This is a repository copy of *Performance of Group Testing Algorithms With Near-Constant Tests-per-Item*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/137695/

Version: Accepted Version

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Performance of Group Testing Algorithms With Near-Constant Tests-per-Item

Oliver Johnson, Matthew Aldridge, and Jonathan Scarlett

Abstract—We consider the nonadaptive group testing with N items, of which $K = \Theta(N^\theta)$ are defective. We study a test design in which each item appears in nearly the same number of tests. For each item, we independently pick L tests uniformly at random with replacement, and place the item in those tests. We analyse the performance of these designs with simple and practical decoding algorithms in a range of sparsity regimes, and show that the performance is consistently improved in comparison with standard Bernoulli designs. We show that our new design requires 23% fewer tests than a Bernoulli design when paired with the simple decoding algorithms known as COMP and DD. This gives the best known nonadaptive group testing performance for $\theta > 0.43$, and the best proven performance with a practical decoding algorithm for all $\theta \in (0, 1)$. We also give a converse result showing that the DD algorithm is optimal for these designs when $\theta > 1/2$. We complement our theoretical results with simulations that show a notable improvement over Bernoulli designs in both sparse and dense regimes.

I. Introduction and Defintions

In group testing, there is a population of items, some of which are 'defective' in some sense. We test subsets of items called 'pools'. In the standard noiseless case we consider in this paper, a test outcome is negative if every item the the pool is nondefective, and is positive if at least one item is defective. Through many such pooled tests, we hope to be able to accurately estimate which items are defective.

The group testing problem was introduced by Dorfman [2], as described in [3, Ch. 1.1]. While a wide variety of problem setups have been considered, they all share common features, and can be considered in a wider class of sparse inference problems including compressed sensing [4]. Group testing has been applied in a wide variety of contexts, including biology [5]–[9], anomaly detection in networks [10], [11], signal processing and data analysis [12], [13], and communications [14]–[16] – although this list is far from exhaustive.

In this paper, we prove rigorous performance bounds for the nonadaptive noiseless group testing problem. 'Nonadaptive' means that the make-up of every test pool is decided on advance, so tests can be performed in parallel. In the common Bernoulli design, each item is placed in each test independently with some fixed probability p. We instead

consider an alternative test design that we call the 'near-constant column weight' design. Here, independently for each item, we choose L tests uniformly at random with replacement and place the item in those tests. We pair our design with practical algorithms for detecting the defective items. Using both rigorous asymptotic results and experimental simulations, we shall see that we can accurately detect the defective items with considerably fewer tests than the Bernoulli design.

We proceed by formalizing the problem and fixing some notation. We have a large number of items N, of which K are defective. We assume that the defective items are rare, with $K = o(N)$ as $N \to \infty$; moreover, for concreteness we follow [1], [17], [18] by taking $K = \Theta(N^\theta)$ for some fixed parameter $\theta \in (0, 1)$. We follow the 'combinatorial model' and suppose that $\mathcal{K}$, the true set of defective items, is chosen uniformly at random from the $\binom{N}{K}$ sets of this size.

We perform a sequence of nonadaptive tests to form an estimate $\widehat{\mathcal{K}}$ of $\mathcal{K}$, and study the tradeoff between maximising the success probability $\mathbb{P}(\widehat{\mathcal{K}} = \mathcal{K})$ and minimising the number of tests T. We could simply take $T = N$, and test each item one by one. However, Dorfman's key insight [2] is that since the problem is sparse, in the sense that $K \ll N$, each test has a negative outcome with high probability, so these tests are not optimally informative. A better procedure considers a series of pools of items that are tested together, where the outcome of each test is positive if and only if it contains at least one defective item.

A group testing procedure requires two parts. First, a *test design* describes which items will be placed in which testing pools. Second, a *decoding algorithm* uses the results of these tests to estimate which items are defective.

Definition 1: We represent the testing pools by a (possibly random) binary matrix $\mathsf{X} \in \{0, 1\}^{T \times N}$, where $x_{ti} = 1$ if test t includes item i and $x_{ti} = 0$ otherwise. The rows of X correspond to tests, and the columns correspond to items.

Definition 2: We consider the standard *noiseless group testing model*. The outcomes of each test are represented by a binary vector $\mathbf{y} = (y_t) \in \{0, 1\}^T$, where a positive outcome $y_t = 1$ occurs if $x_{ti} = 1$ for some $i \in \mathcal{K}$, which is if the test contains a defective item. A negative outcome $y_t = 0$ occurs otherwise.

A commonly used test design is the Bernoulli design – see, for example, [17]–[23].

Definition 3: We define the *Bernoulli testing design* as having a testing matrix X in which each entry x_{ti} is independently set to be 1 with probability p and 0 otherwise, for some fixed parameter $p \in (0, 1)$.

In this paper, we will demonstrate that better performance can be achieved by using a design we call the near-constant column weight design.

Definition 4: We define the *near-constant column weight testing design* as having a testing matrix X in which L entries of each column of are selected uniformly at random with replacement and set to 1, with independence between columns. The remaining entries of X are set to 0. We set $L = \nu T/K$ for some parameter $\nu > 0$.

We now need an algorithm to produce an estimate of the defective set. In analogy with channel coding, we can think of the defective set $\mathcal{K}$ as a 'message' to be decoded from the 'signal' **y**, so we refer to such an algorithm as a 'decoding algorithm'. (For more on connections between group testing and channel coding, see, for example, [4], [5], [17], [22], [24], [25].)

Definition 5: We estimate the defective set by $\widehat{\mathcal{K}} = \widehat{\mathcal{K}}(\mathsf{X}, \mathbf{y})$, and define the (average) success probability

$$\mathbb{P}(\text{suc}) = \frac{1}{\binom{N}{K}} \sum_{|\mathcal{K}|=K} \mathbb{P}(\widehat{\mathcal{K}} = \mathcal{K}),$$

where the probability is over the random test design X.

We will demonstrate the superiority of our new design with two simple decoding algorithms. We define the algorithms here, but postpone detailed discussion to Section III. First, the COMP (Combinatorial Orthogonal Matching Pursuit) algorithm is a very simple algorithm based on the fact that every item in a negative test is definitely nondefective.

Definition 6: The COMP algorithm is given as follows:
1) Mark each item that appears in a negative test as non-defective, and refer to every other item as a Possible Defective (PD) – we write $\mathcal{PD}$ for the set of such items.
2) Mark every item in $\mathcal{PD}$ as defective.

Second, the DD (Definite Defectives) algorithm builds on COMP to find items we can be certain are defective.

Definition 7: The DD algorithm is given as follows:
1) Mark each item that appears in a negative test as non-defective, and refer to every other item as a Possible Defective (PD).
2) For each positive test that contains a single Possible Defective item, mark that item as defective.
3) Mark all remaining items as non-defective.

The main results of this paper concern rigorous bounds on the performance of the near-constant column weight design with various decoding algorithms. We are interested in how many tests are required for the success probability to tend to 1 as N gets large. Specifically, we show the following:

- The COMP algorithm requires 23% fewer tests with a near-constant column weight design than with a Bernoulli design, which for $\theta > 0.77$ is fewer tests than required even for optimal algorithms with Bernoulli designs. (Theorem 2)
- The DD algorithm also requires 23% fewer tests with a near-constant column weight design than with a Bernoulli design, which for $\theta > 0.43$ is fewer tests than required even for optimal algorithms with Bernoulli designs, and for all $\theta \in (0, 1)$ is fewer tests than the best proven results

for practical algorithms with Bernoulli designs. (Theorem 3)
- We give an upper bound on the performance of near-constant column weight designs regardless of the decoding algorithm, showing that DD is optimal for this design when $\theta \geq 1/2$. (Theorem 4)
- We complement our rigorous theoretical results with simulations that show a notable improvement over Bernoulli designs in both sparse and dense regimes. (Subsection II-C)

The structure of the remainder of the paper is as follows. In Section II, we define the rate of group testing (Subsection II-A), formally state our main results of the paper (Subsection II-B), provide simulation results to illustrate the improved performance of our test design (Subsection II-C), and briefly discuss some related work (Subsection II-D). In Section III, we describe the main decoding algorithms used in more detail and introduce some key quantities that control their performance. In Section IV, we deduce the main theorems of the paper, with proofs of some techinical results given the appendices.

II. FURTHER DEFINITIONS AND MAIN RESULTS

A. The rate of group testing

In this paper, we focus on nonadaptive designs, where the entire matrix X is fixed in advance of the tests. In the adaptive case (where the members of each test are chosen using the outcomes of the previous tests), Hwang's generalised binary splitting algorithm [26] recovers the defective set $\mathcal{K}$ using $\log_2 \binom{N}{K} + O(K)$ tests. This can be seen to be essentially optimal by a standard argument based on Fano's inequality (see for example [23]), a strengthened version of which [25] implies that any algorithm using T tests has success probability bounded above by

$$\mathbb{P}(\text{suc}) \leq \frac{2^T}{\binom{N}{K}}. \tag{1}$$

This means that any algorithm with success probability $\mathbb{P}(\text{suc})$ tending to 1 requires at least

$$T = \log_2 \binom{N}{K} \sim K \log_2 \frac{N}{K} \sim (1 - \theta)K \log_2 N \tag{2}$$

tests, where $f(N) \sim g(N)$ means that $\lim_{N \to \infty} f(N)/g(N) = 1$. (See [21, Lemma 25] for details of the asymptotic behaviour of the binomial coefficient.) This motivates the following definition [21] of the rate of an algorithm.

Definition 8: For any algorithm using T tests, we define the *rate* to be

$$\frac{\log_2 \binom{N}{K}}{T}. \tag{3}$$

Given a random matrix design, we say that R is an *achievable rate* if for any $\epsilon > 0$, there exists a group testing algorithm with rate converging to R and success probability at least $1 - \epsilon$ for N sufficiently large. We adopt the terminology *maximum achievable rate* when referring to a given design (e.g., Bernoulli) and/or decoding rule (e.g., COMP).

Intuitively, one can think of the rate as being the number of bits of information learned per test when the recovery is successful.

In this language, the result of [26] shows that, for all $\theta \in (0, 1)$, adaptive group testing has an achievable rate of $R = 1$ in the regime $K = \Theta(N^\theta)$ and is therefore optimal, since by (1), no algorithm can learn more than 1 bit per test. It is an interesting question to consider whether there exists a matrix design and an algorithm with achievable rate $R = 1$ in the nonadaptive case. It appears to be difficult even to design a class of matrices with non-zero achievable rate using combinatorial constructions (see [3], [24] for reviews of the extensive literature on this subject, with key early contributions coming from [27]–[29]). Hence, much recent work on nonadaptive group testing has considered Bernoulli designs (see Definition 3). The maximum achievable rate is known exactly for such designs, as stated in the following.

Theorem 1: The maximum achievable rate for Bernoulli nonadaptive group testing with $K = \Theta(N^\theta)$ defectives, for $\theta \in [0, 1)$, is

$$C(\theta) = \max_{\nu > 0} \min \left\{ \frac{\nu e^{-\nu}}{\ln 2} \frac{1 - \theta}{\theta}, \, h(e^{-\nu}) \right\}, \tag{4}$$

where $h(t) = -t \log_2 t - (1 - t) \log_2(1 - t)$ is the binary entropy function. In particular, for $\theta \leq 1/3$, the maximum achievable rate of Bernoulli designs is 1.

The direct part of Theorem 1 is due to [17] and the converse due to [19]. (The special case $\theta = 0$ is older [30].)

The curve (4) is illustrated in Figure 1 below. For $\theta \geq 1/2$, the paper [21] showed that (4) is achieved by the DD algorithm described above. However, for $\theta < 1/2$, the algorithms known to achieve the bound (4) are based on maximising the likelihood or solving other difficult combinatorial problems, and cannot be considered as practical in a computational sense – see Section III-D for more details. For example, we describe the SSS algorithm in Definition 10 below, which achieves the bound of [19], but is impractical for large values of N and K.

B. Main results

Our main results concern improving on Theorem 1 by using a near-constant column weight design (Defintion 4). Recall that this design has a testing matrix X in which $L = \nu T/K$ entries of each column of are selected uniformly at random with replacement and set to 1, with independence between columns, and the remaining entries of X are set to 0. The tester may choose L to depend on the parameters of the group testing problem.

Since the tests are chosen *with replacement*, some columns may actually have weight slightly less than L due to the same test being picked more than once, hence we use the term 'near-constant'. Since the weight of a column is the number of tests an item is in, we also consider these designs as 'near-constant tests-per-item'. In a preliminary report [1], we used the less precise terminology 'constant column weight' for these same designs. Evidence from simulations and heuristic calculations suggest that truly-constant column weight designs have the same performance as the near-constant designs we consider

here, but the rigorous analysis of such designs seems more difficult.[1]

The main results of this paper are the following three theorems. The COMP and DD algorithms were defined in Defintions 6 and 7 and are discussed further in Section III. The main results are proved in Section IV. We illustrate these results in Figure 1, with our new rates for near-constant column weight designs marked in thick red, and corresponding rates for Bernoulli designs marked in thin blue.

