A Tight Scale-Locality Bound for Partial Detection in Non-Adaptive Group Testing

Nader H. Bshouty[1]

[1]Technion

Abstract

We give a lower bound for randomized non-adaptive group testing when the goal is to find any ℓ defective items but the total number d of defectives is unknown. Bshouty and Haddad-Zaknoon proved an upper bound of $O(\ell \log^2 n)$ tests and a lower bound of

$$\Omega\left(\frac{\ell \log^2 n}{\log \ell + \log\log n}\right).$$

We prove the matching lower bound. More generally, we show that every randomized non-adaptive algorithm that succeeds with constant probability for every defective set must use

$$\Omega\big(\ell \log^2(n/\ell)\big)$$

tests.

The proof is as follows. At a fixed value of d, finding ℓ defectives requires about $\ell \log(n/d)$ bits of information. On the other hand, one fixed group test is informative only when its size is tuned to the scale of d; across all logarithmic scales of d, a single test contributes only $O(1)$ bits. Summing over all scales gives the lower bound.

We also record the matching upper bound

$$O\big(\ell \log^2(n/\ell)\big),$$

obtained by running the known-d algorithm in parallel over dyadic guesses for d. Thus the randomized non-adaptive complexity of unknown-d partial detection is $\Theta\big(\ell \log^2(n/\ell)\big)$ for constant success probability.

1. Introduction

Group testing is a classical method for identifying defective items in a large population using pooled tests. Given a ground set $[n] = \{1, \ldots, n\}$ and an unknown defective set $I \subseteq [n]$, a group test is a subset $Q \subseteq [n]$ whose answer is positive if and only if $Q \cap I \neq \emptyset$. The model was introduced by Dorfman in the context of economical blood testing [15], and has since become a central topic in combinatorics, algorithms, statistics, coding theory, and information theory.

The literature on group testing is extensive. Classical adaptive procedures include the work of Sobel and Groll [42] and Hwang's generalized binary splitting method [30]. Competitive and adaptive variants have been studied in [3, 18, 20, 21, 11, 40, 45]. Non-adaptive group testing is closely connected to pooling designs, superimposed codes, disjunct matrices, cover-free families, and coding theory; see, for example, [2, 17, 16, 19, 32, 22, 10, 25, 39, 37, 38]. Randomized group testing and algorithms with optimal or near-optimal query complexity have been investigated in [7, 8, 14]. The related problem of estimating the number of defectives has been studied in [24, 6, 5].

Group testing has also found many applications. It appears in DNA library screening and pooling designs [17, 16, 19], product quality control and sequential screening [34], image compression [29], data-stream and frequent-item problems [12], and multiaccess communications [44]. More recently, neural and machine-learning motivated forms of group testing have been studied in [35]. During the COVID-19 pandemic, group testing was widely discussed as a way to accelerate PCR testing and reduce laboratory costs [4, 9, 23, 27, 36, 41, 47, 28]. Related motivations also arise in abnormal-event detection

ICTCS 2026: Italian Conference on Theoretical Computer Science, September 7–9, 2026, Udine, Italy

bshouty@cs.technion.ac.il (N. H. Bshouty)

and large-scale image classification, where one may wish to find a small number of positive or abnormal instances among many candidates [33, 43, 46, 35].

In this paper we study a partial-detection variant of group testing. Instead of identifying all $d = |I|$ defective items, the goal is to output any ℓ defective items, where $\ell \leq d$. This problem was considered by Ahlswede, Deppe, and Lebedev [1], who studied the case of finding one defective item and gave bounds for related models. Katona [31] studied the case of finding at least one excellent element, including deterministic non-adaptive and two-round settings. Gerbner and Vizer [26] generalized these results to deterministic r-round algorithms and arbitrary ℓ.

Partial detection is useful in settings where finding a small number of confirmed defectives is already enough to trigger a downstream action, such as triage, screening, or quickly isolating a small infected cluster, and full identification of all d defectives may be unnecessary or too costly.

The systematic study of detecting ℓ defective items from d defectives was carried out by Bshouty and Haddad-Zaknoon in [48]. They considered adaptive and non-adaptive algorithms, deterministic and randomized algorithms, and both the case where d is known up to a constant factor and the case where no prior information about d is available. In the randomized non-adaptive setting with unknown d, they proved an upper bound of $O(\ell \log^2 n)$ tests and a lower bound of

$$\Omega\left(\frac{\ell \log^2 n}{\log \ell + \log\log n}\right).$$

Thus their work left a logarithmic gap in the most difficult randomized non-adaptive unknown-d regime.

Our contribution is to close this gap. We prove that every randomized non-adaptive algorithm which, with constant probability, outputs ℓ defective items for every defective set must use

$$\Omega\big(\ell \log^2(n/\ell)\big)$$

tests. We also record the matching upper bound

$$O\big(\ell \log^2(n/\ell)\big)\,,$$

obtained by running the known-d randomized non-adaptive algorithm of [48] in parallel over dyadic guesses for d. Therefore the randomized non-adaptive complexity of partial detection with unknown d is

$$\Theta\big(\ell \log^2(n/\ell)\big)$$

for constant success probability.

