each sub-collection within this family consists of \(O(d)\) disjoint pools indexed as \(I_{i,k}\).

---

\(^1\)In fact, as long as there is a statistical difference between the probability of a positive test outcome when the number of defective items is within the range \((l, u)\), and outside this range, our approach works. Due to space limitations, in this work we focus on the two models in Figures 1 and 2.
In this example, of items denoted by its connected nodes are measured. A solid edge indicates the top nodes by an edge whose type indicates the test outcome when the union families. The bottom nodes represent reference groups picked uniformly at random. (4) Independently of all prior choices, that a “critical” reference group (with exactly 7 positive). At the other extreme, more than \(3\) defective items (in this example, only 2) is a code-design parameter whose value is specified later). (5) The complete, \(\rho\) and \(N\), \(\beta\) is partitioned uniformly at random into indicator groups of size \(\gamma_2 n/(d-1)\) each (here \(\gamma_2\) is a code-design parameter whose value is specified later). (5) The complete bipartite graph shows the “cross-product” design of threshold tests of the form \(\mathcal{R} \times \mathcal{I}\), in which left (blue) nodes represent reference groups, right (red) nodes represent indicator groups, and edges represent threshold tests corresponding to the union of items denoted by the connected nodes (one reference group with one indicator group).

Each disjoint pool within a sub-collection is referred to as an indicator group. Consequently, each item appears only once within any sub-collection and \(I\) times within the entire family. The indicator group collection serves the role of an item \(\{x_j\}\) described in the preceding paragraphs.

Our algorithm is based on augmenting each indicator group within a sub-collection with the indicator group collections to form pools that are then tested. The idea of transversal design is not new and has been used before in conventional group testing as well. The novelty here is the cross-product, namely, testing indicator groups against a reference group collection resulting in \(\mathcal{R} \times \mathcal{I}\) pools. The question arises as to how to construct these collections and how to find defectives given the test outcomes.

To construct \(\mathcal{R}\) we begin by noting that if one chooses a random group of size \(n l/d\), with probability about \(1/\sqrt{l}\) it has exactly \(l\) defective items in it. This is because of the fact that the expected number of defective items in such a group is exactly \(l\), and standard analysis using Stirling’s approximation of the hypergeometric distribution corresponding to the number of defective items in a group of size \(n l/d\) implies that the probability of hitting this expectation scales as \(1/\sqrt{l}\). This means that if one chooses about \(R = O(\sqrt{l})\) “candidate reference groups” \(\mathcal{R}_r\), each of size \(n l/d\), then “with high probability” at least one group \(\mathcal{R}^*\) will be “critical” (have exactly \(l\) defective items in it). To summarize, we select \(O(\sqrt{l})\) candidate reference groups of size \(n l/d\) each, and \(O(d \log(n))\) indicator groups of size about \(n/d\) each. We then perform threshold group-tests on every pair of the form \(\mathcal{R}_r \cup \mathcal{I}\), for a total of about \(O(\sqrt{l} d \log(n))\) (non-adaptive) tests.

Our decoding algorithm hinges on identifying the critical candidate(s) \(\mathcal{R}^*\). To do so we make use of the statistical difference between reference groups that are critical, and those that are not. When \(\mathcal{R}\) and a randomly-picked indicator group are
Hence one does not need in a group falling within this range with high probability.

Encoder/Testing scheme:

(Instead of demanding that such critical groups contain exactly such defectives and tests with this threshold group tests

\[ R \text{ is an increasing function of the number of defective items in } R \text{ and then use only } \rho \text{ for the first stage tests required. The idea is to use the first stage} \]

if one is allowed to perform threshold group tests

\[ (0) \leq \rho \leq \theta_i \leq \theta_{i1} \]

and if it contains \( v \) defective items, \( \alpha_w = \text{Pr} (y_{p,r,i} = 1 | R_{p,r} \cap D = l, x_j = w) \). Note that \( x_j \) is a binary indicator variable taking values 1 or 0 depending on whether the item is defective or not. For every critical \( R_{p,r} \subseteq \rho \text{ and } x_j \in \rho \) if \( \sum_{y_{p,r,i}} \leq |I|\alpha_0(1 + \Delta) \) where \( \Delta = (\alpha_1 - \alpha_0)/(2\alpha_0) \), the decoder declares \( x_j \) to be non-defective, else declares it to be defective. That is, the decoder declares an item to be non-defective if the empirically observed fraction of positive test outcomes involving that item is “close to” the “expected value” \( (\theta_i) \). The probability of this event is calculated in Lemma 6.