Our first result concerns the simple and practical COMP decoding algorithm (see Defintion 6 and Subsection III-A), based on the fact that all items in negative tests must be negative.

Theorem 2: Consider a near-constant column weight design with an optimised parameter $\nu > 0$. When there are $K = \Theta(N^\theta)$ defectives for $\theta \in [0, 1)$, the COMP algorithm has success probability tending to 1 if $T \geq (1 + \epsilon)T^{\text{COMP}}$, and tending to 0 if $T \leq (1 - \epsilon)T^{\text{COMP}}$. Here

$$T^{\text{COMP}} = \frac{1}{\ln 2} K \log_2 N.$$

Hence, the COMP algorithm has maximum achievable rate $\ln 2 \, (1 - \theta)$ for all $\theta \in [0, 1)$.

This rate

$$\ln 2 \, (1 - \theta) \approx 0.693(1 - \theta)$$

is an improvement by 30.6% on the rate

$$\frac{1}{e \ln 2}(1 - \theta) \approx 0.531(1 - \theta)$$

for COMP with a Bernoulli design [19], [23]. (An improvement in rate of 30.6% corresponds to using 23.4% fewer tests.) Further, for $\theta > 0.766$, Theorem 2 is an improvement on (4), meaning that in dense cases, the very simple COMP algorithm with a near-constant column weight design beats *any* decoding algorithm with a Bernoulli design (see Figure 1).

Some insight on this result can be attained by considering the conditions under which COMP succeeds. Under the choice $\nu = \ln 2$, Bernoulli testing with probability ν/K and a near-constant column weight design with $L = \nu T/K$ both result in roughly half of the tests being positive (e.g., see Lemma 1 below). However, a given non-defective item is placed in roughly $\text{Binomial}\left(\frac{T}{2}, \frac{\ln 2}{K}\right)$ negative tests under Bernoulli testing, and $\text{Binomial}\left(\frac{T \ln 2}{K}, \frac{1}{2}\right)$ negative tests under the near-constant column weight design. While these two distributions have the same expectation, the latter has a much smaller probability of being zero, which is the event under which COMP fails.

It is also interesting to note that while $\nu = \ln 2$ (which is 'maximally informative' in the sense of maximising the entropy of the test outcome) optimises the rate of COMP (as well as DD below) for the near-constant column weight design, COMP [23] and DD [21] with Bernoulli designs are optimised with a fraction $1 - e^{-1} \approx 0.632$ of positive tests.

---

[1] As pointed out by a reviewer, the COMP rate in Theorem 2 can be shown to be achieved by a truly-constant column weight design with little extra difficulty. However, we have not been able to rigorously verify that the same is true for our main result, Theorem 3, or for the algorithm-independent converse, Theorem 4.
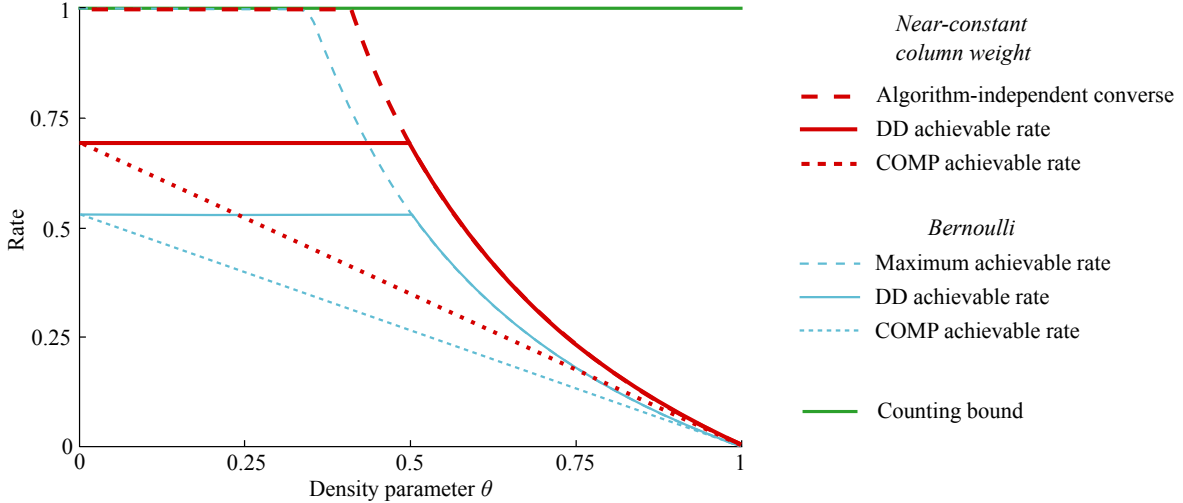
Fig. 1. Rates and bounds for group testing algorithms with Bernoulli designs and near-constant column weight designs. In thick red, we plot the rate bounds for near-constant column weight designs from Theorems 2, 3 and 4. In thin blue, we plot the rate bounds for Bernoulli designs from Theorem 1 and [21]. The green horizontal line represents the universal 'counting bound' arising from (1).

Our second and most important result concerns the practical DD decoding algorithm (see Definition 7 and Subsection III-B).

Theorem 3: Consider a near-constant column weight design with an optimized parameter $\nu > 0$. When there are $K = \Theta(N^\theta)$ defectives for $\theta \in (0, 1)$, the DD algorithm has success probability tending to 1 if

$$T \geq (1 + \epsilon)\frac{1}{\ln 2} \max\left\{K \log_2 \frac{N}{K}, K \log_2 K\right\},$$

and hence has an achievable rate

$$R = \ln 2 \min\left\{1, \frac{1 - \theta}{\theta}\right\} = \begin{cases} \ln 2 & \theta \leq \frac{1}{2} \\ \ln 2 \frac{1 - \theta}{\theta} & \theta > \frac{1}{2}. \end{cases}$$

This rate

$$\ln 2 \min\left\{1, \frac{1 - \theta}{\theta}\right\} \approx 0.693 \min\left\{1, \frac{1 - \theta}{\theta}\right\}$$

is an improvement again by 30.6% on the rate of

$$\frac{1}{e \ln 2} \min\left\{1, \frac{1 - \theta}{\theta}\right\} \approx 0.531 \min\left\{1, \frac{1 - \theta}{\theta}\right\}$$

proved by [21] for DD with Bernoulli designs. In fact, to our knowledge, DD with the near-constant column weight design gives the highest proven *practically achievable* rate for all $\theta \in (0, 1)$. Further, for $\theta > 1/(1 + e(\ln 2)^2) \approx 0.434$, Theorem 3 is an improvement on (4), meaning that in this regime, the practical DD algorithm with a near-constant column weight design beats any decoding algorithm (even impractical ones) with a Bernoulli design.

Our third result is an algorithm-independent converse, showing the maximum possible rate of any decoding algorithm with a near-constant weight design.

Theorem 4: Consider a near-constant column weight design, with $K = \Theta(N^\theta)$ defectives for $\theta \in (0, 1)$. Regardless of the choice of $\nu > 0$, no algorithm can achieve a rate greater than

$$\min\left\{1, \ln 2 \frac{1 - \theta}{\theta}\right\} = \begin{cases} 1 & \theta \leq \theta^* \\ \ln 2 \frac{1 - \theta}{\theta} & \theta > \theta^*, \end{cases} \quad (5)$$

where

$$\theta^* = \frac{\ln 2}{1 + \ln 2} \approx 0.409.$$

Comparing Theorems 3 and 4, we see that if we use a near-constant column weight design, the DD algorithm gives the optimal performance for $\theta \geq 1/2$.

C. Simulations

We complement our rigorous results on the rate, which are asymptotic as $N \to \infty$, with simulations that show near-constant weight designs also improve on Bernoulli designs for finite problem sizes. In Figure 2, we illustrate the performance of these algorithms via simulations in an illustrative sparse case ($N = 500$, $K = 10$) and a denser case ($N = 2000$, $K = 100$). For the sparse case, in addition to plotting performance of COMP and DD, we plot the performance of the SSS algorithm (see Definition 10), which achieves the bounds of Theorem 1 [20, Corollary 4], though is not practical for larger problems. Because of this issue of practicality, we do not consider SSS for the denser case; instead, we plot the performance of a related algorithm called SCOMP, which is described in [21], so we omit a description in this paper for the sake of brevity. (Essentially, it amounts to performing DD followed by greedy refinements.) Our near-constant weight designs provide a consistent notable improvement on Bernoulli designs, particularly in the denser example.

D. Related work

While we are not aware of previous works using our exact near-constant column weight design, closely-related designs
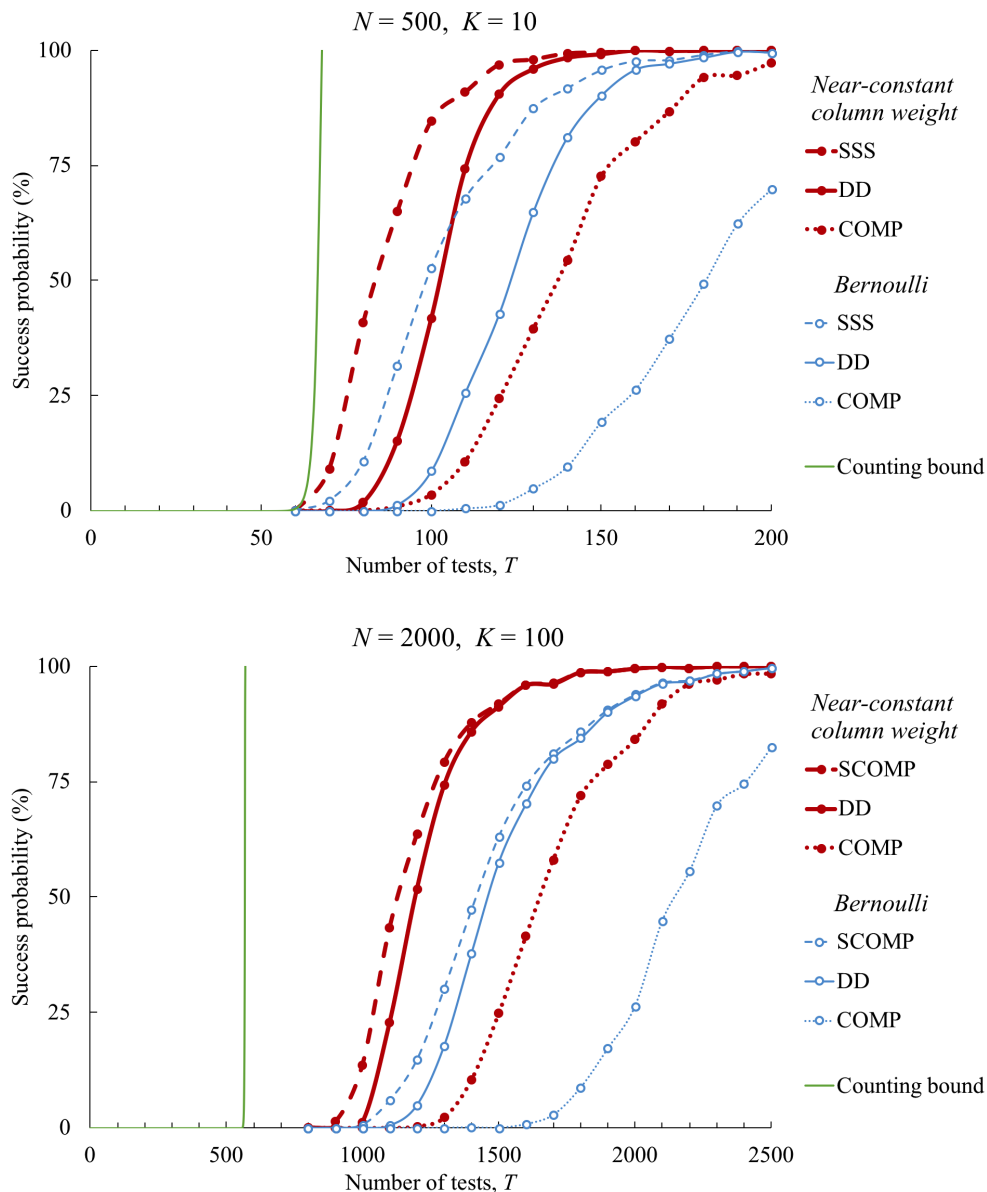
Fig. 2. Empirical performance (each point based on 1000 simulations) of various algorithms for both near-constant column weight and Bernoulli designs, in the cases $N = 500$, $K = 10$ and $N = 2000$, $K = 100$.

have been proposed. Our key contribution is a *rigorous* analysis of such designs in the regime $K = \Theta(N^\theta)$, requiring several novel techniques. In particular, we prove the achievability of rates strictly above those achieved by Bernoulli designs.