The proof is based on a scale-locality principle. At a fixed value of d, outputting ℓ defective items requires about $\ell \log(n/d)$ bits of information. On the other hand, a fixed group test is informative only when its size is tuned to the scale of d: if the test is too small, it is almost always negative; if it is too large, it is almost always positive. Hence a single test contributes only $O(1)$ bits total over all logarithmic scales of d. Summing the fixed-scale information requirements over all scales yields the lower bound. This perspective is related to the scale-locality ideas used in the tight lower bound for non-adaptive estimation of the number of defectives [49], but the argument here is an entropy direct-sum over all scales rather than a two-distribution total-variation argument.

1.1. Our Technique

The proof of the lower bound is based on a scale-locality principle. We consider many possible values of the number of defectives, $d_0, d_1, d_2, \ldots$, where each value is twice the previous one. First fix one such value d. If the defective set I is chosen uniformly among all d-subsets of $[n]$, then any one fixed ℓ-set is contained in I with probability roughly $(d/n)^\ell$. Therefore, in order to output an ℓ-set contained in I with constant probability, the transcript of the algorithm must reveal about $\ell \log(n/d)$ bits of information about I.

We then look at a single fixed test Q. Let $k = |Q|$. If the number of defectives is d, then the expected number of defectives in Q is kd/n. If this quantity is much smaller than 1, then Q probably contains

no defective item, so the answer to the test is almost always 0. If this quantity is much larger than 1, then Q probably contains at least one defective item, so the answer is almost always 1. In both cases the answer is nearly predictable and gives very little information. Thus the test can be informative only for those values of d for which kd/n is close to 1. Since the possible values of d grow geometrically, this can happen for only constantly many scales. Hence one fixed test contributes only $O(1)$ bits total over all logarithmic scales of d.

Finally, we sum over all scales. Each scale d requires about $\ell \log(n/d)$ bits, while each test supplies only $O(1)$ bits in total across all scales. This gives the lower bound

$$\Omega\big(\ell \log^2(n/\ell)\big)\,.$$

This perspective is related to the scale-locality ideas used in the tight lower bound for non-adaptive estimation of the number of defectives [49]. The difference is that the estimation lower bound uses a two-distribution total-variation argument, whereas the present proof uses an entropy direct-sum over all scales.

The upper bound uses a parallel guessing strategy. If a constant-factor estimate of d is known, then the randomized non-adaptive algorithm of [48] finds ℓ defective items with the optimal fixed-scale number of tests. When d is unknown, we run this fixed-scale algorithm in parallel for all dyadic guesses of d, starting from ℓ and going up to n. One of these guesses is within a constant factor of the true value of d, so the corresponding copy of the algorithm succeeds. Since all copies are run in parallel, the resulting algorithm is still non-adaptive. A direct summation of the fixed-scale costs over the dyadic guesses gives the matching upper bound $O(\ell \log^2(n/\ell))$.

2. Model and Main Result

For a defective set $I \subseteq [n]$ and a test $Q \subseteq [n]$, define

$$T_I(Q) = \mathbf{1}[Q \cap I \neq \varnothing].$$

Thus $T_I(Q) = 1$ if Q contains at least one defective item, and $T_I(Q) = 0$ otherwise.

A randomized non-adaptive algorithm that detects ℓ defective items using q tests first samples a random seed S. After $S = s$ is fixed, the algorithm chooses deterministic tests

$$Q_1(s), \dots, Q_q(s) \subseteq [n].$$

Given a defective set I, the answer vector is

$$Y(I, s) = \big(T_I(Q_1(s)), \dots, T_I(Q_q(s))\big) \in \{0,1\}^q.$$

The algorithm then outputs a set

$$L = L(s, Y(I, s)) \subseteq [n].$$

We say that the algorithm *succeeds* on I if

$$|L| = \ell \qquad \text{and} \qquad L \subseteq I.$$

The algorithm has success probability at least p if, for every defective set I with $|I| \geq \ell$, the probability over the random seed S that it succeeds on I is at least p.

Theorem 1 (Main theorem)**.** *There is an absolute constant $c_0 > 0$ such that the following holds. Let*[1] *$1 \leq \ell \leq n/16$. Suppose a randomized non-adaptive group testing algorithm succeeds with probability at least $2/3$ on every defective set $I \subseteq [n]$ with $|I| \geq \ell$. Then the number of tests satisfies*

$$q \geq c_0\, \ell \log^2(n/\ell).$$

[1]The restriction $\ell \leq n/16$ is used only to focus on the sparse partial-detection regime. When $\ell = \Theta(n)$, the expression $\ell \log^2(n/\ell)$ is $\Theta(n)$, and the trivial deterministic non-adaptive algorithm that tests all singletons gives a linear upper bound.

