Johnson, O. (2017). Strong converses for group testing in the finite blocklength regime. IEEE Transactions on Information Theory. DOI: 10.1109/TIT.2017.2697358

Strong converses for group testing from finite blocklength results

Oliver Johnson

***Abstract*—We prove new strong converse results in a variety of group testing settings, generalizing a result of Baldassini, Johnson and Aldridge. First, in the non-adaptive case, we mimic the hypothesis testing argument introduced in the finite blocklength channel coding regime by Polyanskiy, Poor and Verdú, and using joint source–channel coding arguments of Kostina and Verdú. In the adaptive case, we combine this approach with a novel model formulation based on causal probability and directed information theory. In both cases, we prove results which are valid for finite sized problems, and imply capacity results in the asymptotic regime. These results are illustrated graphically for a range of models.**

***Index Terms*—Group testing, converse bounds, finite blocklength, sparse models.**

I. Introduction and group testing model

The group testing problem was introduced by Dorfman [1] in the 1940s, and captures the idea of efficiently isolating a small subset $\mathcal{K}$ of defective items in a larger set containing N items. The models used vary slightly, but fundamentally we perform a sequence of tests, each defined by a testing pool of items, with the outcome of each depending on the number of defective items in the pool. The most basic model, which we refer to as 'standard noiseless group testing' is that the test outcome equals 1 if and only if the testing pool contains at least one defective item. Given T tests, the group testing problem requires us to design test pools and estimation algorithms to maximise $\mathbb{P}(\mathrm{suc})$, the success probability (probability of recovering the defective set exactly), where the randomness enters through the defectivity status of the items (see Definition I.2 below) and any noise in the measurements.

This paper focuses on converse results, giving upper bounds on the $\mathbb{P}(\mathrm{suc})$ that can be achieved by any algorithm given T tests. We generalize the following strong result proved by Baldassini, Johnson and Aldridge [2, Theorem 3.1]:

Theorem I.1. *Suppose the defective set $\mathcal{K}$ is chosen uniformly from the $\binom{N}{K}$ possible sets of given size K. For adaptive or non-adaptive standard noiseless group testing:*

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

As Figure 1 illustrates, Theorem I.1 is a strong result in this context, giving a converse which closely matches the achievability results provided by Hwang's algorithm [3]. This paper extends Theorem I.1 to a variety of settings. We first discuss four dichotomies in the modelling of the group testing problem. There are a number of further variations beyond these, as described in an ever-increasing body of literature.

1) **[Combinatorial vs Probabilistic]** First, consider how the defective items are chosen. Combinatorial group testing (see for example [2], [4], [5], [6]) is the model from Theorem I.1: we suppose the defective set $\mathcal{K}$ is chosen uniformly from the $\binom{N}{K}$ possible sets of fixed size K. In probabilistic group testing (see for example [7], [8]) the ith item is defective independently with probability p_i (with p_i not necessarily identical). In fact, we put both these models in a common setting:

   Definition I.2. *Write $\mathbf{U} \in \{0,1\}^N$ for the (random) defectivity vector, where component U_i is the indicator of the event that the ith item is defective. For any vector $\mathbf{u} \in \{0,1\}^N$ write $P_{\mathbf{U}}(\mathbf{u}) = \mathbb{P}(\mathbf{U} = \mathbf{u})$, and define entropy*

$$H(\mathbf{U}) = -\sum_{\mathbf{u}\in\{0,1\}^N} P_{\mathbf{U}}(\mathbf{u}) \log_2 P_{\mathbf{U}}(\mathbf{u}). \tag{2}$$

   Example I.3. *For the two models as described above:*
   a) *For combinatorial group testing, $\mathbf{U}$ is uniform over $\binom{N}{K}$ outcomes and $H(\mathbf{U}) = \log_2 \binom{N}{K}$.*
   b) *For probabilistic group testing, entropy $H(\mathbf{U}) = \sum_{i=1}^N h(p_i)$, where $h(t)$ is the binary entropy function. If $p_i \equiv p$ then $H(\mathbf{U}) = Nh(p)$.*