2) For each \( i = 1, 2, \cdots, I \), let \( I_{i}^{(0)} \) be a randomly picked indicator group from \( \{ I_{i} \} \), and let \( I_{i}^{(0)} = \{ I_{i}^{(0)} : i = 1, 2, \cdots, I \} \). Let \( y_{p,r,i} \) be the test outcome when the group \( R_{p,r} \cap D = \theta_i \), \( \alpha_w = \text{Pr} (y_{p,r,i} = 1 | R_{p,r} \cap D = l, x_j = w) \). Note that \( x_j \) is a binary indicator variable taking values 1 or 0 depending on whether the item is defective or not. For every critical \( R_{p,r} \subseteq \rho \text{ and } x_j \in \rho \) if \( \sum_{y_{p,r,i}} \leq |I|\alpha_0(1 + \Delta) \) where \( \Delta = (\alpha_1 - \alpha_0)/(2\alpha_0) \), the decoder declares \( x_j \) to be non-defective, else declares it to be defective. That is, the decoder declares an item to be non-defective if the empirically observed fraction of positive test outcomes involving that item is “close to” the “expected value” \( (\theta_i) \). The probability of this event is calculated in Lemma 6.

IV. PROOF OF THEOREM 1

**Definition 1:** Hypergeometric distribution describes the probability of picking \( v \) defective items when we pick \( s \) distinct items from \( n \) items with \( d \) defectives. The probability mass function is given by

\[
\text{Pr}(v, s, n, d) = \binom{d}{v} \binom{n-d}{s-v} / \binom{n}{s}
\]
Lemma 4: The probability of picking \( l \) defective items when we sample \( (nl)/d \) items from \( n \) items with \( d \) defective items is \( \Omega \left( \frac{1}{\sqrt{n}} \right) \).

Probability of picking \( l \) defective items
\[
O \left( \frac{1}{\sqrt{n}} \right)
\]

Fig. 6. Probability of sampling \( l \) defective items as a function of the size of the group undergoing a threshold test.

Proof: The probability that a test of a certain size has exactly \( l \) defective items scales according to the hypergeometric distribution given in Definition [1]. When the number of items in the test equals \( nl/d \), this probability can be shown via Stirling’s approximation [17] that for all \( n \in \mathbb{N}^+ \), \( 1 \leq n/(2\pi n)^{1/2} (e/n)^{-n} \leq e/(2\pi)^{1/2} \), and therefore implying for all \( k \in \mathbb{N}^+, k < n, \sqrt{2\pi/e} \leq \left( e/2\pi \right)^{k(n-k)/n^2} \leq e/(2\pi) \) to scale as
\[
\left( \frac{n}{2} \right)^{1/2} \geq \frac{4\pi^2}{e^\delta} \sqrt{d-l} \sqrt{n/d-l} \sqrt{n-d/n-d}.
\]

Note that the exponential terms from the Stirling’s approximation of the binomial coefficients are exactly cancelled out in [1].

Lemma 5: With probability at least \( 1 - \epsilon_2 \), for each \( \rho \in \{1, \ldots, P\} \), every \( \mathcal{P}_\rho \) has at least one critical reference group when
\[
R > \left( \ln \left( \frac{1}{\epsilon_2} \right) + \ln \left( \frac{2d}{d-l} \right) \right) \frac{e^{\delta}}{4\pi^2} \sqrt{\frac{d-l}{d-l}} \sqrt{\frac{n-d}{n-d}}.
\]

Proof: Let \( E_1(\rho) \) be the event that a specific \( \mathcal{P}_\rho \) has “too many” defective items, i.e., \( \| \mathcal{P}_\rho \cap D \| > (1 + \delta)\gamma(\rho) d \) for any particular \( \rho \). Let \( E_1 \) be the event corresponding to the union of \( E_1(\rho) \), i.e., that at least one division has too many defective items. Let \( E_2 \) be the event that there exists a \( \mathcal{P}_\rho \) which contains no critical reference group. First, we compute the union bound of probability of \( E_1 \) for all \( \rho \) as
\[
P(\Pr(E_1) < \frac{4d}{d-l} \exp \left( \frac{-2(n+2)((\delta\gamma^2)(d-1)/(n-d+1)(d+1))}{(n-d+1)(d+1)} \right)) \leq 4d \exp \left( -\delta \gamma^2 (d-2) \right)
\]

Inequality (2) follows from Hush’s bound [18]. Note that \( \gamma_1 = (d+1)/(2d) \). When \( \delta \) is greater than \( \sqrt{(4d)/(d+1)^2} \ln ((4d)/(d-1)) \), it is bounded from above by a constant \( \epsilon_1 \).