Theorem 2 (Matching upper bound)**.** *There is an absolute constant $C_0 > 0$ such that the following holds. Let $1 \le \ell \le n/16$. There is a randomized non-adaptive group testing algorithm which, without knowing $d = |I|$, succeeds with constant probability on every defective set $I \subseteq [n]$ with $|I| \ge \ell$ and uses at most*

$$C_0\, \ell \log^2(n/\ell)$$

tests.

Corollary 3. *For every $1 \le \ell \le n/16$, the randomized non-adaptive test complexity of unknown-d partial detection is*

$$\Theta\big(\ell \log^2(n/\ell)\big)$$

for a constant success probability.

In particular, for every fixed $c < 1$, if $\ell \le n^c$, then the complexity is

$$\Theta_c(\ell \log^2 n).$$

3. Lower Bound

3.1. Entropy

For a discrete random variable X taking values in a finite set $\mathcal{X}$, its entropy is defined by

$$H(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log \frac{1}{\Pr[X = x]},$$

where terms with $\Pr[X = x] = 0$ are omitted. Equivalently,

$$H(X) = \mathbb{E}\left[\log \frac{1}{\Pr[X]}\right].$$

Entropy measures the average number of bits needed to describe the outcome of X.

For two discrete random variables X and Y, the conditional entropy of X given Y is

$$H(X \mid Y) = \sum_{y} \Pr[Y = y]\, H(X \mid Y = y).$$

Equivalently,

$$H(X \mid Y) = \sum_{y} \Pr[Y = y] \sum_{x} \Pr[X = x \mid Y = y] \log \frac{1}{\Pr[X = x \mid Y = y]},$$

where terms with probability 0 are omitted. The conditional entropy $H(X \mid Y)$ is the average number of additional bits needed to describe X after the value of Y is known.

We use the following standard facts about entropy; see, for example, Cover and Thomas [13].

(i) **Maximum entropy bound**: If X takes at most M possible values, then

$$H(X) \le \log M.$$

Equality holds when X is uniformly distributed over M values.

(ii) **Deterministic dependence**: if X is determined by Y, that is, $X = f(Y)$ for some function f, then

$$H(X \mid Y) = 0.$$

(iii) **Conditioning reduces entropy**: $H(X \mid Y) \le H(X)$.

(iv) **Chain rule for entropy**: $H(X|Y) - H(X|Y,Z) = H(Z|Y) - H(Z|X,Y)$.

(v) **Subadditivity of entropy**: If $X = (X_1, \dots, X_q)$, then

$$H(X) \le H(X_1) + \cdots + H(X_q).$$

(vi) If B is a 0-1 random variable with $\Pr[B = 1] = p$, then

$$H(B) = h(p) := p \log(1/p) + (1-p) \log(1/(1-p)).$$

We also use the standard estimates

$$h(p) \le Cp \log(1/p) \qquad (0 < p \le 1/2) \tag{1}$$

for an absolute constant C, and the symmetric bound $h(p) = h(1-p)$. Indeed, for $0 < p \le 1/2$,

$$h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \le p \log \frac{1}{p} + 2p \le Cp \log \frac{1}{p},$$

for an absolute constant C, since $\log(1/p) \ge \log 2$.

3.2. One Fixed Scale Requires Many Bits

First fix the number of defectives d. Let I be uniformly random among all d-subsets of $[n]$.

We call $Z = (S, Y)$ the *transcript* of the algorithm. It contains the random seed S and all test answers Y. The output of the algorithm is then a function of the transcript, $L = L(Z)$.

The next lemma says that if an algorithm can output ℓ elements of I with constant probability, then its transcript must contain $\Omega(\ell \log(n/d))$ bits about I. Equivalently, $H(I)$ is the average number of bits needed to describe I before seeing the transcript, $H(I \mid Z)$ is the average number of bits still needed after seeing Z, and $H(I) - H(I \mid Z) = \Omega(\ell \log(n/d))$ is the number of bits about I revealed by Z.

Lemma 4 (Information needed at one scale)**.** *Let $\ell \le d \le n/4$. Let I be uniformly random among all d-subsets of $[n]$. Let Z be any transcript, and suppose that from Z one can compute an ℓ-set $L = L(Z)$ such that*

$$\Pr[L \subseteq I] \ge 2/3.$$

Then

$$H(I) - H(I \mid Z) = \Omega(\ell \log(n/d)).$$

In words, the transcript Z reveals $\Omega(\ell \log(n/d))$ bits of information about I.

Proof. Let G be the event that the output is correct: $G = \{L \subseteq I\}$. By assumption, $\Pr[G] \ge 2/3$.