Kautz and Singleton [31] observed that a construction based on a concatenation of constant-weight codes gives matrices with the so-called K-disjunctness property (the union of any K columns does not contain any other column). Such matrices give group testing designs guaranteeing that K defectives can be recovered with zero probability of error in noiseless group testing (see for example [3, Chapter 7]). However, the group testing designs resulting from the construction of [31] require $T = O(K^2 (\log N)^2)$ tests. This is an example of the fact that the zero-error criterion requires considerably more tests than the $T = O(K \log N)$ required for the 'error probability tending to zero' criterion (see Definition 8) that we study here.

Similarly, other subsequent papers have proposed forms of constant or near-constant column weight designs [32]–[35], but to our knowledge, none of these works provide non-trivial achievable rates for the vanishing error probability criterion, which is the focus of the present paper. Chan *et al.* [23] considered constant *row* weight designs, and found no improvement over Bernoulli designs.

Mézard *et al.* [36] considered randomised designs with both constant row and column weights, and with constant column weights only. The paper used heuristics from statistical physics to suggest that such designs may beat Bernoulli designs. In our notation, they suggest the maximum achievable rate of these constant weight designs may be equal to our converse bound (5) for all θ. (Our Theorem 3 rigorously proves this for $\theta \geq 1/2$ under our slightly different design.) The work of [36] contains some non-rigorous steps; in particular, they make use

of a 'no short loops' assumption that is only verified for $\theta > \frac{5}{6}$ and conjectured for $\theta > \frac{2}{3}$, while experimentally being shown to *fail* for smaller values such as $\theta = \frac{1}{3}$.

D'yachkov *et al.* [37] studied list decoding with (exactly) constant column weight designs, and setting their list size to 1 corresponds to insisting that COMP succeeds. However, they only considered the case that $K = O(1)$. In the limit as K gets large, the rate $\ln 2$ obtained [37, Claim 2] matches the rate for COMP given here in Theorem 2 in the limit $\theta \to 0$.

A distinct line of works has sought designs that not only require a low number of tests, but also near-optimal *decoding complexity* (e.g., $K\mathrm{poly}(\log N)$) [38]–[41]. However, our focus in this paper is on the required number of tests, for which the existing guarantees of such algorithms contain loose constants or extra logarithmic factors.

III. DECODING ALGORITHMS: FURTHER DETAILS

In this section we discuss the COMP and DD algorithms in more detail, and introduce the SSS (Smallest Satisfying Set) algorithm. We discuss the conditions under which these algorithms succeed. These algorithms were previously studied in [23] and [21] with Bernoulli designs. Before continuing, we present another key definition:

Definition 9: Consider an item i and a set of items $\mathcal{L}$ not including i. We say that item i is *masked* by $\mathcal{L}$ if every test that includes i also includes at least one member of $\mathcal{L}$.

A. COMP algorithm

Recall the COMP algorithm from Definition 6:

1) Mark each item that appears in a negative test as non-defective, and refer to every other item as a Possible Defective (PD) – we write $\mathcal{PD}$ for the set of such items.
2) Mark every item in $\mathcal{PD}$ as defective.

This is based on a simple inference: Any negative test only contains non-defective items, so any item in a negative test can be marked as non-defective. Given enough negative tests, we might hope to correctly infer every member of $\mathcal{K}^{\mathsf{c}}$ in this way. The name COMP (Combinatorial Orthogonal Matching Pursuit) was coined in [23], although the method itself appeared much earlier – see, for example, [5], [24], [31], [42].

Clearly, the first step will not make any mistakes (every item marked as non-defective will indeed be non-defective), so errors will only occur in the second step. As a result COMP will always estimate $\mathcal{K}$ by a set $\widehat{\mathcal{K}}_{\mathrm{COMP}}$ with $\mathcal{K} \subseteq \widehat{\mathcal{K}}_{\mathrm{COMP}}$.

As in [21], a quantity of particular interest is $G := |\mathcal{PD} \setminus \mathcal{K}| = |\mathcal{PD}| - K$, the number of non-defective items masked by the defective set $\mathcal{K}$. So G is the number of non-defective items that do not appear in any negative test. It is clear that COMP succeeds (recovers the defective set exactly) if and only if $G = 0$, so that

$$\mathbb{P}^{\mathrm{COMP}}(\mathrm{suc}) = \mathbb{P}(G = 0). \qquad (6)$$

We use this in the proof of Theorem 2 in Section IV.

B. DD algorithm

Recall the DD algorithm from Definition 7, which builds on COMP to find items that are definitely defective:

1) Mark each item that appears in a negative test as non-defective, and refer to every other item as a Possible Defective (PD).
2) For each positive test that contains a single Possible Defective item, mark that item as defective.
3) Mark all remaining items as non-defective.

The performance of the DD algorithm with Bernoulli designs was studied in detail by Aldridge, Baldassini and Johnson [21].

Again, the first step will not make any mistakes, and since every positive test must contain at least one defective item, the second step is also certainly correct. Hence, any errors due to DD come from marking a true defective as non-defective in the third step, meaning that the estimate $\widehat{\mathcal{K}}_{\mathrm{DD}}$ satisfies $\widehat{\mathcal{K}}_{\mathrm{DD}} \subseteq \mathcal{K}$. The choice to mark all remaining items as non-defective is motivated by the sparsity of the problem, since *a priori* an item is much less likely to be defective than non-defective.

We analyse DD rigorously in Section IV, using the following notation, used in [21]. For each $i \in \mathcal{K}$, we write:

- M_i for the number of tests containing defective item i and no other defective;
- L_i for the number of tests containing defective item i and no other possible defective item (no other member of $\mathcal{PD}$).

In the terminology of Definition 9, we see that DD succeeds if and only if no defective item $i \in \mathcal{K}$ is masked by $\mathcal{PD} \setminus \{i\}$. Further, since item i is masked by $\mathcal{PD} \setminus \{i\}$ if and only if $L_i = 0$, we can write

$$\mathbb{P}^{\mathrm{DD}}(\mathrm{suc}) = 1 - \mathbb{P}\left(\bigcup_{i \in \mathcal{K}} \{L_i = 0\}\right). \qquad (7)$$

For a given defective item $i \in \mathcal{K}$, we write $\mathcal{K}^{(i)} = \mathcal{K} \setminus \{i\}$ for the set of defectives with i removed. For a given set $\mathcal{M}$, we write $W^{(\mathcal{M})}$ for the total number of tests containing at least one item from $\mathcal{M}$. The random variable $W^{\mathcal{K} \setminus \{i\}}$ (the total number of tests containing at least one item in $\mathcal{K}^{(i)}$), henceforth denoted by $W^{(\mathcal{K} \setminus i)}$, will be of particular interest.

To understand the distributions of these quantities, it is helpful to think of the process by which elements of the columns are sampled as a *coupon collector* problem, where each coupon corresponds to one of the T tests. For a single item, $W^{(\{i\})}$ is the number of distinct coupons selected when L coupons are chosen uniformly at random from a population of T coupons. In general, for a set $\mathcal{M}$ of size M, the independence of distinct columns means that $W^{(\mathcal{M})}$ is the number of distinct coupons collected when choosing ML coupons uniformly at random from a population of T coupons.

Hence, as described in more detail in Section IV, we can first give a concentration of measure result for $W^{(\mathcal{K} \setminus i)}$ (see Lemma 1), then characterise the distribution of M_i given $W^{(\mathcal{K} \setminus i)}$ (see Lemma 4). Following this, we can state the distribution of G conditioned on $W^{(\mathcal{K})} = W^{(\mathcal{K} \setminus i)} + M_i$ (see Lemma 5), and finally deduce the distribution of L_i conditioned on G and $W^{(\mathcal{K})}$ (see Lemma 6). This allows us to deduce bounds on (7).

C. SSS algorithm

We describe one more algorithm, which we call the SSS (Smallest Satisfying Set) algorithm, following [21]. This algorithm is not directly mentioned in the statement of our main results, but its analysis will be important in proving Theorem 4.

Definition 10: We say that a putative defective set $\mathcal{J}$ is *satisfying* if:

1) No negative test contains a member of $\mathcal{J}$.
2) Every positive test contains at least one member of $\mathcal{J}$.

The SSS algorithm simply finds the smallest satisfying set (breaking ties arbitrarily), and takes that as the estimate $\widehat{\mathcal{K}}_{\mathrm{SSS}}$.

Note that the true defective set $\mathcal{K}$ is certainly a satisfying set, and hence SSS is guaranteed to return a set of no larger size, so $|\widehat{\mathcal{K}}_{\mathrm{SSS}}| \leq |\mathcal{K}|$. However, it may not be the case that $\widehat{\mathcal{K}}_{\mathrm{SSS}} \subseteq \mathcal{K}$. We can identify a particular failure event for SSS: If a defective item $i \in \mathcal{K}$ is masked by the other defective items $\mathcal{K} \setminus \{i\}$ (in the sense of Definition 9) then $\mathcal{K} \setminus \{i\}$ will be a smaller satisfying set, so SSS is certain to fail.

Hence, writing A_i for the event that item i is masked by $\mathcal{K} \setminus \{i\}$, we can use the Bonferroni inequality to obtain a lower bound on the SSS error probability $\mathbb{P}^{\mathrm{SSS}}(\mathrm{err})$ of the form

$$\mathbb{P}^{\mathrm{SSS}}(\mathrm{err}) \geq \mathbb{P}\left(\bigcup_{i \in \mathcal{K}} A_i\right) \geq \sum_{i \in \mathcal{K}} \mathbb{P}(A_i) - \frac{1}{2} \sum_{i \neq j \in \mathcal{K}} \mathbb{P}\left(A_i \cap A_j\right). \tag{8}$$

This serves as a starting point for upper bounding the rate of the SSS algorithm, which in turn will be used to infer our general converse (Theorem 4).

D. Note on practical feasibility

We refer to COMP and DD as 'practical' algorithms, since they can be implemented with low run-time and storage. For example, COMP simply requires us to take one pass through the test matrix and outcomes, requiring no more than $O(N)$ storage beyond the matrix itself, and $O(TN)$ runtime. Similarly, DD builds on COMP, requiring two passes through the test matrix and outcomes and can be performed with the same amount of storage and runtime.

In contrast, we can interpret SSS as an integer programming problem, meaning that it is unlikely to be practical to run for large problems. We think of it as the 'best possible' algorithm without knowing K, and use a rigorous form of this statement [19] to obtain algorithm-independent performance bounds. Note that although the SSS algorithm may be considered to be infeasible in practice, the papers [43], [44] show that a relaxation of the integer programming problem to the real numbers can give good performance.

Furthermore, the decoding algorithms we consider here do not require exact, or even approximate, knowledge of K. This is in contrast to the optimal maximum likelihood decoder of [17], which requires the exact value of K. Note, however, that the optimal choice of the parameter $\nu = (\ln 2)T/K$ in the *design* stage does require knowing K.

IV. PROOFS OF MAIN RESULTS

The main goal of this section is to prove our achievable rate for the DD algorithm (Theorem 3). Along the way, we will also prove the COMP rate (Theorem 2) and the algorithm-independent upper bound (Theorem 4); the former will essentially come 'for free', though the latter will require non-trivial additional effort.

A. Concentration of $W^{(\mathcal{M})}$

Recall that $W^{(\mathcal{M})}$ corresponds to the total number of tests in which items from $\mathcal{M}$ are placed. The following lemma shows that this quantity concentrates around its mean.

Lemma 1: Let $M = |\mathcal{M}|$, and fix the constants $\alpha > 0$ and $\epsilon \in (0, 1)$. When making $LM = \alpha T$ draws with replacement from a total of T coupons, the total number of distinct coupons $W^{(\mathcal{M})}$ satisfies

$$\mathbb{P}\left(\left|W^{(\mathcal{M})} - (1 - e^{-\alpha})T\right| \geq \delta\right) \leq 2\exp\left(-\frac{\delta^2}{\alpha T}\right) \tag{9}$$

for T sufficiently large.

Proof: We first characterise the expectation of $W^{(\mathcal{M})}$, and then show concentration about that expectation. By the linearity of expectation, we have

$$\mathbb{E}W^{(\mathcal{M})} = \sum_{j=1}^{T} \mathbb{P}(\text{coupon } j \text{ in first } LM \text{ selections})$$

$$= \sum_{j=1}^{T} \left(1 - \left(1 - \frac{1}{T}\right)^{LM}\right)$$

$$= \left(1 - \left(1 - \frac{1}{T}\right)^{\alpha T}\right)T.$$

It follows that $\mathbb{E}W^{(\mathcal{M})} = (1 - e^{-\alpha})T + o(T)$ as $T \to \infty$.