Let \( \Pr(v) = \Pr(|\mathcal{P}_\rho \cap D| = 1 \mid |\mathcal{P}_\rho \cap D| = v) \), i.e. the probability that we pick a critical reference from \( \mathcal{P}_\rho \) given \( \mathcal{P}_\rho \) contains \( v \) defective items. Let \( \Pr(v) = \Pr(|\mathcal{P}_\rho \cap D| = v) \), i.e. the probability that we pick a critical reference from \( \mathcal{P}_\rho \). We wish to compute \( \Pr_1((1 + \delta)\gamma_1d) \) and \( \Pr_1((1 - \delta)\gamma_1d) \). The ratio of \( \Pr_1(\gamma_1d) \) to \( \Pr_1((1 + \delta)\gamma_1d) \) can be computed as
\[
\frac{\Pr_1((1 + \delta)\gamma_1d)}{\Pr_1((1 + \delta)\gamma_1d)} = \frac{\gamma_1d - \gamma_1d}{(1 + \delta)\gamma_1d - \gamma_1d} \left( \frac{n}{n-d} \right) \left( \frac{n}{n-d} \right).
\]

Equality (4) follows by noting that \( \Pr_1(v) \) follows hypergeometric distribution. Inequality (5) follows from \( (1 + x)^y \leq e^{xy} \) for \( |x| < 1 \) and \( y \geq 0 \). For \( \delta = o(1/\sqrt{d}) \) and \( l = o(d) \), (5) is bounded from above by \( e \). The same technique applied to the ratio of \( \Pr_1(\gamma_1d) \) to \( \Pr_1((1 - \delta)\gamma_1d) \) implies that it is less than \( e \). These bounds on these ratios, together with Lemma [4] gives us that
\[
\Pr_1((1 + \delta)\gamma_1d) < \frac{\Pr_1((1 + \delta)\gamma_1d)}{\Pr_1((1 + \delta)\gamma_1d)} \leq \frac{\epsilon_1}{4d \exp \left( -\delta \gamma^2 (d-2) \right)}
\]

Finally, we bound the probability of \( E_2 \) (note that \( E_2 \) is defined in the beginning of the proof) occurring as
\[
\Pr(E_2) < P \sum_{v=0}^{d} \Pr(v) (1 - \Pr_1(v))^R
\]

We now substitute equations (3) and (4) into (7). Note that for large enough \( d \), the constant \( \epsilon_1 \) (bounded from above by the quantity in (3)) can be made arbitrarily small. Hence, if we bound (7) from above by constant \( \epsilon_2 \), for large enough \( d \) and \( R \), (7) can be made smaller than any \( \epsilon_2 \), and we obtain Lemma [5].

Lemma 6: With probability at least \( 1 - \epsilon_3 \), the decoder correctly determines whether a given \( \mathcal{R}_{\rho,\tau} \) is critical or not when
\[
I > \sqrt{8\epsilon_2^2 \left( \ln(RP) + \ln \left( \frac{1}{\epsilon_3} \right) \right)}
\]

Proof: We have three type of reference groups. We call a reference group promising if it contains at most \( l-1 \) defective items, and call it misleading if it contains at least \( l+1 \) defective items.
items. Finally, recall that a reference group is critical if it contains exactly \( l \) defective items. The error events include four kinds of misclassification. We denote the probability of misclassifying a promising \( \mathcal{R}_{p,r} \) to be critical by \( \Pr_{e}^{p,c} \), the probability of misclassifying a critical \( \mathcal{R}_{p,r} \) to be promising by \( \Pr_{e}^{p,p} \), the probability of misclassifying a misleading \( \mathcal{R}_{p,r} \) to be critical by \( \Pr_{e}^{m,c} \), and the probability of misclassifying a critical \( \mathcal{R}_{p,r} \) to be misleading by \( \Pr_{e}^{m,m} \). Each of the error probabilities can be bounded by a binomial distribution. The first error event can be computed as

\[
\Pr_{e}^{p,c} \leq \Pr \left( \bigcup_{v=0}^{l} \sum_{i=0}^{l} g_{v,i}^{(0)} \geq |\mathcal{I}(0)| \theta_{l}(1 - \Delta_{c}) \right) \\
\leq RP \sum_{t=0}^{l} \frac{l}{t} \theta_{l-1}(1 - \theta_{l-1})^{l-t} \\
\leq RP \exp \left( -2I(\theta_{l}(1 - \Delta_{c}) - \theta_{l-1}^{2}) \right). 
\]