Before seeing any transcript, the defective set I is uniform over $\binom{n}{d}$ possibilities, so, by (i),

$$H(I) = \log \binom{n}{d}.$$

Now imagine that we are only told the output set L. If the event G occurs, then I must contain all elements of L. Once those ℓ elements are fixed, the remaining $d - \ell$ defective items can be chosen in at most

$$\binom{n-\ell}{d-\ell}$$

ways. If G does not occur, we use the trivial bound that I has at most $\binom{n}{d}$ possibilities. Also, revealing whether G occurred costs at most one extra bit. Therefore

$$H(I \mid L) \le 1 + \Pr[G] \log \binom{n-\ell}{d-\ell} + (1 - \Pr[G]) \log \binom{n}{d}.$$

Thus

$$\begin{aligned} H(I) - H(I \mid L) &\geq \Pr[G] \left(\log \binom{n}{d} - \log \binom{n-\ell}{d-\ell} \right) - 1 \\ &= \Pr[G] \log \left(\frac{\binom{n}{d}}{\binom{n-\ell}{d-\ell}} \right) - 1 \\ &= \Pr[G] \log \left(\frac{\binom{n}{\ell}}{\binom{d}{\ell}} \right) - 1. \end{aligned}$$

Since $d \leq n/4$ and $\ell \leq d$, the ratio $\binom{n}{\ell}/\binom{d}{\ell}$ is at least a constant multiple of $(n/d)^\ell$. Hence

$$H(I) - H(I \mid L) = \Omega(\ell \log(n/d)).$$

Finally, L is computed from Z, so knowing Z can only reduce the uncertainty about I further:

$$H(I \mid Z) \leq H(I \mid L).$$

Therefore

$$H(I) - H(I \mid Z) \geq H(I) - H(I \mid L) = \Omega(\ell \log(n/d)).$$

This proves the lemma. □

3.3. One Test Is Useful at Only Constantly Many Scales

Now we let d vary over many possible values. Define

$$d_j = 4\ell \cdot 2^j, \qquad j = 0, 1, \ldots, m,$$

where m is the largest integer with $d_m \leq n/4$. Since d_j doubles at every step,

$$m = \lfloor \log(n/\ell) \rfloor - 4 = \Theta(\log(n/\ell)).$$

For each j, let I_j be uniformly random among all d_j-subsets of $[n]$.

Fix a single test $Q \subseteq [n]$, and let $k = |Q|$. At scale j, define

$$B_j(Q) = T_{I_j}(Q) = \mathbf{1}[Q \cap I_j \neq \varnothing].$$

The following lemma makes precise the idea that a single test has only $O(1)$ total usefulness across all scales.

Lemma 5 (Scale locality for one test)**.** *For every fixed test $Q \subseteq [n]$,*

$$\sum_{j=0}^{m} H(B_j(Q)) = O(1).$$

Proof. Let $x_j = kd_j/n$. This is the expected number of defectives in Q when there are d_j defectives. Since $d_{j+1} = 2d_j$, we have $x_{j+1} = 2x_j$. So the sequence x_j doubles as j increases.

Let

$$p_j = \Pr[B_j(Q) = 1] = \Pr[Q \cap I_j \neq \varnothing].$$

We split into three cases.

Small scales: indices j with $x_j \leq 1/2$. By Markov's inequality,

$$p_j = \Pr[|Q \cap I_j| \geq 1] \leq \mathbb{E}\,[|Q \cap I_j|] = x_j.$$

Hence, using (1),

$$H(B_j(Q)) = h(p_j) \le C x_j \log(1/x_j).$$

Since the positive numbers x_j double from one scale to the next, the sum of $x_j \log(1/x_j)$ over all j with $x_j \le 1/2$ is bounded by an absolute constant.

Middle scales: indices j with $1/2 < x_j < 2$. There are at most three such indices j, because x_j doubles each time. Each contributes at most one bit, so the total contribution is $O(1)$.

Large scales: indices j with $x_j \ge 2$. Let

$$r_j = \Pr[B_j(Q) = 0] = \Pr[Q \cap I_j = \varnothing].$$

Then

$$r_j = \frac{\binom{n-k}{d_j}}{\binom{n}{d_j}} \le \left(1 - \frac{k}{n}\right)^{d_j} \le e^{-kd_j/n} = e^{-x_j}.$$

Since $H(B_j(Q)) = h(r_j)$ and $r_j \le e^{-x_j} \le e^{-2} < 1/2$, again using (1),

$$H(B_j(Q)) \le C x_j e^{-x_j}.$$

The sum of $x_j e^{-x_j}$ over a sequence that doubles at every step and starts at least 2 is bounded by an absolute constant.

Adding the three ranges gives

$$\sum_{j=0}^{m} H(B_j(Q)) = O(1).$$

□

3.4. Proof of the Main Theorem

We now prove Theorem 1.