To establish concentration about the mean, we use Mc-Diarmid's inequality [45], which characterises the concentration of functions of independent random variables when the bounded difference property is satisfied. Write $Y_1, Y_2, \ldots, Y_c$ for the labels of the selected coupons, and $W(c) = f(Y_1, Y_2, \ldots, Y_c)$ for the number of distinct coupons. Note that here we have the bounded difference property, in that

$$\left|f(Y_1, \ldots, Y_j, \ldots, Y_c) - f(Y_1, \ldots, \hat{Y}_j, \ldots, Y_c)\right| \leq 1$$

for any j, $Y_1, \ldots, Y_c$, and $\hat{Y}_j$, since the largest difference we can make is swapping a distinct coupon Y_j for a non-distinct one $\hat{Y}_j$, or vice versa. McDiarmid's inequality [45] gives that

$$\mathbb{P}\left(\left|f(Y_1, \ldots, Y_c) - \mathbb{E}f(Y_1, \ldots, Y_c)\right| \geq \delta\right) \leq 2\exp\left(-\frac{2\delta^2}{c}\right).$$

Setting $c = \alpha T$ gives the desired result; we crudely remove the factor of 2 from the exponent to account for the fact that we are considering deviations from the asymptotic value of the mean of $W^{(\mathcal{M})}$ rather than the exact value, which amounts to the replacement of α by $\alpha(1 + o(1))$. ∎

B. Proof of the algorithm-independent converse

The above concentration result plays an important role in the proof of Theorem 4.

Proof of Theorem 4: We divide the proof into three steps. First, we begin with an overview of some preliminary results that will be used throughout the proof; second, we bound the error probability of the SSS algorithm; and third, we bound a key quantity that arises in the proof.

1) Preliminaries: The initial steps follow the proof of a similar result for Bernoulli testing in [19]. As shown there, if $\mathbb{P}^{\text{SSS}}(\text{err}) + \mathbb{P}^{\text{COMP}}(\text{err}) > 1 + \epsilon$ for some $\epsilon > 0$ that remains bounded away from zero as $N \to \infty$, then the error probability is also bounded away from zero for an arbitrary algorithm. We know the condition under which $\mathbb{P}^{\text{COMP}}(\text{err}) \to 1$ from Theorem 2 (which will be proved later), and it is easy to see that the corresponding bound is weaker than that of Theorem 4, since

$$1 - \theta \le \min\left\{1, \frac{1-\theta}{\theta}\right\}.$$

Hence, it suffices to show that the error probability of the SSS algorithm is bounded away from zero; we do so in the remainder of the proof.

The upper bound of 1 on the rate is well-known for arbitrary test designs (this follows from (1), for example), so we only need to obtain the other term in (5). To do so, we claim that it suffices to show that for any choice of $\nu > 0$ (such that $L = \nu T/K$) the error probability is bounded away from zero for some T satisfying

$$T = \frac{K \ln K}{-\nu \ln(1 - e^{-\nu})}(1 + o(1)). \tag{10}$$

To see that it suffices to choose T in this way, first note that $-\nu \ln(1 - e^{-\nu})$ attains its maximum of $(\ln 2)^2$ at $\nu = \ln 2$, in which case the rate corresponding to (10) is $\ln 2 \frac{1-\theta}{\theta}$, as required. For other choices of ν, the choice (10) corresponds to more tests than dictated by the rate of Theorem 4, but this is allowed for the purpose of proving a converse, since additional tests can never hurt the SSS algorithm.

We can also assume that ν is constant, since it is straightforward to verify that the cases $\nu \to 0$ or $\nu \to \infty$ fail to even yield the correct scaling $T = \Theta(K \ln N)$. This is because, in such cases, the probability of a given test being positive tends to either 0 or 1, and hence the entropy of the test vanishes.

Finally, we note the following concentration result: Lemma 1 above shows that for both $M = K - 2$ and $M = K - 1$, choosing $\alpha = LM/T = \nu M/K \to \nu$ reveals that $W^{(\mathcal{M})}$ is exponentially concentrated around $(1 - e^{-\nu})T$. In particular, there exists a constant $c_0 > 0$ such that

$$\Pr\left(\left|W^{(\mathcal{M})} - (1 - e^{-\nu})T\right| \ge \sqrt{c_0 T \ln T}\right) \le \frac{1}{T^3} \tag{11}$$

for sufficiently large T.

2) Bounding the error probability of SSS: We start with the lower bound on the error probability given in (8), which we repeat here for convenience:

$$\mathbb{P}^{\text{SSS}}(\text{err}) \ge \mathbb{P}\left(\bigcup_{i \in \mathcal{K}} A_i\right) \ge \sum_{i \in \mathcal{K}} \mathbb{P}(A_i) - \frac{1}{2} \sum_{i \ne j \in \mathcal{K}} \mathbb{P}(A_i \cap A_j). \tag{12}$$

We will show that for any constant $\nu > 0$, the right-hand side of (12) is bounded away from zero as $N \to \infty$ under some number of tests T satisfying (10). We begin by bounding the two terms on the right-hand side of (12).

Lemma 2: Under the preceding definitions, and under a near-constant column weight design with parameter $\nu > 0$, we have for any constants $c_1, c_2 > 0$ and $\epsilon_1 \in (0, 1)$ that

$$K\mathbb{P}(A_i) \ge Kc_1^L \, \mathbb{P}\left(W^{(\mathcal{K}\setminus i)} \ge Tc_1\right) \tag{13}$$

$$\binom{K}{2}\mathbb{P}(A_i \cap A_j) \le \frac{K^2}{2(1-\epsilon_1)}\left(c_2 + \frac{\nu}{K}\right)^{2L} \mathbb{P}\left(Tc_1 \le W^{(\mathcal{K}\setminus i,j)} \le Tc_2\right)$$
$$+ \binom{K}{2}\mathbb{P}\left(W^{(\mathcal{K}\setminus i,j)} < Tc_1\right) + \binom{K}{2}\mathbb{P}\left(W^{(\mathcal{K}\setminus i,j)} > Tc_2\right) \tag{14}$$

when N is sufficiently large.

The idea of the proof is to lower bound $\mathbb{P}(A_i)$ by restricting attention to the event $W^{(\mathcal{K}\setminus i)} \ge Tc_1$ and applying counting arguments, and to upper bound $\mathbb{P}(A_i \cap A_j)$ by one when the suitable bounds on $W^{(\mathcal{K}\setminus i,j)}$ fail to hold, while upper bounding it using counting arguments otherwise. The details are given in Appendix C.

From (11), if we choose

$$c_1 = 1 - e^{-\nu} - \sqrt{\frac{c_0 \ln T}{T}} \tag{15}$$

$$c_2 = 1 - e^{-\nu} + \sqrt{\frac{c_0 \ln T}{T}}, \tag{16}$$

then, recalling from (10) that $T = \Theta(K \ln N)$, the final two terms in (14) vanish at rate $O(T^{-1})$ as $N \to \infty$. Thus, overall, (13) and (14) simplify to

$$K\mathbb{P}(A_i) \ge Kc_1^L - O\left(\frac{1}{T}\right) \tag{17}$$

$$\binom{K}{2}\mathbb{P}(A_i \cap A_j) \le \frac{K^2}{2(1-\epsilon_1)}\left(c_2 + \frac{\nu}{K}\right)^{2L} + O\left(\frac{1}{T}\right). \tag{18}$$

Combining these, we find that (12) yields

$$\mathbb{P}^{\text{SSS}}(\text{err}) \ge Kc_1^L\left(1 - \frac{K(c_2 + \nu/K)^{2L}}{2c_1^L(1-\epsilon_1)}\right) - O\left(\frac{1}{T}\right) \tag{19}$$

$$= Kc_1^{\nu T/K}\left(1 - \frac{K(c_2 + \nu/K)^{2\nu T/K}}{2c_1^{\nu T/K}(1-\epsilon_1)}\right) - O\left(\frac{1}{T}\right), \tag{20}$$

where we have used the fact that $L = \nu T/K$.

3) Bounding the right-hand side of (20): With the lower bound (20) on the error probability in place, the completion of the proof amounts to two simple but tedious steps:

1) Equate the large bracketed term with $1 - \frac{1}{2(1-\epsilon_1)} \approx \frac{1}{2}$, and show that solving for T yields an expression of the form (10);

2) Show that under any choice of T of the form (10), the remaining term $Kc_1^{\nu T/K}$ approaches 1 as $N \to \infty$.

These steps are summarised in the following lemma, whose proof is relegated to Appendix C.

Lemma 3: Choosing c_1 and c_2 as in (15) and (16), there exists a choice of T satisfying (10) for which (20) can be weakened to

$$\mathbb{P}^{\text{SSS}}(\text{err}) \geq (1 - o(1))\left(1 - \frac{1}{2(1 - \epsilon_1)}\right) - O\left(\frac{1}{T}\right). \qquad (21)$$

We conclude the proof of Theorem 4 by noting that the right-hand side can be made arbitrarily close to $\frac{1}{2}$ for sufficiently large K and T, since ϵ_1 can be chosen arbitrarily small. ∎

C. Conditional distributions of M_i and G

Recall that M_i denotes the number of tests containing defective item i but no other defective items. The following lemma gives the distribution of this quantity conditioned on $W^{(\mathcal{K}\backslash i)}$, the number of tests covered by those other defectives. It is written in terms of the following definition: For any integers n, k the *Stirling number of the second kind* is given by

$$\left\{ {n \atop k} \right\} := \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n, \qquad (22)$$

and equals the number of partitions of a set of size n into k nonempty subsets (see for example [46, eq. (8)]).

Lemma 4:

1) We can write the conditional distribution of $M_i \mid W^{(\mathcal{K}\backslash i)}$ explicitly as

$$\mathbb{P}(M_i = j \mid W^{(\mathcal{K}\backslash i)} = w)$$
$$= \frac{(T - w)_{(j)}}{T^L} \sum_{s=0}^{L-j} \binom{L}{s} \left\{ {L - s \atop j} \right\} w^s, \qquad (23)$$

where $(n)_{(j)} := (n)!/(n-j)!$ denotes the falling factorial.

2) For fixed L and w, there exists an explicit value $C = C(L, w) := \exp(L^2/4w)$, independent of j, such that

$$\mathbb{P}(M_i = j \mid W^{(\mathcal{K}\backslash i)} = w) \leq C \binom{L}{j} \left(1 - \frac{w}{T}\right)^j \left(\frac{w}{T}\right)^{L-j}.$$

That is, the probability is upper bounded by a multiple of the $\text{Bin}(L, 1 - w/T)$ mass function.

Proof: See Appendix A. ∎

Next, we observe that we can write $W^{(\mathcal{K})} = W^{(\mathcal{K}\backslash i)} + M_i$. Recall that G is the number of non-defectives masked by the defective set $\mathcal{K}$. Since an item is only counted in G if each of the tests appearing in the corresponding column are in the set of size $W^{(\mathcal{K})}$, we have the following.

Lemma 5: Conditional on $W^{(\mathcal{K})} = x$, we have

$$G \mid \{W^{(\mathcal{K})} = x\} \sim \text{Bin}(N - K, (x/T)^L).$$

D. Proof of COMP maximum achievable rate

We can now prove Theorem 2 using the above results.

Proof of Theorem 2: We start with the achievability part, for which we set $\nu = \ln 2$. We consider the regime $T = \gamma_{\text{COMP}} K \ln N$, where $\gamma_{\text{COMP}} = (1 + \epsilon)/(\ln 2)^2$. As

mentioned in (7), COMP succeeds if and only if $G = 0$. Using Lemma 5, we know that

$$\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = x) = \left(1 - \left(\frac{x}{T}\right)^L\right)^{N-K}, \qquad (24)$$

which is a decreasing function in x. Hence, given δ, for all $x \leq (1/2 + \delta)T$, we have

$$\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = x) \geq \mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = (1/2 + \delta)T)$$
$$= \left(1 - (1/2 + \delta)^L\right)^{N-K}.$$

Next, using the fact that

$$L = \frac{T \ln 2}{K} = \gamma_{\text{COMP}} \ln 2 \ln N = (1 + \epsilon)\frac{1}{\ln 2} \ln N,$$

we find that for any ϵ, we can choose δ sufficiently small that $(\frac{1}{2} + \delta)^L \leq N^{-(1+\epsilon/2)}$, and hence

$$\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = x) \geq \left(1 - N^{-(1+\epsilon/2)}\right)^{N-K}.$$

We deduce that the success probability $\mathbb{P}^{\text{COMP}}(\text{suc})$ is lower bounded as follows:

$$\mathbb{P}^{\text{COMP}}(\text{suc}) = \sum_x \mathbb{P}(W^{(\mathcal{K})} = x)\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = x)$$
$$\geq \sum_{x \leq (1/2+\delta)T} \mathbb{P}(W^{(\mathcal{K})} = x)\left(1 - N^{-(1+\epsilon/2)}\right)^{N-K}$$
$$= \left(1 - N^{-(1+\epsilon/2)}\right)^{N-K} \left(1 - \mathbb{P}(W^{(\mathcal{K})} \geq (1/2 + \delta)T)\right),$$

which is seen to tend to 1 by taking $\alpha = \ln 2$ in Lemma 1 (since we collect a total of $KL = T \ln 2$ coupons).