   We prove in Corollary IV.4 that results resembling Theorem I.1 hold for general sources satisfying the Shannon-McMillan-Breiman theorem. This includes settings where $\mathbf{U}$ is generated by a stationary ergodic Markov chain, which is a natural model of a setting where nearest neighbours are susceptible to infection.
2) **[Binary vs Non-binary]** Second, consider the set of possible outcomes $\mathcal{Y}$ of each test. We refer to $\mathcal{Y}$ as the alphabet, since in this paper (as in [4], [5] and other papers) we consider an analogy between group testing and channel coding problems. It is most standard to consider the binary case, where $\mathcal{Y} = \{0,1\}$, though other models are possible (see [9, Section 6.3] for a detailed review). For brevity this paper will only consider the binary case, though our techniques will work in a more general setting, and write $\mathbf{Y} \in \mathcal{Y}^T = \{0,1\}^T$ for the outcome of the group testing process.
3) **[Noisy vs Noiseless]** Third, consider how the outcome of each test is formed. To fix notation, we perform a sequence of T tests defined by test pools $\mathcal{X}_1, \ldots, \mathcal{X}_T$, where each $\mathcal{X}_t \subseteq \{1, 2, \ldots, N\}$. We represent this by a binary test matrix $\mathcal{X} = (x_{it} : i = 1, \ldots, N \text{ and } t =$

The author is with the School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK Email: maotj@bristol.ac.uk

$1, \dots, T$), where $x_{it} = 1$ if and only if item i is included in the t-thpool ($\mathcal{X}$ concatenates the column vectors given by the indicator functions of the T test pools). Since the test design may be random, we write X for a random variable giving a test matrix of this form.

For a test matrix X and defectivity vector $\mathbf{u}$, a key object of interest is the vector $\mathbf{K} = \mathbf{u}^T\mathsf{X}$, with t-thcomponent $K_t = \sum_{i=1}^{N} x_{it}\mathbb{I}(\text{item } i \in \mathcal{K})$, the total number of defective items appearing in the t-thtest. Observe that K_t is a deterministic function of $\mathbf{u}$ and X_t (and does not depend on any other variables). It is useful to define $\mathbf{X}$ via $X_t = \mathbb{I}(K_t \geq 1)$. We assume that the group testing model is static, memoryless and satisfies the 'Only Defects Matter' property introduced by Aldridge [10], [9]:

Definition I.4 (Only Defects Matter)**.** *Assume the t-thtest outcome Y_t is a random function of K_t (so $\mathbf{Y}$ is conditionally independent of $\mathbf{U}$ given $\mathbf{K}$, and Y_t is conditionally independent of $(K_s)_{s\neq t}$ given K_t). Further, for some fixed transition matrix P, we assume*

$$\mathbb{P}(Y_t = y|K_t = k) = P(y|k), \quad \textit{for all } y, k, t. \tag{3}$$

Note that Definition I.4 includes the noiseless standard group testing case, where we simply take $\mathbf{Y} = \mathbf{X}$. To understand Definition I.4, consider feeding $\mathbf{X}$ symbol-by-symbol through a memoryless channel, independent of defectivity vector $\mathbf{U}$, and group testing design X. In the notation of (3), we assume that $P(1|k) \equiv P(1|1)$ for all $k \geq 1$; in the noiseless case we take $P(1|k) \equiv 1$ and $P(0|0) = 1$. However, Definition I.4 allows a wider range of noise models, including the dilution channel of Atia and Saligrama [5], where we take $P(0|k) = (1-u)^k$ for some u.

For a fixed test matrix $\mathsf{X} = \mathcal{X}$, as in [2], in the noiseless case the testing procedure naturally defines a mapping $\boldsymbol{\theta}(\cdot, \mathcal{X}) : \{0,1\}^N \to \{0,1\}^T$. That is, given defectivity vector $\mathbf{u} \in \{0,1\}^N$, we write the vector function $\boldsymbol{\theta}$ with components given by scalar function θ defined as $\theta(\mathbf{u}, \mathcal{X}_t) = \mathbb{I}(K_t \geq 1) = X_t$. That is:

$$\boldsymbol{\theta}(\mathbf{u}, \mathcal{X}) = (\theta(\mathbf{u}, \mathcal{X}_1), \theta(\mathbf{u}, \mathcal{X}_2), \dots, , \theta(\mathbf{u}, \mathcal{X}_T)) . \tag{4}$$

4) **[Adaptive vs Non-adaptive]** The final distinction is whether we design the test matrix using an adaptive or a non-adaptive strategy. In the non-adaptive case the entire test matrix $\mathsf{X} = \mathcal{X}$ needs to be chosen in advance of the tests. In contrast, in the adaptive case, the $(t+1)$st test pool $\mathcal{X}_{t+1}$ is chosen based on a knowledge of previous test pools $\mathcal{X}_{1,t} := \{\mathcal{X}_1, \dots, \mathcal{X}_t\}$ and test outcomes $\mathbf{Y}_{1,t} := \{Y_1, \dots, Y_t\}$. We can think (see [9]) that adaptive group testing corresponds to joint source–channel coding with feedback, and non-adaptive group testing to coding with no feedback. Clearly (see [2]), we can do no worse in the adaptive setting than for non-adaptive group testing, but it remains an open and interesting question to determine precisely by how much adaptivity can improve performance.