Assume the algorithm uses q tests. For each seed s, the tests are fixed sets $Q_1(s), \dots, Q_q(s)$. For scale j, let $Y_j = \big(T_{I_j}(Q_1(S)), \dots, T_{I_j}(Q_q(S))\big)$ be the answer vector. The algorithm sees both S and Y_j, and then outputs

$$L_j = L(S, Y_j).$$

Since the algorithm succeeds with probability at least $2/3$ on every defective set, it also succeeds with probability at least $2/3$ when I_j is uniformly random among all d_j-sets. Therefore Lemma 4, applied with transcript $Z = (S, Y_j)$, gives

$$H(I_j) - H(I_j \mid S, Y_j) = \Omega(\ell \log(n/d_j)).$$

Since the random seed S is independent of I_j, we have $H(I_j) = H(I_j \mid S)$. By the chain rule for entropy ((iv)),

$$H(I_j \mid S) - H(I_j \mid S, Y_j) = H(Y_j \mid S) - H(Y_j \mid I_j, S) \le H(Y_j \mid S).$$

Therefore $H(I_j) - H(I_j \mid S, Y_j) \le H(Y_j \mid S)$. Hence, for every j, $H(Y_j \mid S) = \Omega(\ell \log(n/d_j))$. Summing over all scales,

$$\sum_{j=0}^{m} H(Y_j \mid S) \ge \Omega\left(\ell \sum_{j=0}^{m} \log(n/d_j)\right). \tag{2}$$

We next upper-bound the same left-hand side. Fix the seed $S = s$. By subadditivity of entropy (v),

$$H(Y_j \mid S = s) \le \sum_{i=1}^{q} H\big(T_{I_j}(Q_i(s))\big).$$

Therefore

$$\sum_{j=0}^{m} H(Y_j \mid S=s) \le \sum_{j=0}^{m}\sum_{i=1}^{q} H\big(T_{I_j}(Q_i(s))\big) = \sum_{i=1}^{q}\sum_{j=0}^{m} H\big(T_{I_j}(Q_i(s))\big).$$

For each fixed test $Q_i(s)$, Lemma 5 says that the inner sum over j is $O(1)$. Thus

$$\sum_{j=0}^{m} H(Y_j \mid S=s) = O(q).$$

Averaging over S gives

$$\sum_{j=0}^{m} H(Y_j \mid S) = O(q). \tag{3}$$

Combining (2) and (3),

$$q \ge \Omega\left(\ell \sum_{j=0}^{m} \log(n/d_j)\right).$$

Recall that $d_j = 4\ell \cdot 2^j$. Therefore $\log(n/d_j) = \log(n/4\ell) - j$. By the definition of m, we have $m = \lfloor \log(n/\ell) \rfloor - 4$. Hence

$$\sum_{j=0}^{m} \log(n/d_j) = \sum_{j=0}^{m} (\log(n/4\ell) - j) \ge \sum_{j=0}^{\lfloor m/2 \rfloor} (\log(n/4\ell) - j) = \Omega(\log^2(n/4\ell)).$$

Therefore

$$q \ge \Omega\left(\ell \sum_{j=0}^{m} \log(n/d_j)\right) = \Omega\big(\ell \log^2(n/\ell)\big).$$

This proves Theorem 1.

4. A Matching Upper Bound

We now prove the matching upper bound. We use two ingredients from [48]: Theorem 11, which gives the randomized non-adaptive upper bound when a constant-factor estimate of d is known, and Lemma 9, which gives a randomized non-adaptive constant-factor estimator for d. We briefly recall the black-box tools from [48]. The known-D algorithm repeatedly isolates a single defective by random sampling at the appropriate scale $1/D$, while the estimator probes geometrically many sampling rates and identifies a constant-factor estimate of d from the transition between mostly negative and mostly positive answers.

First, by Lemma 9 in [48], if a number D is known such that $d/4 \le D \le 4d$, then there is a randomized non-adaptive algorithm that detects ℓ defective items using $O(\ell \log(n/d))$ tests for constant success probability. Since D and d differ by at most a constant factor, this may also be written as

$$O\Big(\ell \log \frac{n}{D}\Big).$$

Second, by Lemma 9 of [48], there is a randomized non-adaptive estimator which, using $O(\log(1/\delta) \log n)$ tests, returns a value $\widehat{D}$ satisfying $d/2 < \widehat{D} < 2d$ with probability at least $1 - \delta$.

The algorithm runs the estimator and, in parallel, runs the known-D algorithm for all dyadic guesses

$$D_i = 2^i \ell, \qquad i = 0, 1, \ldots, \lceil \log(n/\ell) \rceil.$$

After all test answers are received, the algorithm looks at the estimate $\widehat{D}$ and chooses an index i such that D_i is within a factor 2 of $\widehat{D}$. If the estimator is correct, then this D_i is within a constant factor of the true value of d, and therefore the corresponding known-D_i algorithm succeeds with constant probability. Thus the algorithm does not need to verify which of the parallel outputs is correct; it uses the estimator to select the appropriate output.

The total number of tests is the sum of the costs over all guesses:

$$\sum_{i=0}^{\lceil \log(n/\ell) \rceil} O\Big(\ell \log \frac{n}{2^i \ell}\Big) = O\big(\ell \log^2(n/\ell)\big)\,.$$

The estimator contributes only $O(\log n)$ additional tests for constant success probability, which is dominated by the above bound. Hence the total number of tests is $O\big(\ell \log^2(n/\ell)\big)$.