In inequality (8), we take union bound over all \( \mathcal{R}_{p,r} \), and the summation is over the tail of a binomial distribution (corresponding to the event that a promising reference group “behaves like” a critical reference group). Inequality (9) follows from the Chernoff bound. Similarly, for the other error events we have that

\[
\Pr_{e}^{p,p} \leq RP \exp \left( -2I(\theta_{l}(1 - \Delta_{c})^{2}) \right), \\
\Pr_{e}^{m,c} \leq RP \exp \left( -2I(\theta_{l+1} - \theta_{l}(1 + \Delta_{c}))^{2} \right), \\
\Pr_{e}^{m,m} \leq RP \exp \left( -2I(\theta_{l}(1 - \Delta_{c})^{2}) \right). 
\]

Within the valid range of \( \Delta_{c} \), \( \Pr_{e}^{p,c} \) is strictly increasing as a function of \( \Delta_{c} \); conversely, \( \Pr_{e}^{p,p} \) is strictly increasing as a function of \( \Delta_{c} \); \( \Delta_{c} \) is one that allows for a “small” choice of \( l \), while still keeping both \( \Pr_{e}^{p,c} \) and \( \Pr_{e}^{p,p} \) “small”. The same argument holds for \( \Delta_{l} \) to \( \Pr_{e}^{m,c} \) and \( \Pr_{e}^{m,m} \). Some specific choices of \( \Delta_{c} \) and \( \Delta_{l} \) that work, and that we use, are

\[
\Delta_{c} = (\theta_{l} - \theta_{l-1})/(2\theta_{l}), \\
\Delta_{l} = (\theta_{l+1} - \theta_{l})/(2\theta_{l}), 
\]

so that (10) balances (11), and (12) balances (13).

Let \( \Pr_{v|w} = \Pr \left( \left( \mathcal{I}_{(0)} \setminus \mathcal{R}_{p,r} \right) \cap D \mid v \right) \). (This is the probability that, conditioned on the reference group \( \mathcal{R}_{p,r} \) containing \( v \) items, a randomly chosen indicator set from the \( i \)-th family \( \mathcal{I}_{(0)} \) contains exactly \( w \) defective items that are not contained in the reference group \( \mathcal{R}_{p,r} \)).

Hence the conditional probabilities of giving a positive outcome can be expanded as

\[
\theta_{l-1} = \frac{1}{2} \left( 1 - \Pr_{v|l-1}^{(0)} - \Pr_{v|l-1}^{(0)} + \sum_{v=g}^{d} \Pr_{v|l-1}^{(0)} \right) \\
\theta_{l} = \frac{1}{2} \left( 1 - \Pr_{v|l}^{(0)} + \sum_{v=g}^{d} \Pr_{v|l}^{(0)} \right) \\
\theta_{l+1} = \frac{1}{2} \left( 1 + \sum_{v=g}^{d} \Pr_{v|l+1}^{(0)} \right). 
\]

Since the summations in each equation (15)-(17) are “close” to each other, we ignore them in the following calculations (since only their pairwise differences required, and the summations only contribute lower-order terms). For example, the difference between \( \Pr_{v|l}^{(0)} \) and \( \Pr_{v|l-1}^{(0)} \) can be computed as

\[
\Pr_{v|l}^{(0)} - \Pr_{v|l-1}^{(0)} = \Pr_{v|l}^{(0)} \left( 1 - \frac{\Pr_{v|l-1}^{(0)}}{\Pr_{v|l}^{(0)}} \right) \\
\leq \Pr_{v|l}^{(0)} \left( \frac{v}{d-l+v} \right) 
\]

When \( v = o(d) \), (18) is bounded from above by \( v/(d-l+v) \), which is asymptotically negligible as \( d \) grows without bound.