The converse proceeds similarly, except that we need to consider a general choice of the parameter ν. By a similar argument to the one above, we deduce that the success probability $\mathbb{P}^{\text{COMP}}(\text{suc})$ is given by

$$\sum_x \mathbb{P}(W^{(\mathcal{K})} = x)\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = x)$$
$$\leq \mathbb{P}(W^{(\mathcal{K})} \leq (1 - e^{-\nu} - \delta)T)$$
$$+ \sum_{x \geq (1-e^{-\nu}-\delta)T} \mathbb{P}(W^{(\mathcal{K})} = x)\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = x)$$
$$\leq \delta + \mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = (1 - e^{-\nu} - \delta)T)$$

for N sufficiently large, where we have used Lemma 1.

Using (24) again, but with $L = \nu T/K$, we have

$$\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = (1 - e^{-\nu} - \delta)T) = \left(1 - (1 - e^{-\nu} - \delta)^{\nu T/K}\right)^{N-K} \qquad (25)$$

Since $(1 - e^{-\nu})^\nu$ is minimised at $\nu = \ln 2$, the same is true of the right-hand side when $\delta = 0$. More generally, we can choose some δ' (as a function of δ) such that $\delta' \to 0$ as $\delta \to 0$, and continue as follows:

$$\mathbb{P}(G = 0 \mid W^{(\mathcal{K})} = (1 - e^{-\nu} - \delta)T)$$
$$\leq \left(1 - \left(\frac{1}{2} - \delta'\right)^{(\ln 2 + \delta')T/K}\right)^{N-K}.$$

Since

$$\lim_{N \to \infty} \left(1 - \frac{c}{N - K}\right)^{N-K} = e^{-c}$$

for any $c > 0$, we find that the error probability is upper bounded by $e^{-c}(1 + o(1))$ when $(N - K)(\frac{1}{2} - \delta')^{(\ln 2 + \delta')T/K} \geq c$. Taking the logarithm and noting that c and δ' can be arbitrarily small, we find that the success probability vanishes when

$$\ln 2 \frac{(\ln 2)T}{K \ln N} \leq 1 - \epsilon,$$

which is precisely when $T \leq (1 - \epsilon)T^{\text{COMP}}$. ∎

E. Conditional distribution of L_i

Recalling that L_i denotes the number of tests containing defective item i and no other "possible defective" (item from $\mathcal{PD}$), we have the following.

Lemma 6: For any g, w, j, we have

$$\mathbb{P}\big(L_i = 0 \mid G = g, W^{(\mathcal{K}\backslash i)} = w, M_i = j\big) = \phi_j\left(\frac{1}{w + j}, gL\right),$$

where

$$\phi_j(s, V) = \sum_{\ell=0}^{j} (-1)^\ell \binom{j}{\ell}(1 - \ell s)^V. \qquad (26)$$

Proof: See Appendix A. ∎

Note that the function $\phi_j(s, V)$ also appeared in [21]; however, our analysis here requires using it very differently. We make use of the following properties, the proofs of which are deferred to Appendix B.

Lemma 7: For all values of j, s and V, the function $\phi_j(s, V)$ introduced in (26) has the properties that:

- $\phi_j(s, V)$ is increasing in s,
- $\phi_j(s, V)$ is increasing in V,
- $\phi_j(s, V)$ is decreasing in j.

Lemma 8: If $sVj \leq 2$, then

$$\phi_j(s, V) \leq \frac{V! s^j}{(V - j)!} \leq \exp\big(j \ln(Vs)\big).$$

F. Proof of the DD achievable rate

We put the above results together to prove Theorem 3, giving a lower bound on the achievable rate of the DD algorithm. The key is to express the success probability $\mathbb{P}^{\text{DD}}(\text{suc})$ in terms of an expectation of the function ϕ, and to show that this expectation is concentrated in a regime where ϕ takes favourable values.

Proof of Theorem 3: We consider the regime where $T = \gamma_{\text{DD}} mK \ln N$, with $\gamma_{\text{DD}} = (1 + \epsilon)/(\ln 2)^2$ and $m = \max\{\theta, 1 - \theta\}$. In addition, we choose the parameter $v = \ln 2$. As a result, $L = vT/K$ satisfies the following:

$$L \ln 2 = \frac{T(\ln 2)^2}{K} = m(1 + \epsilon) \ln N.$$

As in [21], writing $\mathbb{P}^{\text{DD}}(\text{suc})$ for the success probability of DD and applying the union bound to (7) we know that

$$\mathbb{P}^{\text{DD}}(\text{suc}) = 1 - \mathbb{P}\left(\bigcup_{i \in \mathcal{K}} \{L_i = 0\}\right) \geq 1 - \sum_{i \in \mathcal{K}} \mathbb{P}(L_i = 0), \quad (27)$$

so that $\mathbb{P}^{\text{DD}}(\text{suc})$ will tend to 1 (as required) if, for a particular defective item $i \in \mathcal{K}$,

$$K\mathbb{P}(L_i = 0) \to 0, \qquad (28)$$

since symmetry means that $\mathbb{P}(L_i = 0)$ is equal for each $i \in \mathcal{K}$. The stated value for the rate then follows upon substituting the choice $T = \gamma_{\text{DD}} mK \ln N$ and (2) into (3).

In order to characterise $\mathbb{P}(L_i = 0)$, we define $A = \{w_- \leq W^{(\mathcal{K}\backslash i)} \leq w_+\}$ and $B = \{G \leq g^*\}$, for some w_-, w_+ and g^* to be chosen shortly. Using Lemma 6, we have the terms at in the large displayed equations at the top of the following page, where:

- (29) follows because, by Lemma 7, on the event $\{A \cap B\}$ the bound $\phi_j(1/w, gL) \leq \phi_j(1/w_-, g^*L)$ holds, and everywhere else $\phi \leq 1$ (since ϕ represents a probability);
- (30) follows since $\mathbb{P}(A \cap B^c) = \mathbb{P}(B^c \mid A)\mathbb{P}(A) \leq \mathbb{P}(B^c \mid A)$.

We consider the terms of (30) separately, taking $w_- = T(1 - \delta)/2$, $w_+ = T(1 + \delta)/2$, and $g^* = N(1/2 + \delta)^L$, where $\delta = (\epsilon \ln 2)/4(1 + \epsilon)$.

The first term of (30) can be bounded as follows. Combining $L = (\ln 2)T/K$ and $w_- = T(1 - \delta)/2$ gives $L/w_- = 2 \ln 2/(K(1 - \delta))$, and recalling that $g^* = N(1/2 + \delta)^L$ and $m = \max(\theta, 1 - \theta)$, it follows that

$$\beta := \ln\left(\frac{g^* L}{w_-}\right)$$
$$= (1 - \theta) \ln N + L \ln(1/2 + \delta) + \ln\left(\frac{2 \ln 2}{1 - \delta}\right)$$
$$\leq m\left(1 + (1 + \epsilon)\left(-1 + \frac{2\delta}{\ln 2}\right)\right) \ln N + \ln\left(\frac{2 \ln 2}{1 - \delta}\right)$$
$$\leq m\left(-\frac{\epsilon}{2}\right) \ln N + \ln\left(\frac{2 \ln 2}{1 - \delta}\right), \qquad (31)$$

where the second line follows by combining $L = m(1 + \epsilon) \ln N/\ln 2$ and $\ln(1/2 + \delta) \leq -\ln 2 + 2\delta$, and the third line follows since $1 + (1 + \epsilon)(-1 + 2\delta/\ln 2) \leq -\epsilon/2$ under the above choice $\delta = (\epsilon \ln 2)/4(1 + \epsilon)$. We claim that (31) implies $jg^* L/w_- \leq 2$ for all $j \leq L$. Indeed, we have $T = \Theta(K \log N)$ and $L = \Theta(T/K)$, so that $L = \Theta(\log N)$, whereas (31) implies that $g^* L/w_-$ decays to zero strictly faster than $1/\log N$. This implies that

$$\phi_j(1/w_-, g^* L) \leq \exp\left(j \ln\left(\frac{g^* L}{w_-}\right)\right) = e^{j\beta}, \qquad (32)$$

since the conditions of Lemma 8 are satisfied under these arguments. Writing $\phi(j) = \phi_j(1/w_-, g^* L)$ (which is decreasing in j by the third part f Lemma 7), we can bound K times the inner sum of (30) as follows:

$$K \sum_{j=0}^{L} \mathbb{P}\big(M_k = j \mid W^{(\mathcal{K}\backslash i)} = w\big) \phi(j)$$
$$\leq KC(L, w) \sum_{j=0}^{L} \mathbb{P}\big(\text{Bin}(L, 1 - w/T) = j\big) \phi(j) \qquad (33)$$
$$= KC(L, w) \sum_{j=0}^{L} \mathbb{P}\big(\text{Bin}(L, 1 - w_+/T) = j\big)$$
$$\times \left(\frac{\mathbb{P}(\text{Bin}(L, 1 - w/T) = j)}{\mathbb{P}(\text{Bin}(L, 1 - w_+/T) = j)}\right) \phi(j) \qquad (34)$$
$$\leq KC(L, w) \sum_{j=0}^{L} \mathbb{P}\big(\text{Bin}(L, 1 - w_+/T) = j\big) \phi(j) \qquad (35)$$

$$\mathbb{P}(L_i = 0) = \sum_{w,j,g} \mathbb{P}\big(W^{(\mathcal{K}\setminus i)} = w, M_i = j, G = g\big) \times \big(\mathbb{I}[A \cap B] + \mathbb{I}\big[(A \cap B)^{\mathsf{c}}\big]\big)\phi_j\left(\frac{1}{w+j}, gL\right)$$

$$\leq \sum_{w \in [w_-, w_+]} \mathbb{P}\big(W^{(\mathcal{K}\setminus i)} = w\big) \sum_{j=0}^{L} \mathbb{P}\big(M_i = j \mid W^{(\mathcal{K}\setminus i)} = w\big)\phi_j(1/w_-, g^*L) + \mathbb{P}\big((A \cap B)^{\mathsf{c}}\big) \tag{29}$$

$$= \sum_{w \in [w_-, w_+]} \mathbb{P}\big(W^{(\mathcal{K}\setminus i)} = w\big) \sum_{j=0}^{L} \mathbb{P}\big(M_i = j \mid W^{(\mathcal{K}\setminus i)} = w\big)\phi_j(1/w_-, g^*L) + \mathbb{P}(A^{\mathsf{c}}) + \mathbb{P}(A \cap B^{\mathsf{c}})$$

$$\leq \sum_{w \in [w_-, w_+]} \mathbb{P}\big(W^{(\mathcal{K}\setminus i)} = w\big) \sum_{j=0}^{L} \mathbb{P}\big(M_i = j \mid W^{(\mathcal{K}\setminus i)} = w\big)\phi_j(1/w_-, g^*L) \tag{30}$$
$$+ \mathbb{P}\big(W^{(\mathcal{K}\setminus i)} \notin [w_-, w_+]\big) + \mathbb{P}\big(G > g^* \mid W^{(\mathcal{K}\setminus i)} \in [w_-, w_+]\big),$$

$$\leq KC(L, w_-)\left(\frac{w_+}{T} + \frac{T - w_+}{T}\,\mathrm{e}^{\beta}\right)^{L} \tag{36}$$

Here:

- (33) follows from the second part of Lemma 4.
- We deduce (35) using the following argument: The bracketed term in (34) is easily verified to be increasing in j by substituting the Binomial mass function and noting $1 - w/T \geq 1 - w_+/T$, and we already know from Lemma 7 that $\phi(j)$ is decreasing. Hence, (34) is the expectation of the product of an increasing and decreasing function, and so by 'Chebyshev's other inequality' [47, eq. (1.7)], it is bounded above by the product of the expectations of those functions.[2]
- (36) follows by upper bounding $C(L, w) \leq C(L, w_-)$ and $\phi(j) \leq \mathrm{e}^{j\beta}$ from (32), and then evaluating the sum exactly.