Theorem 6 (Matching upper bound)**.** *Let $1 \le \ell \le n/16$ and let $0 < \delta < 1$. There is a randomized non-adaptive group testing algorithm which, without knowing $d = |I|$, outputs ℓ defective items with probability at least $1 - \delta$ for every defective set $I \subseteq [n]$ satisfying $|I| \ge \ell$, using*

$$O\big((\ell + \log(1/\delta)) \log^2(n/\ell) + \log(1/\delta) \log n\big)$$

tests.

In particular, this yields Theorem 2. For a constant success probability, the required number of tests is $O\big(\ell \log^2(n/\ell)\big)$.

Proof. We use two ingredients from [48]. The first is Theorem 11 there: if a number D is known in advance and $d/4 \le D \le 4d$, then there is a randomized non-adaptive algorithm that detects ℓ defective items with probability at least $1 - \delta$ using $O((\ell + \log(1/\delta)) \log(n/d))$ tests. Since D and d differ by at most a constant factor, this can be written as $O((\ell + \log(1/\delta)) \log(2n/D))$.

The second ingredient is Lemma 9 of [48]: there is a randomized non-adaptive estimator which, using $O(\log(1/\delta) \log n)$ tests, returns a value $\widehat{D}$ satisfying $d/2 < \widehat{D} < 2d$ with probability at least $1 - \delta$.

We now construct the unknown-d algorithm. Let

$$M = \lceil \log(n/\ell) \rceil, \qquad D_i = \min\{2^i \ell, n\}, \qquad i = 0, 1, \ldots, M.$$

Since the algorithm is non-adaptive, we cannot first run the estimator, wait for $\widehat{D}$, and then choose the tests for the known-D algorithm. Therefore, before seeing any answers, we run in parallel:

1. the estimator, with failure probability at most $\delta/2$;
2. for every $i = 0, 1, \ldots, M$, the known-D_i algorithm, also with failure probability at most $\delta/2$.

All tests are chosen before any answers are observed, so the whole algorithm is non-adaptive.

After all answers are received, compute the estimate $\widehat{D}$. Choose an index i such that D_i is within a factor 2 of $\widehat{D}$, and output the set produced by the corresponding known-D_i algorithm.

If the estimator succeeds, then $d/2 < \widehat{D} < 2d$. Since D_i is chosen within a factor 2 of $\widehat{D}$, we get $d/4 < D_i < 4d$. Thus the selected known-D_i algorithm has a valid constant-factor estimate of d, and therefore it succeeds with probability at least $1 - \delta/2$. By the union bound, the estimator succeeds and the selected known-D_i algorithm succeeds with probability at least $1 - \delta$.

It remains to count the tests. The estimator uses $O(\log(1/\delta) \log n)$ tests. The known-D_i algorithms use altogether

$$\sum_{i=0}^{M} O((\ell + \log(1/\delta)) \log(2n/D_i)) = O\left((\ell + \log(1/\delta)) \sum_{i=0}^{M} \log \frac{2n}{2^i \ell}\right)$$

$$= O\big((\ell + \log(1/\delta)) \log^2(n/\ell)\big)\,.$$

Therefore the total number of tests is $O\big((\ell + \log(1/\delta)) \log^2(n/\ell) + \log(1/\delta) \log n\big)$. For constant δ, this becomes $O\big(\ell \log^2(n/\ell)\big)$. □

Combining Theorems 1 and 2, the randomized non-adaptive complexity of unknown-d partial detection, for constant success probability, is $\Theta\big(\ell \log^2(n/\ell)\big)$.

5. Conclusion

We determined the randomized non-adaptive test complexity of detecting ℓ defective items when the total number d of defectives is unknown. The main result is the tight bound $\Theta\big(\ell \log^2(n/\ell)\big)$ for constant success probability.

The lower bound follows from a scale-locality argument. At a fixed value of d, any successful algorithm must reveal enough information to identify ℓ defective items, which costs about $\ell \log(n/d)$ bits. However, a single group test is useful only for a small range of possible values of d: if the test is too small it is almost always negative, and if it is too large it is almost always positive. Thus each test contributes only $O(1)$ useful bits over all logarithmic scales of d. Summing the information required at all scales gives the lower bound $\Omega\big(\ell \log^2(n/\ell)\big)$.

The matching upper bound is obtained by running, in parallel, the known-d randomized non-adaptive algorithm for all dyadic guesses of d, together with a non-adaptive estimator for d that is used only after all answers are known to select the appropriate copy. This keeps the algorithm non-adaptive and gives the same order of tests as the lower bound.

This closes the gap left by the previous bounds for randomized non-adaptive partial detection with unknown d.

Open problem. Develop a scale-locality theory for partial detection under threshold variants, gap-threshold tests, and noisy group-testing models. In particular, determine whether the unknown-d cost remains an additional logarithmic scale factor, and identify the optimal dependence on the threshold, the gap size, or the noise rate.