The quantity \( 2\theta_{l} \Delta_{c} \) in (12) is then bounded by using (14), (16) and (17), and noting that

\[
4\theta_{l} \Delta_{c} = 2(\theta_{l+1} - \theta_{l}) \\
= \Pr_{v|l}^{(0)} \\
= \prod_{i=0}^{d} \left( 1 - \frac{d-l}{n-i} \right) \\
> \left( 1 - \frac{d-l}{n - \frac{\gamma n}{d-1}} \right)^{\frac{\gamma n}{d-1}} \\
\geq \exp \left( -\frac{\gamma n}{n - \frac{\gamma n}{d-1} - d-l} \right) 
\]

Inequality (19) follows from the fact that \( (1+x)^{-y} \geq e^{-y} \) for \( |x| < 1 \) and \( y \geq 0 \).

The quantity \( 2\theta_{l} \Delta_{l} \) in (10) can be bounded in a similar manner as

\[
4\theta_{l} \Delta_{l} = 2(\theta_{l} - \theta_{l-1}) \\
> \Pr_{v|l-1}^{(0)} \\
= \left( \frac{d-l-1}{d-l} \right)^{\frac{n}{d-1}} \\
= \gamma 2 d - l + 1 - \frac{\gamma n}{d-1} \prod_{i=0}^{d} \left( 1 - \frac{d-l}{n-i} \right) \\
> \gamma 2 \left( 1 - \frac{d-l}{n - \frac{\gamma n}{d-1} + 1} \right)^{\frac{\gamma n}{d-1}-1} \\
\geq \gamma 2 \exp \left( -\frac{\gamma n}{n - \frac{\gamma n}{d-1} + d-l} \right). 
\]

Finally we substitute the results from (13)-(20) into (9)-(12). The requirement that error probability of misclassification of any reference groups be at most \( \epsilon_{3} \) implies

\[
I > 8 \max \left( \left( \Pr_{v|l}^{(0)} \right)^{-2}, \left( \Pr_{v|l-1}^{(0)} \right)^{-2} \right) \left( \ln(RP) + \ln \left( \frac{1}{\epsilon_{3}} \right) \right). 
\]

For \( d = o(n) \) and “large enough” \( n \), (19) “behaves” as \( \gamma_{2} e^{-\epsilon_{2}} \), and (20) “behaves” as \( \gamma_{2} e^{-\epsilon_{2}} \). The quantity \( I \) is minimized when \( \gamma_{2} \) is 1. Therefore we obtain the result in Lemma [4].
Lemma 7: With error probability at most $\epsilon_4$, the decoder correctly determines whether an item $x_j$ is defective or non-defective when

$$I > 8e^2 \left( \ln(n) + \ln \left( \frac{1}{\epsilon_4} \right) \right).$$

Proof: As the rule for deciding whether an item is defective is not, and also the rule for deciding whether a reference group is critical or not, both depend on matching empirically observed test outcome statistics with precomputed thresholds, the proof here is essentially the same as in the proof of Lemma 6. We outline the major changes below.

The error event includes both false positives (misclassifying non-defective items to be defective) and false negatives (misclassifying defective items to be non-defective). The probability of false negatives can be computed as

$$P_e^- = \Pr \left( \bigcup_{j:x_j=1} \sum_i y_{r,i}^{(j)} \leq |\mathcal{I}^{(j)}| \alpha_0 (1 + \Delta) \right) \leq n \exp \left( -2I(\alpha_1 - \alpha_0 (1 - \Delta))^2 \right). \tag{21}$$

In a similar manner, the probability of false positives can be computed as

$$P_e^+ = \Pr \left( \bigcup_{j:x_j=0} \sum_i y_{r,i}^{(j)} > |\mathcal{I}^{(j)}| \alpha_0 (1 + \Delta) \right) \leq n \exp \left( -2I(\alpha_0 \Delta)^2 \right). \tag{22}$$

Let $\Pr_{v|\text{ind}}^{(j)} = \Pr(|\mathcal{I}^{(j)}_v \cap \mathcal{D}| = v \mid x_j = 0)$ and $\Pr_{v|\text{d}}^{(j)} = \Pr(|\mathcal{I}^{(j)}_v \cap \mathcal{D}| = v \mid x_j = 1)$. A good choice of $\Delta$ is

$$\Delta = (\alpha_1 - \alpha_0)/(2\alpha_0). \tag{23}$$

We may expand $\alpha_0$ and $\alpha_1$ in terms of $\Pr_{v|\text{ind}}^{(j)}$ and $\Pr_{v|\text{d}}^{(j)}$ as