We can simplify (36) using the following:

$$KC(L, w_-)\left(\frac{w_+}{T} + \frac{T - w_+}{T}\,\mathrm{e}^{\beta}\right)^{L}$$
$$= \frac{KC(L, w_-)}{2^L}\big(1 + \delta + \mathrm{e}^{\beta}(1 - \delta)\big)^{L} \tag{37}$$
$$\leq C(L, w_-) \cdot c\exp(-m\epsilon \ln N)\exp\big(L(\delta + \mathrm{e}^{\beta}(1 - \delta))\big) \tag{38}$$
$$\leq C(L, w_-) \cdot c\exp\left(\left(-m\epsilon + \frac{m(1+\epsilon)}{\ln 2}(\delta + \mathrm{e}^{\beta}(1 - \delta))\right)\ln N\right) \tag{39}$$

Here:

- (37) follows by substituting $w_+ = T(1 + \delta)/2$.
- (38) follows from $1 + \zeta \leq \mathrm{e}^{\zeta}$, along with the fact that

$$\frac{K}{2^L} \leq c\exp\big((m - m(1 + \epsilon))\ln N\big) \leq c\exp(-m\epsilon \ln N)$$

by $L = m(1 + \epsilon)\ln N/\ln 2$ and by $K = \Theta(N^{\theta})$ giving $K \leq cN^{\theta} \leq cN^{m}$ for some $c = \Theta(1)$.
- (39) follows by again using $L = m(1 + \epsilon)\ln N/\ln 2$.

We conclude that (39) acts as an upper bound on K times the first term of (30). Overall (39) tends to zero for δ sufficiently small, since $C(L, w_-) = \exp(L^2/4w_-)$ tends to 1 in this regime.

[2]In fact, [47, eq. (1.7)] concerns $\mathbb{E}[f(X)g(X)]$ for two increasing functions, but we can transform this to $\mathbb{E}[f(X)h(X)]$ for decreasing h by simply defining $h(x) = L - g(x)$.

The second term of (30) decays to zero exponentially fast in T by Lemma 1. More precisely, we make $(K - 1)L$ draws with replacement, so that $\alpha = (K - 1)\ln 2/K \to \ln 2$, meaning that we can take $\epsilon = \delta/3$ in Lemma 1 to obtain

$$\limsup_{N \to \infty} K\,\mathbb{P}\big(W^{(\mathcal{K}\setminus i)} \notin (w_-, w_+)\big)$$
$$\leq \limsup_{N \to \infty} K\,\mathbb{P}\left(\big|W^{(\mathcal{K}\setminus i)} - (1 - \mathrm{e}^{-\alpha})T\big| \geq \epsilon T\right)$$
$$\leq 2\limsup_{N \to \infty} K\exp\left(-\frac{\epsilon^2 T}{\alpha}\right)$$
$$= 2c\limsup_{N \to \infty} \exp\left(\ln N\left(\theta - \frac{\epsilon^2 \gamma_{\mathrm{DD}} mK}{\alpha}\right)\right),$$

since $T = \gamma_{\mathrm{DD}} mK \ln N$ and $K = \Theta(N^{\theta})$ (and hence $K \leq cN^{\theta}$ for some $c = \Theta(1)$). We conclude that this term tends to zero, since the exponent behaves as $-K \ln N$.

To control the third term in (30), observe that if $W^{(\mathcal{K}\setminus i)} \leq w_+$, then

$$\frac{W^{(\mathcal{K})}}{T} \leq \frac{1 + \delta}{2} + \frac{L}{T} = \frac{1 + \delta}{2} + \frac{\ln 2}{K} \leq \frac{1}{2} + \frac{3\delta}{4},$$

where the first inequality holds since $W^{(\mathcal{K})} \leq W^{(\mathcal{K}\setminus i)} + L$ and $w_+ = T(1 + \delta)/2$, the equality holds since $L = \ln 2T/K$, and the final inequality holds for K sufficiently large. Hence, and defining $p = (1/2 + 3\delta/4)^L$, Lemma 5 gives

$$\mathbb{P}\big(G > g^* \mid W^{(\mathcal{K}\setminus i)} \in (w_-, w_+)\big)$$
$$\leq \mathbb{P}\big(\mathrm{Bin}(N, p) > g^*\big)$$
$$\leq \exp\left(-\frac{(g^*)^2}{2(Np + g^*/3)}\right) \tag{40}$$
$$= \exp\left(-N\frac{(1/2 + \delta)^{2L}}{2((1/2 + 3\delta/4)^L + (1/2 + \delta)^L/3)}\right) \tag{41}$$
$$= \exp\left(-N\frac{(1/2 + \delta)^L}{2(1/3 + o(1))}\right), \tag{42}$$

where (40) follows from Bernstein's inequality [48, eq. (2.10)], (41) follows from $p = (1/2 + 3\delta/4)^L$ and $g^* = N(1/2 + \delta)^L$, and (42) follows since the ratio of $(1/2 + 3\delta/4)^L$ to $(1/2 + \delta)^L$ tends to zero as $N \to \infty$ (and hence $L \to \infty$, since $L = \Theta(\log N)$).

Finally, since $L = m(1 + \epsilon)\log N/\log 2$, we find that $(1/2 + \delta)^L$ behaves as N^{-c} for some c that can be made

arbitrarily close to $m = \max\{\theta, 1 - \theta\}$ by choosing δ and ϵ sufficiently small. By definition, $m < 1$, and the bound in (42) is exponential in N^{1-c}, meaning that it vanishes even when multiplied by K. ∎

V. CONCLUSIONS AND OPEN QUESTIONS

We have studied nonadaptive group testing with near-constant column weight designs. We have seen that:

- The very simple COMP algorithms requires 23.4% fewer tests with a near-constant column weight design than with a Bernoulli design, performing even better than optimal algorithms with Bernoulli designs for $\theta > 0.766$.
- Using a near-constant column weight design, the practical DD algorithm again uses 23.4% fewer tests than with Bernoulli designs, outperforms any possible algorithm for Bernoulli designs for $\theta > 0.434$, and beats the best-known theoretical guarantees of existing practical algorithms for all $\theta \in (0, 1)$.
- An upper bound on performance of near-constant column weight designs shows that DD is optimal for this design when $\theta \geq 1/2$.
- Numerical experiments demonstrated a notable improvement over Bernoulli designs in both sparse and dense regimes.

We briefly mention some interesting open problems connected with this paper, which we hope to address in future work:

1) It remains open to determine the maximum achievable rate of constant or near-constant column weight designs for $\theta \leq 1/2$, in the spirit of Theorem 1. We conjecture that the value is (5) (i.e., Theorem 4 is sharp), and is achievable by the maximum likelihood algorithm (as well as the SSS algorithm of [21]). This is the value suggested by a non-rigorous result of Mézard, Tarzia and Toninelli [36].
2) It is an important open problem to determine whether 'practical' algorithms can improve on the performance of DD. For example, the SCOMP algorithm of [21] and approaches based on linear programming both have a rate at least as large as DD [44]. However, we do not know the *best possible* rate of DD for $\theta < 1/2$, nor how to determine whether these algorithms or others can have a higher rate than DD.
3) It remains of great interest to determine whether a rate of 1 can be achieved for values of θ beyond $1/3$ using constant or near-constant column weights or some other design. The conjecture above would imply that near-constant column-weight designs achieve rate 1 for $\theta < 0.409$. More generally, it is an open problem as to whether there exists an 'adaptivity gap', i.e., a choice of $\theta < 1$ such that any nonadaptive design must have rate less than 1, despite the rate of 1 being achievable with adaptive testing.

APPENDIX A
PROPERTIES OF THE DISTRIBUTION OF M_i

A. Proof of Lemma 4

Proof of Lemma 4: We prove the first part of the lemma directly. Suppose that we pick L coupons from a population of T coupons, w of which were previously chosen. Clearly, the probability of the event that exactly s of the coupons picked were previously chosen is $\mathbb{P}(\text{Bin}(L, w/T) = s)$.

Conditioning on this event, we calculate the probability that we pick $L - s$ coupons out of a population of $T - w$ coupons and obtain exactly j distinct new coupons. Clearly we require $L - s \geq j$, or $s \leq L - j$. By a standard counting argument, we can choose these j coupons $\binom{T-w}{j}$ ways, then $\left\{{L-s \atop j}\right\}$ ways of placing the $L-s$ coupons into j unlabelled bins such that none of them are empty (see [46, p. 204]), and finally $j!$ different labellings of the bins. Moreover, overall there are $(T - w)^{L-s}$ assignments of the coupons.

Putting this all together and recalling the definition

$$(T - w)_{(j)} = \frac{(T - w)!}{(T - w - j)!} = \binom{T - w}{j} j!,$$

we have

$$\mathbb{P}\big(M_i = j \mid W^{(\mathcal{K}\backslash i)} = w\big)$$
$$= \sum_{s=0}^{L-j} \binom{L}{s} \left(\frac{w}{T}\right)^s \left(1 - \frac{w}{T}\right)^{L-s} \binom{T - w}{j} \left\{{L - s \atop j}\right\} j! \frac{1}{(T - w)^{L-s}}$$
$$= \sum_{s=0}^{L-j} \binom{L}{s} \frac{w^s}{T^L} (T - w)_{(j)} \left\{{L - s \atop j}\right\},$$

as required.

We now prove the second part of the lemma. Relabelling $t = L - j - s \geq 0$ and using the fact that $\left\{{t+j \atop j}\right\} \leq \binom{t+j}{j} j^t$ (see [49]), we obtain that the inner sum of (23) is:

$$\sum_{t=0}^{L-j} \binom{L}{L - j - t} \left\{{t + j \atop j}\right\} w^{L-j-t}$$
$$\leq w^{L-j} \sum_{t=0}^{L-j} \binom{L}{L - j - t} \binom{t + j}{j} \left(\frac{j}{w}\right)^t$$
$$= w^{L-j} \binom{L}{j} \sum_{t=0}^{L-j} \binom{L - j}{t} \left(\frac{j}{w}\right)^t$$
$$= w^{L-j} \binom{L}{j} \left(1 + \frac{j}{w}\right)^{L-j}$$
$$\leq w^{L-j} \binom{L}{j} C,$$

where the third line follows by explicitly evaluating the summation, and the final line holds with $C = \exp(L^2/4w)$ since

$$\left(1 + \frac{j}{w}\right)^{L-j} \leq \exp\left(\frac{j(L - j)}{w}\right) \leq \exp\left(\frac{L^2}{4w}\right).$$

This allows us to deduce that the whole of (23) satisfies

$$\mathbb{P}\big(M_i = j \mid W^{(\mathcal{K}\setminus i)} = w\big) \le C \frac{(T-w)_{(j)}}{T^L} w^{L-j} \binom{L}{j}$$

$$\le C \frac{(T-w)^j}{T^L} w^{L-j} \binom{L}{j},$$

as required. ∎

B. Proof of Lemma 6

Proof of Lemma 6: The case $j = 0$ is trivial, so we assume here that $j \ge 1$.

We have conditioned on three events: on $W^{(\mathcal{K}\setminus i)} = w$, meaning there are w tests containing one or more item from $\mathcal{K} \setminus i$; on event $M_i = j$, meaning there are j tests that contain item i and no member of $\mathcal{K} \setminus i$; and on $G = g$, meaning there are g items labelled as possibly defective but not in $\mathcal{K}$.

By relabelling, without loss of generality, we can assume that tests $1, \ldots, j$ are the ones that contain defective item i and no other defective item. We write A_s for the event that test s does not have any element of $\mathcal{PD} \setminus \mathcal{K}$ in it.

If an item is in $\mathcal{PD} \setminus \mathcal{K}$, then the tests that it appears in are chosen uniformly among those which already contain a defective. Hence, for any set $S \subseteq \{1, \ldots, j\}$ of size ℓ, we have $\mathbb{P}(\bigcap_{r \in S} A_r) = (1 - \ell/(w+j))^{gL}$. This is because we require that the L coupons of each of g items in $\mathcal{PD} \setminus \mathcal{K}$ take values in the set of positive tests ($W^{(\mathcal{K}\setminus i)} + M_i = w + j$ in total), but avoid the specified ℓ tests. Thus,

$$\mathbb{P}\big(L_i = 0 \mid G = g, W^{(\mathcal{K}\setminus i)} = w, M_i = j\big)$$

$$= \mathbb{P}\left(\bigcap_{s=1}^{j} A_s^{\mathsf{c}}\right)$$

$$= 1 - \mathbb{P}\left(\bigcup_{s=1}^{j} A_s\right)$$

$$= \sum_{\ell=0}^{j} (-1)^{\ell} \sum_{\substack{S \subseteq \{1,\ldots,j\} \\ |S|=\ell}} \mathbb{P}\left(\bigcap_{j \in S} A_j\right)$$

$$= \sum_{\ell=0}^{j} (-1)^{\ell} \binom{j}{\ell} \left(1 - \frac{\ell}{w+j}\right)^{gL}, \quad (43)$$

and the result follows. ∎

APPENDIX B
PROPERTIES OF THE ϕ FUNCTION

A. Proof of Lemma 7

Proof of Lemma 7: We deduce the results using the expression

$$\phi_j(s, V) = \sum_{\ell=0}^{j} (-1)^{\ell} \binom{j}{\ell} (1 - \ell s)^V$$

from (26).