Acknowledgements. We thank the anonymous reviewers of ICTCS 2026 for their careful reading of the paper and for their helpful comments and suggestions. Their feedback helped us improve the presentation and clarify several points.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

[1] R. Ahlswede, C. Deppe, and V. Lebedev. Finding one of d defective elements in some group testing models. *Problems of Information Transmission*, 48, 2012. https://doi.org/10.1134/S0032946012020068.

[2] D. J. Balding, W. J. Bruno, D. Torney, and E. Knill. A comparative survey of non-adaptive pooling designs. In T. Speed and M. S. Waterman, editors, *Genetic Mapping and DNA Sequencing*, pages 133–154. Springer, New York, 1996.

[3] A. Bar-Noy, F. K. Hwang, I. Kessler, and S. Kutten. A new competitive algorithm for group testing. *Discrete Applied Mathematics*, 52(1):29–38, 1994. https://doi.org/10.1016/0166-218X(92)00185-O.

[4] R. Ben-Ami et al. Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clinical Microbiology and Infection*, 26(9):1248–1253, 2020. https://doi.org/10.1016/j.cmi.2020.06.009.

[5] N. H. Bshouty. Lower bound for non-adaptive estimation of the number of defective items. In *30th International Symposium on Algorithms and Computation, ISAAC 2019*, pages 2:1–2:9, 2019. https://doi.org/10.4230/LIPIcs.ISAAC.2019.2.

[6] N. H. Bshouty, V. E. Bshouty-Hurani, G. Haddad, T. Hashem, F. Khoury, and O. Sharafy. Adaptive group testing algorithms to estimate the number of defectives. In *Algorithmic Learning Theory, ALT 2018*, pages 93–110, 2018.

[7] N. H. Bshouty, N. Diab, S. R. Kawar, and R. J. Shahla. Non-adaptive randomized algorithm for group testing. In *International Conference on Algorithmic Learning Theory, ALT 2017*, pages 109–128, 2017.

[8] N. H. Bshouty, G. Haddad, and C. A. Haddad-Zaknoon. Bounds for the number of tests in non-adaptive randomized algorithms for group testing. In *SOFSEM 2020: Theory and Practice of Computer Science*, pages 101–112, 2020. https://doi.org/10.1007/978-3-030-38919-2_9.

[9] J. J. Cabrera Alvargonzalez, S. Rey Cao, S. Pérez Castro, L. Martinez Lamas, O. Cores Calvo, J. Torres Piñon, J. Porteiro Fresco, J. Garcia Comesãna, and B. Regueiro Garcia. Pooling for SARS-CoV-2 control in care institutions. *BMC Infectious Diseases*, 20(1):1–6, 2020.

[10] H. Chen and F. K. Hwang. Exploring the missing link among d-separable, $\bar{d}$-separable and d-disjunct matrices. *Discrete Applied Mathematics*, 155(5):662–664, 2007. https://doi.org/10.1016/j.dam.2006.10.009.

[11] Y. Cheng, D. Du, and Y. Xu. A zig-zag approach for competitive group testing. *INFORMS Journal on Computing*, 26(4):677–689, 2014. https://doi.org/10.1287/ijoc.2014.0591.

[12] G. Cormode and S. Muthukrishnan. What's hot and what's not: Tracking most frequent items dynamically. *ACM Transactions on Database Systems*, 30(1):249–278, 2005. https://doi.org/10.1145/1061318.1061325.

[13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Second edition, Wiley-Interscience, 2006.

[14] P. Damaschke and A. S. Muhammad. Randomized group testing both query-optimal and minimal adaptive. In *SOFSEM 2012: Theory and Practice of Computer Science*, pages 214–225, 2012. https://doi.org/10.1007/978-3-642-27660-6_18.

[15] R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

[16] D. Du and F. K. Hwang. *Pooling Design and Nonadaptive Group Testing: Important Tools for DNA Sequencing*. World Scientific, 2006.

[17] D. Z. Du and F. K. Hwang. *Combinatorial Group Testing and Its Applications*. World Scientific, 1993.

[18] D. Du and F. K. Hwang. Competitive group testing. *Discrete Applied Mathematics*, 45(3):221–232, 1993. https://doi.org/10.1016/0166-218X(93)90011-C.

[19] D. Z. Du and F. K. Hwang. *Pooling Designs and Nonadaptive Group Testing: Important Tools for DNA Sequencing*. World Scientific, 2006.

[20] D. Du and H. Park. On competitive group testing. *SIAM Journal on Computing*, 23(5):1019–1025, 1994. https://doi.org/10.1137/S0097539793246690.

[21] D. Du, G. Xue, S. Sun, and S. Cheng. Modifications of competitive group testing. *SIAM Journal on Computing*, 23(1):82–96, 1994. https://doi.org/10.1137/S0097539792227612.

[22] A. G. D'yachkov and V. V. Rykov. Bounds on the length of disjunctive codes. *Problems of Information Transmission*, 18(3):166–171, 1982.

[23] A. M. Eis-Hübinger et al. Ad hoc laboratory-based surveillance of SARS-CoV-2 by real-time RT-PCR using minipools of RNA prepared from routine respiratory samples. *Journal of Clinical Virology*, 127:104381, 2020. https://doi.org/10.1016/j.jcv.2020.104381.