We argue that a key tool in understanding adaptive group testing is directed information theory. This was first introduced by Marko [11] in the 1970s, with interest revived by the work of Massey [12] in the 1990s, and developed further by authors such as Kramer [13]. In particular, as described by Massey [12], many authors make an incorrect probabilistic formulation of such simple objects as discrete memoryless channels with feedback. A correct formulation requires the use of the causal conditional probability distribution. We use the notation of the review paper [14, Equation (7)], that for sequences $\mathbf{x} = (x_1, \dots, x_T)$ and $\mathbf{y} = (y_1, \dots, y_T)$, and subsequences $\mathbf{y}_{1,t-1} = (y_1, \dots, y_{t-1})$,

$$P_{\mathbf{X}\|\mathbf{Y}}(\mathbf{x}\|\mathbf{y}) := \prod_{t=1}^{T} P_{X_t|\mathbf{X}_{1,t-1},\mathbf{Y}_{1,t}}(x_t|\mathbf{x}_{1,t-1}, \mathbf{y}_{1,t}) . \tag{5}$$

For any fixed $\mathbf{y}$, the fact that (5) is formed as a product of probability distributions means that $\sum_{\mathbf{x}} P_{\mathbf{X}\|\mathbf{Y}}(\mathbf{x}\|\mathbf{y}) = 1$. Using this probability distribution implies the form of the directed information of Marko [11].

In Lemma IV.1 below, assuming the Only Defects Matter property Definition I.4, we decompose the joint probability of $(\mathbf{U}, \mathcal{X}, \mathbf{Y})$ in the general adaptive setting, using the term $P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-)$, which is defined in (16). Here we adapty the causal conditional probability notation (5) above, with superscript $\mathbf{y}^-$ referring to the fact that there is a lag in the index of $\mathbf{y}$ (we choose the set $\mathcal{X}_i$ based on a knowledge of the previous sets $\mathcal{X}_{1,t-1}$ and test outcomes $\mathbf{y}_{1,t-1}$).

Lemma IV.1 shows the t-thoutput symbol Y_t is conditionally independent of $\mathbf{U}$, given values $K_{1,t}$ and previous outputs $\mathbf{Y}_{1,t-1}$. This is precisely the definition of a causal system between $\mathbf{K}$ and $\mathbf{Y}$ given by Massey in [12, Equation (8)], under which condition feedback does not increase the capacity of a discrete memoryless channel.

Regardless of these variations, we always make an estimate $\mathbf{Z} = \widehat{\mathbf{U}}$, based only on a knowledge of outputs $\mathbf{Y} = \mathbf{y}$ and test matrix $\mathsf{X} = \mathcal{X}$, using a probabilistic estimator (decoder) that gives $\mathbf{Z} = \mathbf{z}$ with probability

$$P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{z}|\mathbf{y}, \mathcal{X}) . \tag{6}$$

The main results of the paper are Theorem III.2, which gives an upper bound on $\mathbb{P}(\mathrm{suc})$ in the non-adaptive case, and Theorem IV.2, which gives the corresponding result in the adaptive case. The strength of these results is illustrated in results such as Example V.4, where we calculate bounds on the success probability in the case where $\mathbf{X}$ forms the input and $\mathbf{Y}$ the output of a binary symmetric channel with error probability p, with the resulting bounds being plotted in Figure 3.

The structure of the paper is as follows. In Section II we review existing results concerning group testing converses. In Section III we use an argument based on the papers [15] and [16] to prove Theorem III.2, which implies a strong converse for non-adaptive group testing. In Section IV we discuss the adaptive case, by extending these arguments using the causal probability formulation described above. We prove a bound (Theorem IV.2) which specializes in the noiseless case to give

a result (Theorem IV.3) which generalizes Theorem I.1. We consider examples of this noiseless result in the probabilistic case in Section V-A. Finally in Section V-B we apply Theorem IV.2 in the noisy adaptive case. The proofs of the main theorems are given in Appendices.