$$\alpha_0 = \frac{1}{2} \left( 1 - \Pr_{0|\text{ind}}^{(j)} + \sum_{v=0}^d \Pr_{v|\text{ind}}^{(j)} \right),$$

$$\alpha_1 = \frac{1}{2} \left( 1 + d \Pr_{v|\text{d}}^{(j)} \right).$$

The difference between $\alpha_0$ and $\alpha_1$ can be computed as

$$2(\alpha_1 - \alpha_0) > \Pr_{0|\text{ind}}^{(j)} - \sum_{v=0}^d \Pr_{v|\text{ind}}^{(j)} \tag{24}$$

$$\geq \prod_{t=0}^{\frac{\gamma_2 n - d + l}{\frac{\gamma_2 n - d + l}{n - \frac{\gamma_2 n}{d}}}} \left( 1 - \frac{d - l}{n - \frac{\gamma_2 n}{d} - 1} \right)^{\frac{\gamma_2 n - d + l}{n - \frac{\gamma_2 n}{d} - 1}} \geq \exp \left( -\gamma_2 n - d + l \right) \tag{25}$$

Equality (24) ignores the small difference came from the summation terms. Inequality (25) follows from the fact that $(1 + x)^y \geq e^{-xy}$ for $|x| < 1$ and $y \geq 0$.

Finally we substitute (21) and (22) into (21) and (22), and $\gamma_2$ is set to be 1 in Lemma 6. The requirement that the error probability of misclassification of any item be at most $\epsilon_4$ implies the result in Lemma 7.

Proof of Theorem 1: A sufficient condition for high probability decoding all items is when Lemmas 5 and 7 are satisfied. Therefore, with error probability at most $\epsilon_2 + \epsilon_3 + \epsilon_4$, the total number of tests $T = R P(d - 1)I$, where $P = (2d)/(d - l)$, is specified in Lemma 5 and 1 required to satisfy both Lemma 6 and Lemma 7 is set as

$$I > 8e^2 \left( \ln(n) + \ln \left( \frac{1}{\epsilon_3} \right) + \ln \left( \frac{1}{\epsilon_4} \right) \right).$$

Explicitly, $T$ is at least $(4e^8 \ln(2)/\pi^2) \ln(1/\epsilon_2) \sqrt{d \ln(n)} + O(\ln(1/\epsilon_3 \epsilon_4) d \ln(T))$. As to the computational complexity of decoding, recall that the first decoding step decodes a reference group by counting the empirical fraction of positive outcomes from $I$ indicator groups, and the second decoding step decodes an item by doing the same thing. Therefore, given there are $PR$ reference groups and $n$ items, the complexity is $I(n + PR)$, which is $O(n \ln(n)) + O(\ln(1/\epsilon_2) \ln(n)) + O(n \ln(1/\epsilon_3 \epsilon_4))$. Let $\epsilon = \max(\epsilon_2, \epsilon_3, \epsilon_4)$, we obtain Theorem 1.

V. PROOF SKETCHES OF THEOREMS 2 AND 3

Slight modifications of TGT-BERN-NONA can result in an adaptive algorithm (as in Theorem 2), and also an algorithm for threshold group testing models where the probability of giving a positive outcome is a monotonically nondecreasing function of the number of defective items being measured. As a demonstration, we show a two-stage adaptive algorithm TGT-BERN-ADA, and a non-adaptive algorithm TGT-LINNONA that works under the “linear model” of stochastic threshold group testing (as in Figure 2).

A. TGT-BERN-ADA

In the first stage, we aim to find multiple critical reference groups. This is done by first performing the encoding step 1 of TGT-BERN-NONA, and then obtaining a set of indicator groups called $\mathcal{I}^{(1)}$. The construction of $\mathcal{I}^{(1)}$ is however
In this case, the notation for the total number of families in the set of indicator items is slightly different from the definition given in TGT-BERN-NONA. Here $\mathcal{I}^{(0)} = \{I_i^{(0)} : i = 1, 2, \cdots, I_1\}$, where each $I_i^{(0)}$ is a group of $\gamma_2 n/d$ distinct items randomly picked from $N$. For every pair of a reference group from $\mathcal{R}$ and an indicator group from $\mathcal{I}^{(0)}$, the two groups are pooled together and a threshold group test is performed. As to inference of whether particular reference groups are critical or not, this follows the decoding step 1 of TGT-BERN-NONA, but using the definition/parameters of $\mathcal{I}^{(0)}$ provided in this paragraph.