First we show that $\phi_j(s, V)$ is increasing in s. As in [21, Lemma 32], a direct calculation using the fact that $\ell\binom{j}{\ell} = j\binom{j-1}{\ell-1}$ gives

$$\frac{\partial}{\partial s} \phi_j(s, V)$$

$$= Vj \sum_{\ell=1}^{j} (-1)^{\ell-1} \binom{j-1}{\ell-1} (1 - \ell s)^{V-1} \quad (44)$$

$$= (1-s)^{V-1} Vj \sum_{\ell=1}^{j} (-1)^{\ell-1} \binom{j-1}{\ell-1} \left(1 - \frac{(\ell-1)s}{1-s}\right)^{V-1}$$

$$= (1-s)^{V-1} Vj\, \phi_{j-1}\left(\frac{s}{1-s}, V-1\right) \quad (45)$$

$$\ge 0,$$

where the second line uses the fact that

$$(1 - \ell s) = (1-s)\left(1 - \frac{(\ell-1)s}{1-s}\right),$$

and the third line above follows by relabelling $\ell' = \ell - 1$.

Second, we show that $\phi_j(s, V)$ is increasing in V. Again using $\ell\binom{j}{\ell} = j\binom{j-1}{\ell-1}$, we can write

$$\phi_j(s, V) - \phi_j(s, V-1)$$

$$= \sum_{\ell=0}^{j} (-1)^{\ell} \binom{j}{l} (1 - \ell s)^{V-1}\big((1 - \ell s) - 1\big)$$

$$= sj \sum_{\ell=1}^{j} (-1)^{\ell-1} \binom{j-1}{\ell-1} (1 - \ell s)^{V-1}$$

$$= \frac{s}{V} \frac{\partial}{\partial s} \phi_j(s, V)$$

$$\ge 0,$$

where the third line follows from (44).

Third, we show that $\phi_j(s, V)$ is decreasing in j. By expanding $\binom{j}{\ell} = \binom{j-1}{\ell} + \binom{j-1}{\ell-1}$, we can write

$$\phi_j(s, V) = \sum_{\ell=0}^{j} (-1)^{\ell} \left(\binom{j-1}{\ell} + \binom{j-1}{\ell-1}\right) (1 - \ell s)^V$$

$$= \phi_{j-1}(s, V) - \frac{1}{(V+1)j} \frac{\partial}{\partial s} \phi_j(s, V+1)$$

$$\le \phi_{j-1}(s, V),$$

again using (44). ∎

B. Proof of Lemma 8

We now prove Lemma 8, first giving two preliminary lemmas.

Lemma 9: We can expand $\phi_j(s, V)$ (as defined in (26)) as a polynomial in s of degree V as follows:

$$\phi_j(s, V) = \frac{V! s^j}{(V-j)!} \sum_{u=0}^{V-j} (-1)^u s^u \frac{j!(V-j)!}{(u+j)!(V-u-j)!} \begin{Bmatrix} j+u \\ j \end{Bmatrix},$$
$$(46)$$

where we again write $\begin{Bmatrix} j+u \\ j \end{Bmatrix}$ for the Stirling number of the second kind.

Proof: We can expand

$$\phi_j(s,V) = \sum_{\ell=0}^{j}(-1)^\ell \binom{j}{\ell}(1-\ell s)^V$$

$$= \sum_{\ell=0}^{j}(-1)^\ell \binom{j}{\ell}\sum_{t=0}^{V}\binom{V}{t}(-s)^t \ell^t$$

$$= \sum_{t=0}^{V}\binom{V}{t}(-s)^t \sum_{\ell=0}^{j}(-1)^\ell \binom{j}{\ell}\ell^t \qquad (47)$$

$$= \sum_{t=0}^{V}\binom{V}{t}(-s)^t \begin{Bmatrix} t \\ j \end{Bmatrix} j!\,(-1)^j$$

$$= \sum_{u=0}^{V-j}\binom{V}{j+u}(-s)^{j+u}\begin{Bmatrix} j+u \\ j \end{Bmatrix} j!\,(-1)^j,$$

where the second line can be seen by directly evaluating the summation, the fourth line follows by recognising the inner sum in (47) as a multiple of the Stirling number using (22), and the last line follows since by relabelling $t = j + u$ and noting that $\begin{Bmatrix} t \\ j \end{Bmatrix}$ is non-zero only when $t \geq j$. The result now follows by writing

$$\binom{V}{j+u} = \frac{V!}{(V-u-j)!\,(u+j)!}\frac{(V-j)!}{(V-j)!}$$

and $(-s)^{j+u}(-1)^j = (-1)^u s^j s^u$. ∎

We also use the following result from [50, Theorem 4.4].

Lemma 10: The Stirling numbers of the second kind are log-concave in their first argument, that is for any $j, u \in \mathbb{Z}_+$:

$$\begin{Bmatrix} j+u+1 \\ j \end{Bmatrix}^2 \geq \begin{Bmatrix} j+u \\ j \end{Bmatrix}\begin{Bmatrix} j+u+2 \\ j \end{Bmatrix}.$$

We are now in a position to prove Lemma 8.

Proof of Lemma 8: Using Lemma 9, we consider $\phi_j(s, V)$ as a sum of the form

$$\phi_j(s,V) = \frac{s^j V!}{(V-j)!}\sum_{u=0}^{V-j}(-1)^u a_u,$$

where

$$a_u = s^u \frac{j!\,(V-j)!}{(u+j)!\,(V-u-j)!}\begin{Bmatrix} j+u \\ j \end{Bmatrix}.$$

By the alternating series test, if a_u is a monotonically decreasing sequence, we can bound $\sum_{u=0}^{V-j}(-1)^u a_u \leq a_0 = 1$, and the result follows. We can verify that a_u is indeed monotonically decreasing by considering the ratio

$$\frac{a_{u+1}}{a_u} = s\frac{V-j-u}{j+u+1}\frac{\begin{Bmatrix} j+u+1 \\ j \end{Bmatrix}}{\begin{Bmatrix} j+u \\ j \end{Bmatrix}}. \qquad (48)$$

The first fraction in (48) is trivially decreasing in u, and the second fraction in (48) is decreasing in u by Lemma 10. Hence, since the ratio (48) is decreasing in u, it is sufficient to verify that $a_1/a_0 \leq 1$. Since $\begin{Bmatrix} j \\ j \end{Bmatrix} = 1$ and $\begin{Bmatrix} j+1 \\ j \end{Bmatrix} = j(j+1)/2$, direct substitution in (48) gives that $a_1/a_0 = s(V-j)j/2$, so it is sufficient to assume that $s(V-j)j/2 \leq 1$. ∎

APPENDIX C
AUXILIARY RESULTS FOR THE ALGORITHM-INDEPENDENT CONVERSE

A. Proof of Lemma 2

Fixing the index i of some defective item, we note that conditioned on $W^{(\mathcal{K}\backslash i)} = w$, the event A_i occurs if each test that item i occurs in is contained in the w 'already hit' tests. Hence, for any $c_1 > 0$, we can write

$$\mathbb{P}(A_i) = \sum_w \mathbb{P}\big(A_i \mid W^{(\mathcal{K}\backslash i)} = w\big)\mathbb{P}\big(W^{(\mathcal{K}\backslash i)} = w\big)$$

$$= \sum_w \left(\frac{w}{T}\right)^L \mathbb{P}\big(W^{(\mathcal{K}\backslash i)} = w\big)$$

$$\geq \sum_w \left(\frac{w}{T}\right)^L \mathbb{P}\big(W^{(\mathcal{K}\backslash i)} = w\big)\mathbb{I}[w \geq Tc_1]$$

$$\geq c_1^L\,\mathbb{P}\big(W^{(\mathcal{K}\backslash i)} \geq Tc_1\big),$$

which proves (13).

The analysis of the event $A_i \cap A_j$ for $i \neq j$ is more challenging; we show in Section C-B below that if $w \geq Tc_1$ with the same $c_1 > 0$ as above, then for arbitrarily small $\epsilon_1 > 0$ it holds for sufficiently large N that

$$\mathbb{P}\big(A_i \cap A_j \mid W^{(\mathcal{K}\backslash i,j)} = w\big) \leq \left(\frac{w+L}{T}\right)^{2L}\frac{1}{1-\epsilon_1}. \qquad (49)$$

Hence, for any $c_1, c_2 > 0$, we have

$$\binom{K}{2}\mathbb{P}(A_i \cap A_j)$$

$$= \binom{K}{2}\sum_w \mathbb{P}\big(A_i \cap A_j \mid W^{(\mathcal{K}\backslash i,j)} = w\big)\mathbb{P}\big(W^{(\mathcal{K}\backslash i,j)} = w\big)$$

$$\leq \frac{1}{1-\epsilon_1}\binom{K}{2}\sum_{Tc_1 \leq w \leq Tc_2}\left(\frac{w+L}{T}\right)^{2L}\mathbb{P}\big(W^{(\mathcal{K}\backslash i,j)} = w\big)$$

$$\quad + \binom{K}{2}\mathbb{P}\big(W^{(\mathcal{K}\backslash i,j)} \notin [Tc_1, Tc_2]\big)$$

$$\leq \frac{K^2}{2(1-\epsilon_1)}\left(c_2 + \frac{\nu}{K}\right)^{2L}\mathbb{P}\big(Tc_1 \leq W^{(\mathcal{K}\backslash i,j)} \leq Tc_2\big)$$

$$\quad + \binom{K}{2}\mathbb{P}\big(W^{(\mathcal{K}\backslash i,j)} < Tc_1\big) + \binom{K}{2}\mathbb{P}\big(W^{(\mathcal{K}\backslash i,j)} > Tc_2\big),$$

where we have used the fact that $L/T = \nu/K$. This proves (14).

B. Proof of (49)

Recall that we condition on $W^{(\mathcal{K}\backslash i,j)} = w$, and seek to bound the probability of $A_i \cap A_j$ for two defective items i, j. Here A_i is the event that item i is masked by the remaining defective items $\mathcal{K}\backslash\{i\}$ (one of which is j), and analogously for A_j.

In contrast to the rest of the paper, in this section, we represent the columns of the test matrix X corresponding to items i and j by lists $\mathcal{T}_i = (t_{i,1}, \ldots t_{i,L})$ and $\mathcal{T}_j = (t_{j,1}, \ldots t_{j,L})$. Each list entry is obtained by choosing $t \in \{1, \ldots, T\}$ uniformly at random with replacement, so duplicates may occur. Any given list occurs with probability $1/T^L$.

Without loss of generality, we assume that the w tests containing items from $\mathcal{K}\backslash\{i,j\}$ are those indexed by $1, \ldots, w$.

Letting $\mathcal{A}_i$ be the set of list pairs $(\mathcal{T}_i, \mathcal{T}_j)$ under which the event A_i occurs, and similarly for $\mathcal{A}_j$, we have

$$\Pr(A_i \cap A_j \mid W^{(\mathcal{K}\backslash i,j)} = w) = \frac{N_{ij}}{T^{2L}}, \tag{50}$$

where

$$N_{ij} = \sum_{\mathcal{T}_i} \sum_{\mathcal{T}_j} \mathbb{I}\{(\mathcal{T}_i, \mathcal{T}_j) \in \mathcal{A}_i \cap \mathcal{A}_j\} \tag{51}$$

is the number of pairs of lists in $\mathcal{A}_i \cap \mathcal{A}_j$. Here the sets $\mathcal{A}_i$ and $\mathcal{A}_j$ implicitly depend on w.

To bound N_{ij}, we separately consider the number of 'new positive tests' caused by items i and j; that is, those not among the first w. Specifically, letting $N_{ij}(\ell)$ be defined as in (51) with the summation limited to the case that there are ℓ such new positive tests, we have

$$N_{ij} = \sum_{\ell=0}^{L} N_{ij}(\ell).$$

The summation only goes up to L due to the fact that any new positive test containing i must also contain j and vice versa, otherwise the masking under consideration would not occur.

To bound $N_{ij}(\ell)$, we consider the following procedure for choosing the lists:

- From $T-w$ tests, choose ℓ of them to be the new defective tests. This is one of $\binom{T-w}{\ell}$ options.
- For both i and j, assign one list index from $\{1, \dots, L\}$ to each of the ℓ new defective tests. This is at most L^{ℓ} options each, for $L^{2\ell}$ in total.
- For both i and j, the remaining $L-\ell$ list entries are chosen arbitrarily from the $w+\ell$ positive tests. This is $(w+\ell)^{L-\ell}$ options each, for $(w+\ell)^{2(L-\ell)}$ in total.