[24] M. Falahatgar, A. Jafarpour, A. Orlitsky, V. Pichapati, and A. T. Suresh. Estimating the number of defectives with group testing. In *IEEE International Symposium on Information Theory, ISIT 2016*, pages 1376–1380, 2016. https://doi.org/10.1109/ISIT.2016.7541524.

[25] Z. Füredi. On r-cover-free families. *Journal of Combinatorial Theory, Series A*, 73(1):172–173, 1996. https://doi.org/10.1006/jcta.1996.0012.

[26] D. Gerbner and M. Vizer. Rounds in a combinatorial search problem. *Discrete Applied Mathematics*,

276:60–68, 2020. https://doi.org/10.1016/j.dam.2019.11.016.

[27] C. Gollier and O. Gossner. Group testing against COVID-19. *Covid Economics*, pages 32–42, 2020.

[28] C. A. Haddad-Zaknoon. Heuristic random designs for exact identification of defectives using single round non-adaptive group testing and compressed sensing. In *The Fourteenth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, BIOTECHNO*, 2022.

[29] E. S. Hong and R. E. Ladner. Group testing for image compression. *IEEE Transactions on Image Processing*, 11(8):901–911, 2002.

[30] F. K. Hwang. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association*, 67:605–608, 1972.

[31] G. O. Katona. Finding at least one excellent element in two rounds. *Journal of Statistical Planning and Inference*, 141(8):2946–2952, 2011. https://doi.org/10.1016/j.jspi.2011.03.019.

[32] W. Kautz and R. Singleton. Nonrandom binary superimposed codes. *IEEE Transactions on Information Theory*, 10(4):363–377, 1964.

[33] P. Kuppusamy and V. Bharathi. Human abnormal behavior detection using CNNs in crowded and uncrowded surveillance: A survey. *Measurement: Sensors*, 24:100510, 2022. https://doi.org/10.1016/j.measen.2022.100510.

[34] C. Li. A sequential method for screening experimental variables. *Journal of the American Statistical Association*, 57:455–477, 1962. https://doi.org/10.1080/01621459.1962.10480672.

[35] W. Liang and J. Zou. Neural group testing to accelerate deep learning. In *IEEE International Symposium on Information Theory, ISIT 2021*. IEEE, 2021.

[36] C. Mentus, M. Romeo, and C. DiPaola. Analysis and applications of adaptive group testing methods for COVID-19. *medRxiv*, 2020.

[37] E. Porat and A. Rothschild. Explicit nonadaptive combinatorial group testing schemes. *IEEE Transactions on Information Theory*, 57(12):7982–7989, 2011. https://doi.org/10.1109/TIT.2011.2163296.

[38] R. M. Roth. *Introduction to Coding Theory*. Cambridge University Press, 2006.

[39] M. Ruszinkó. On the upper bound of the size of the r-cover-free families. *Journal of Combinatorial Theory, Series A*, 66(2):302–310, 1994.

[40] J. Schlaghoff and E. Triesch. Improved results for competitive group testing. *Combinatorics, Probability and Computing*, 14(1–2):191–202, 2005. https://doi.org/10.1017/S0963548304006649.

[41] H. Shani-Narkiss, O. D. Gilday, N. Yayon, and I. D. Landau. Efficient and practical sample pooling for high-throughput PCR diagnosis of COVID-19. *medRxiv*, 2020.

[42] M. Sobel and P. A. Groll. Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Technical Journal*, 38:1179–1252, 1959.

[43] W. Wang and K. Siau. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management*, 30(1):61–79, 2019.

[44] J. Wolf. Born again group testing: Multiaccess communications. *IEEE Transactions on Information Theory*, 31(2):185–191, 1985.

[45] J. Wu, Y. Cheng, and D. Du. An improved zig zag approach for competitive group testing. *Discrete Optimization*, 43:100687, 2022. https://doi.org/10.1016/j.disopt.2022.100687.

[46] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1492–1500, 2017.

[47] I. Yelin, N. Aharony, E. Shaer-Tamar, A. Argoetti, E. Messer, D. Berenbaum, E. Shafran, A. Kuzli, N. Gandali, T. Hashimshony, Y. Mandel-Gutfreund, M. Halberthal, Y. Geffen, M. Szwarcwort-Cohen, and R. Kishony. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. *medRxiv*, 2020.

[48] N. H. Bshouty and C. A. Haddad-Zaknoon. On detecting some defective items in group testing. *CoRR*, abs/2307.04822, 2023. https://doi.org/10.48550/arXiv.2307.04822. and *Journal of Combinatorial Optimization*, 51(5):53, 2026. DOI: https://doi.org/10.1007/S10878-026-01394-8.

[49] N. H. Bshouty, T.-M. Cheung, G. Harcos, H. Hatami, and A. Ostuni. A tight lower bound on non-adaptive group testing estimation. *Discrete Applied Mathematics*, 366:1–15, 2025.