The second stage follows steps 2 and 3 of TGT-BERN-NONA. However, only the set of reference groups decoded to be critical in the first stage are tested in step 3, hence the multiplicative factor of $\sqrt{l}$ is missing from the overall number of tests. To avoid confusion with TGT-BERN-NONA, notation for the total number of families in the set of indicator groups (denoted $I$ in TGT-BERN-NONA) is replaced by $I_2$.

Finally, to decode whether individual items are defective or not, TGT-BERN-ADA uses decoding step 2 in TGT-BERN-NONA.

**Proof sketch of Theorem 2**: A sufficient condition for high probability decoding all items is when $R$ satisfies Lemma 5 $I_1$ satisfies Lemma 6 and $I_2$ satisfies Lemma 7. Therefore, with error probability at most $\epsilon_2 + \epsilon_3 + \epsilon_4$, the total number of tests $T$ is $RPI_2 + P(d - l)I_2$, which is at least \(16c^2d \ln(n) + O(\ln(1/\epsilon_2)\sqrt{l}\ln(l)) + O(\ln(1/\epsilon_3)\sqrt{l}) + O(\ln(1/\epsilon_4)\sqrt{l})\). The computational complexity of decoding is $nI_2 + PRI_1$, which is $O(n \ln(n)) + O(\ln(1/\epsilon_2)\sqrt{l}\ln(l)) + O(\ln(1/\epsilon_3)\sqrt{l}) + O(n\ln(1/\epsilon_4)n)$. Setting $\epsilon = \max(\epsilon_2, \epsilon_3, \epsilon_4)$, we obtain Theorem 2.

### B. TGT-LIN-NONA

The testing scheme follows that of TGT-BERN-NONA exactly. The decoding scheme is based on TGT-BERN-NONA, but has the following changes:

1. Estimation of number of defectives in a reference group: We first estimate the number of defective items in a single reference group, according to the empirical probability the reference group resulting in positive test outcomes. More precisely, let $\theta_v = \Pr(y_{p, r}^{(0)} = 1 \mid R_{p, r} \cap D = v)$ be the expected fraction of positive test outcomes. The decoder declares $R_{p, r}$ contains $v$ defective items if $|\mathcal{I}^{(0)}|\theta_v(1 - \Delta_{<v}) \leq \sum_i y_{p, r, i}^{(0)} \leq |\mathcal{I}^{(0)}|\theta_v(1 + \Delta_{>v})$, where the “variation” $\Delta_{<v}$ around the expectation is set to equal $(\theta_v - \theta_{v-1})/(2\theta_v)$ and $\Delta_{>v} = (\theta_{v+1} - \theta_v)/(2\theta_v)$.

2. Estimation of defectiveness of items: In this case, the threshold for estimating that an item is defective is different than in TGT-BERN-NONA, since the variation between the empirical probability of observing positive testing outcomes in tests including the reference group is different. Specifically, let $\alpha_{w, v} = \Pr(y_{p, r, i}^{(j)} = 1 \mid R_{p, r} \cap D = v, x_j = w)$. For every $R_{p, r} \subseteq P_{\rho}$ which contains $v \in (l, u)$ defective items and $x_j \in P_{\rho}$, if $\sum_i y_{p, r, i}^{(0)} \leq |\mathcal{I}^{(0)}|\alpha_{v_0}(1 + \Delta)$ where $\Delta_v = (\alpha_{v+1} - \alpha_{v_0})/(2\alpha_{v_0})$, the decoder declares $x_j$ to be non-defective, else declares it to be defective.

**Proof sketch of Theorem 3**: We note that Lemma 5 is not required, since any reference groups $R$ can be used to decode items, as long as the number of defective items in $R$ is between $l$ and $u$, so that there is some statistical difference between the probability of a positive test outcome if a group has $i$ defective items, or if it has $i + 1$ defective items). With high probability, a reference group with a suitably chosen size $(n(u + l)/2d)$ satisfies this relaxed condition. That is, the number of reference groups $R$ required is a constant.

A sufficient condition of high probability decoding of all items is when modified versions of Lemmas 6 and 7 are satisfied. The modifications in Lemma 6 and 7 correspond to the fact that the required number of indicator groups increase by a factor of $g^2$. This is because the decoder is based on estimating the probability difference of giving a positive outcome when the test has an additional defective item. Hence the difference between two different probabilities scales as $1/g$. By the Chernoff bound, to estimate such a probability difference sufficiently accurately requires a multiplicative factor of $g^2$ in the number of tests.