Combining these terms gives

$$\begin{aligned} N_{ij}(\ell) &\le \binom{T-w}{\ell} \cdot L^{2\ell} \cdot (w+\ell)^{2(L-\ell)} \\ &\le (T-w)^{\ell} \cdot L^{2\ell} \cdot (w+L)^{2(L-\ell)} \\ &= (w+L)^{2L} \cdot \left(\frac{L^2(T-w)}{(w+L)^2}\right)^{\ell}. \end{aligned}$$

Now, since $w \ge c_1 T$ by assumption, and recalling that $T = \Theta(K \log N)$, that $K = \Theta(N^{\theta})$ with $\theta \in (0,1)$, and that $L = \nu T/K = \Theta(\log N)$, we find that the bracketed term is less than any fixed $\epsilon_1 > 0$ for sufficiently large N. Hence, summing over ℓ gives

$$\begin{aligned} N_{ij} &\le \sum_{\ell=0}^{L} (w+L)^{2L} \cdot \left(\frac{L^2(T-w)}{(w+L)^2}\right)^{\ell} \\ &\le (w+L)^{2L} \sum_{\ell=0}^{\infty} \epsilon_1^{\ell} \\ &= (w+L)^{2L} \frac{1}{1-\epsilon_1}. \end{aligned}$$

Substituting into (50), we conclude that

$$\Pr(A_i \cap A_j \mid W^{(\mathcal{K}\backslash i,j)} = w) \le \left(\frac{w+L}{T}\right)^{2L} \cdot \frac{1}{1-\epsilon_1},$$

as desired.

C. Proof of Lemma 3

We consider the procedure of selecting T such that the fraction in the bracketed term in (20) equates to $1 - \frac{1}{2(1-\epsilon_1)}$:

$$1 - \frac{K(c_2 + \nu/K)^{2\nu T/K}}{2c_1^{\nu T/K}(1-\epsilon_1)} = 1 - \frac{1}{2(1-\epsilon_1)}, \tag{52}$$

or equivalently

$$K \cdot \left(\frac{(c_2 + \nu/K)^2}{c_1}\right)^{\nu T/K} = 1. \tag{53}$$

Substituting $c_1 = 1 - e^{-\nu} - \sqrt{\frac{c_0 \ln T}{T}}$ and $c_2 = 1 - e^{-\nu} + \sqrt{\frac{c_0 \ln T}{T}}$ and performing some simple rearrangements, we obtain

$$\begin{aligned} \frac{\nu T}{K} &= \frac{\ln K}{\ln \frac{1 - e^{-\nu} - \sqrt{(c_0 \ln T)/T}}{\left(1 - e^{-\nu} + \nu/K + \sqrt{(c_0 \ln T)/T}\right)^2}} \\ &= \frac{\ln K}{\ln\left(1 - e^{-\nu} - \sqrt{\frac{c_0 \ln T}{T}}\right) - 2\ln\left(1 - e^{-\nu} + \frac{\nu}{K} + \sqrt{\frac{c_0 \ln T}{T}}\right)}. \end{aligned}$$

Applying Taylor expansions and noting that the terms $\sqrt{\frac{c_0 \ln T}{T}}$ and ν/K both decay as $O\left(\sqrt{\frac{\ln T}{T}}\right) = O(1/\sqrt{K})$, we obtain

$$\frac{\nu T}{K} = \frac{\ln K}{-\ln(1 - e^{-\nu})}\left(1 + O\left(\frac{1}{\sqrt{K}}\right)\right). \tag{54}$$

This choice is consistent with the assumed condition on T in (10).

Having established (54), we characterise the term $c_1^{\nu T/K}$ appearing in (19), with $c_1 = 1 - e^{-\nu} - \sqrt{\frac{c_0 \ln T}{T}}$ as above:

$$\begin{aligned} \ln(c_1^{\nu T/K}) &= \frac{\ln K \cdot \ln\left(1 - e^{-\nu} - \sqrt{\frac{c_0 \ln T}{T}}\right)}{-\ln(1 - e^{-\nu})}\left(1 + O\left(\frac{1}{\sqrt{K}}\right)\right) \\ &= (-\ln K)\left(1 + O\left(\frac{1}{\sqrt{K}}\right)\right), \end{aligned}$$

where we have again applied standard Taylor expansions. Taking the exponential of both sides gives $c_1^{\nu T/K} = 1/K^{1+O(1/\sqrt{K})}$, which yields $Kc_1^{\nu T/K} = 1 - o(1)$ since $K^{1/K^c} \to 1$ as $K \to \infty$ for any $c > 0$. Substituting into (20) and recalling (52), we obtain the desired bound (21).

REFERENCES

[1] M. Aldridge, O. Johnson, and J. Scarlett, "Improved group testing rates with constant column weight designs," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 1381–1385.

[2] R. Dorfman, "The detection of defective members of large populations," *Ann. Math. Statist.*, vol. 14, no. 4, pp. 436–440, 12 1943.

[3] D. Du and F. Hwang., *Combinatorial Group Testing and Its Applications*, ser. Series on Applied Mathematics. World Scientific, 1993.

[4] C. Aksoylar, G. Atia, and V. Saligrama, "Sparse signal processing with linear and non-linear observations: A unified shannon theoretic approach," in *2013 IEEE Information Theory Workshop (ITW)*, Sept 2013, pp. 1–5.

[5] H.-B. Chen and F. K. Hwang, "A survey on nonadaptive group testing algorithms through the angle of decoding," *Journal of Combinatorial Optimization*, vol. 15, no. 1, pp. 49–59, Jan 2008.

[6] J. L. Gastwirth and P. A. Hammick, "Estimation of the prevalence of a rare disease, preserving the anonymity of the subjects by group testing: application to estimating the prevalence of AIDS antibodies in blood donors," *Journal of Statistical Planning and Inference*, vol. 22, no. 1, pp. 15–27, 1989.

[7] R. Mourad, Z. Dawy, and F. Morcos, "Designing pooling systems for noisy high-throughput protein-protein interaction experiments using Boolean compressed sensing," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 6, pp. 1478–1490, 2013.

[8] K. H. Thompson, "Estimation of the proportion of vectors in a natural population of insects," *Biometrics*, vol. 18, no. 4, pp. 568–578, 1962.

[9] S. D. Walter, S. W. Hildreth, and B. J. Beaty, "Estimation of infection rates in populations of organisms using pools of variable size," *American Journal of Epidemiology*, vol. 112, no. 1, pp. 124–128, 1980.

[10] M. T. Goodrich and D. S. Hirschberg, "Improved adaptive group testing algorithms with applications to multiple access channels and dead sensor diagnosis," *Journal of Combinatorial Optimization*, vol. 15, no. 1, pp. 95–121, 2008.

[11] C. Lo, M. Liu, J. P. Lynch, and A. C. Gilbert, "Efficient sensor fault detection using combinatorial group testing," in *2013 IEEE International Conference on Distributed Computing in Sensor Systems*, May 2013, pp. 199–206.

[12] A. Emad and O. Milenkovic, "Poisson group testing: A probabilistic model for nonadaptive streaming boolean compressed sensing," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3335–3339.

[13] A. C. Gilbert, M. A. Iwen, and M. J. Strauss, "Group testing and sparse signal recovery," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, 2008, pp. 1059–1063.

[14] M. K. Varanasi, "Group detection for synchronous Gaussian code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 1083–1096, 1995.

[15] J. K. Wolf, "Born again group testing: multiaccess communications," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 185–191, 1985.

[16] S. Wu, S. Wei, Y. Wang, R. Vaidyanathan, and J. Yuan, "Achievable partition information rate over noisy multi-access boolean channel," in *2014 IEEE International Symposium on Information Theory*, June 2014, pp. 1206–1210.

[17] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: an information-theoretic framework," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 593–620, Jan 2017.

[18] ——, "Phase transitions in group testing," in *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, 2016, pp. 40–53.

[19] M. Aldridge, "The capacity of Bernoulli nonadaptive group testing," *IEEE Transactions on Information Theory*, vol. 63, no. 11, pp. 7142–7148, Nov 2017.

[20] M. Aldridge, L. Baldassini, and K. Gunderson, "Almost separable matrices," *Journal of Combinatorial Optimization*, vol. 33, no. 1, pp. 215–236, Jan 2017.

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

[22] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1880–1901, March 2012.

[23] C. L. Chan, P. H. Che, S. Jaggi, and V. Saligrama, "Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms," in *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2011, pp. 1832–1839.

[24] M. Malyutov, "Search for sparse active inputs: a review," in *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede*, H. Aydinian, F. Cicalese, and C. Deppe, Eds. Springer, 2013, pp. 609–647.

[25] L. Baldassini, O. Johnson, and M. Aldridge, "The capacity of adaptive group testing," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 2676–2680.

[26] F. K. Hwang, "A method for detecting all defective members in a population by group testing," *Journal of the American Statistical Association*, vol. 67, no. 339, pp. 605–608, 1972.

[27] A. G. D'yachkov and V. V. Rykov, "Bounds on the length of disjunctive codes," *Problemy Peredachi Informatsii*, vol. 18, no. 3, pp. 7–13, 1982, translation: *Problems of Information Transmission*. 18(3): 166–171.

[28] A. G. D'yachkov, V. V. Rykov, and A. M. Rashad, "Superimposed distance codes," *Problems of Control and Information Theory*, vol. 18, no. 4, pp. 237–250, 1989.

[29] M. B. Malyutov, "The separating property of random matrices," *Matematicheskie Zametki*, vol. 23, no. 1, pp. 155–167, 1978, translation: *Mathematical Notes of the Academy of Sciences of the USSR*. 23(1): 84–91.

[30] V. L. Freidlina, "On a design problem for screening experiments," *Teoriya Veroyatnostei i ee Primeneniya*, vol. 20, no. 1, pp. 100–114, 1975, translation: *Theory of Probability & Its Applications*, vol. 20, no. 1, pp. 102–115, 1975.

[31] W. H. Kautz and R. C. Singleton, "Nonrandom binary superimposed codes," *IEEE Transactions on Information Theory*, vol. 10, no. 4, pp. 363–377, 1964.

[32] D. J. Balding, W. J. Bruno, D. Torney, and E. Knill, *A comparative survey of non-adaptive pooling designs*. Springer, 1996, pp. 133–154.

[33] W. Bruno, E. Knill, D. Balding, D. Bruce, N. Doggett, W. Sawhill, R. Stallings, C. Whittaker, and D. Torney, "Efficient pooling designs for library screening," *Genomics*, vol. 26, no. 1, pp. 21 – 30, 1995.

[34] F. K. Hwang, "Random *k*-set pool designs with distinct columns," *Probability in the Engineering and Informational Sciences*, vol. 14, no. 1, p. 4956, 2000.

[35] A. J. Macula, "Probabilistic nonadaptive and two-stage group testing with relatively small pools and DNA library screening," *Journal of Combinatorial Optimization*, vol. 2, no. 4, pp. 385–397, Dec 1998.

[36] M. Mézard, M. Tarzia, and C. Toninelli, "Group testing with random pools: Phase transitions and optimal strategy," *Journal of Statistical Physics*, vol. 131, no. 5, pp. 783–801, 2008.

[37] A. G. D'yachkov, I. V. Vorobev, N. A. Polyansky, and V. Y. Shchukin, "Almost disjunctive list-decoding codes," *Problemy Peredachi Informatsii*, vol. 51, no. 2, pp. 27–49, 2015, translation: *Problems of Information Transmission*. 51(2): 110–131.

[38] S. Cai, M. Jahangoshahi, M. Bakshi, and S. Jaggi, "Efficient algorithms for noisy group testing," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2113–2136, 2017.

[39] K. Lee, R. Pedarsani, and K. Ramchandran, "SAFFRON: A fast, efficient, and robust framework for group testing based on sparse-graph codes," in *IEEE International Symposium on Information Theory*, 2016.

[40] H. Q. Ngo, E. Porat, and A. Rudra, "Efficiently decodable error-correcting list disjunct matrices and applications," in *International Colloquium on Automata, Languages and Programming*, 2011.

[41] P. Indyk, H. Q. Ngo, and A. Rudra, "Efficiently decodable non-adaptive group testing," in *ACM-SIAM Symposium on Discrete Algorithms*, 2010.

[42] J. Luo and D. Guo, "Neighbor discovery in wireless ad hoc networks based on group testing," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2008, pp. 791–797.

[43] D. Malioutov and M. Malyutov, "Boolean compressed sensing: Lp relaxation for group testing," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 3305–3308.

[44] M. Aldridge, "On the optimality of some group testing algorithms," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 3085–3089.

[45] C. McDiarmid, *On the method of bounded differences*, ser. London Mathematical Society Lecture Note Series. Cambridge University Press, 1989, p. 148188.

[46] E. H. Lieb, "Concavity properties and a generating function for Stirling numbers," *Journal of Combinatorial Theory*, vol. 5, no. 2, pp. 203 – 206, 1968.

[47] J. F. C. Kingman, "Uses of exchangeability," *Ann. Probab.*, vol. 6, no. 2, pp. 183–197, 04 1978.

[48] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[49] B. Rennie and A. Dobson, "On Stirling numbers of the second kind," *Journal of Combinatorial Theory*, vol. 7, no. 2, pp. 116 – 121, 1969.

[50] E. Neuman, "On generalized symmetric means and Stirling numbers of the second kind." *Applicationes Mathematicae*, vol. 18, no. 4, pp. 645–656, 1985.