The rest of the argument is as in the proof of Theorem 1.

**REFERENCES**

[1] R. Dorfman, “The detection of defective members of large populations,” Annals of Mathematical Statistics, vol. 14, pp. 436–441, 1943.

[2] D. Du and F. K. Hwang, Combinatorial Group Testing and Its Applications, 2nd ed. World Scientific Publishing Company, 2000.

[3] D. Z. Du and F. K. Hwang, Pooling designs and nonadaptive group testing: important tools for DNA sequencing. World Scientific Publishing Company, 2006.

[4] H.-B. Chen and F. K. Hwang, “A survey on nonadaptive group testing algorithms through the angle of decoding,” J. Comb. Optim., vol. 15, no. 1, pp. 49–59, 2008.

[5] P. Damaschke, “Threshold group testing,” in General Theory of Information Transfer and Combinatorics, ser. Lecture Notes in Computer Science, R. Ahlswede, L. Bujner, N. Cai, H. Aydinian, V. Blinovsky, C. Deppe, and H. Mashurian, Eds. Springer Berlin Heidelberg, 2006, vol. 4123, pp. 707–718.

[6] F. Chin, H. Leung, and S. Yiu, “Non-adaptive complex group testing with inhibitors,” in Theory and Applications of Models of Computation, ser. Lecture Notes in Computer Science, M. Oghara and J. Tarui, Eds. Springer Berlin Heidelberg, 2011, vol. 6648, pp. 172–183.

[7] H. Chen and A. D. Bonis, “An almost optimal algorithm for generalized threshold group testing with inhibitors,” Journal of Computational Biology, vol. 18, no. 6, pp. 851 – 864, June 2011.

[8] A. Emad and O. Milenkovic, “Semi-Quantitative Group Testing: a General Paradigm with Applications in Genotyping,” Arkiv e-prints, Oct. 2012.

[9] M. Cheraghchi, “Improved constructions for non-daptive threshold group testing,” in Proceedings of the 37th international colloquium conference on Automata, languages and programming, ser. ICALP’10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 552-564.

[10] H. Chen, D. Du, and F. Hwang, “An unexpected meeting of four seemingly unrelated problems: graph testing, dna complex screening, superimposed codes and secure key distribution,” Journal of Combinatorial Optimization, vol. 14, no. 2-3, pp. 121–129, 2007. [Online]. Available: http://dx.doi.org/10.1007/s10878-007-9067-3
[11] H. Chang, H. Chen, H. Fu, and C. Shi, “Reconstruction of hidden graphs and threshold group testing,” J. Comb. Optim., vol. 22, no. 2, pp. 270–281, Aug. 2011. [Online]. Available: http://dx.doi.org/10.1007/s10878-010-9291-0

[12] H.-B. Chen and H.-L. Fu, “Nonadaptive algorithms for threshold group testing,” Discrete Applied Mathematics, vol. 157, no. 7, pp. 1581 – 1585, 2009.

[13] N. Alon, V. Guruswami, T. Kaufman, and M. Sudan, “Guessing secrets efficiently via list decoding,” ACM Trans. Algorithms, vol. 3, no. 4, Nov. 2007.

[14] R. Ahlswede, C. Deppe, and V. Lebedev, “Bounds for threshold and majority group testing,” in Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, 2011, pp. 69–73.

[15] G. Maltezos, M. Johnston, K. Taganov, C. Srichuntaratamee, J. Gorman, and D. Baltimore, “Exploring the limits of ultrafast polymerase chain reaction using liquid for thermal heat exchange: A proof of principle,” Applied Physics Letters, vol. 97, 2010.

[16] D. Balding, W. Bruno, D. Torney, and E. Knill, “A comparative survey of non-adaptive pooling designs,” in Genetic Mapping and DNA Sequencing, ser. The IMA Volumes in Mathematics and its Applications, T. Speed and M. Waterman, Eds. Springer New York, 1996, vol. 81, pp. 133–154.

[17] M. Mitzenmacher and E. Upfal, Probability and Computing: Randomized Algorithms and Probabilistic Analysis. Cambridge University Press, 2005.

[18] D. Hush and C. Scovel, “Concentration of the hypergeometric distribution,” Statistics and Probability Letters, vol. 75, no. 2, pp. 127 – 132, 2005.