While this paper only considers group testing, we remark that group testing lies in the area of sparse inference, which includes problems such as compressed sensing and matrix completion, as reviewed by [17]. It is likely that results proved here will extend to more general settings. Group testing itself has a number of applications, including cognitive radios [10], [18], [7], network tomography [19] and efficient gene sequencing [20], [21]. The bounds proved here should provide fundamental performance limits in these contexts.

II. Existing converse results

Information-theoretic considerations mean that to find all the defectives in the noiseless case will require at least $T^* = H(\mathbf{U})$ (the "magic number") tests. In the language of channel coding, the focus of this paper is on converse results; that is given $O(T^*)$ tests, we give strong upper bounds on the success probability $\mathbb{P}(\mathrm{suc})$ of any possible algorithm.

There has been considerable work on the achievability part of the problem, in developing group testing algorithms and proving performance guarantees. Early work on group testing considered algorithms which could be proved to be order optimal (see for example the analysis of [22], [23], [24]), often using combinatorial properties such as separability or disjunctness. More recently there has been interest (see for example [4], [5], [6], [8], [25], [26], [27], [28], [29], [30]) in finding the best possible constant, that is to find algorithms which succeed with high probability using $T = cT^* = cH(\mathbf{U})$ tests, for c as small as possible. In this context, the paper [2] defined the capacity of combinatorial group testing problems, a definition extended to both combinatorial and probabilistic group testing in [7]. We state this definition for both weak and strong capacity in the sense of Wolfowitz:

Definition II.1. *Consider a sequence of group testing problems where the ith problem has defectivity vector $\mathbf{U}^{(i)}$, and consider algorithms which are given $T(i)$ tests. We think of $H(\mathbf{U}^{(i)})/T(i)$ (the number of bits of information learned per test) as the rate of the algorithm and refer to a constant C as the weak group testing capacity if for any $\epsilon > 0$:*

1) *any sequence of algorithms with*

$$\liminf_{i\to\infty} \frac{H(\mathbf{U}^{(i)})}{T(i)} \geq C + \epsilon, \tag{7}$$

has success probability satisfying $\limsup_{i\to\infty} \mathbb{P}(\mathrm{suc}) < 1$,

2) *and there exists a sequence of algorithms with*

$$\liminf_{i\to\infty} \frac{H(\mathbf{U}^{(i)})}{T(i)} \geq C - \epsilon \tag{8}$$

with success probability satisfying $\lim_{i\to\infty} \mathbb{P}(\mathrm{suc}) = 1$.

C is the strong capacity if $\lim_{i\to\infty} \mathbb{P}(\mathrm{suc}) = 0$ for any sequence of algorithms satisfying (7)*.*

For example, in [2] we proved that noiseless adaptive combinatorial group testing has strong capacity 1. This result is proved by combining Hwang's Generalized Binary Splitting Algorithm [3] (which is essentially optimal – see also [2], [23] for a discussion of this) with the converse result Theorem I.1. However, even in the noiseless non-adaptive case the capacity remains unknown in general, although some results are known in some regimes, under assumptions about the distribution of X (see for example [25], [4], [28], [29]).

Capacity results are asymptotic in character, whereas we will consider the finite blocklength regime (in the spirit of [31], [15]) and prove bounds on $\mathbb{P}(\mathrm{suc})$ for any size of problem. We briefly review existing converse results. First, we mention that results (often referred to as folklore) can be proved using arguments based on Fano's inequality.

Lemma II.2. *Using T tests in the noiseless case:*

1) *For combinatorial group testing Chan et al. [6, Theorem 1] give*

$$\mathbb{P}(\mathrm{suc}) \leq \frac{T}{\log_2 \binom{N}{K}}. \tag{9}$$

2) *For probabilistic group testing Li et al. [8, Theorem 1] give*

$$\mathbb{P}(\mathrm{suc}) \leq \frac{T}{Nh(p)}. \tag{10}$$

In order to understand the relationship between (9) and Theorem I.1; fix $\delta > 0$ and use $T = T^*(1-\delta)$ tests, in a regime where $\log_2 \binom{N}{K} \to \infty$ and hence $T^* \to \infty$. Chan et al.'s result (9) gives that $\mathbb{P}(\mathrm{suc}) \leq (1-\delta)$, whereas (1) implies $\mathbb{P}(\mathrm{suc}) \leq 2^{-\delta T^*}$. In the language of Definition II.1, Chan et al. [6] give a weak converse whereas Baldassini, Johnson and Aldridge [2] give a strong converse. In fact, (1) shows that the success probability converges to zero exponentially fast.

To understand why Chan et al's result (9) is not as strong as Theorem I.1 we examine the proof, which uses Fano's inequality, bounding the entropy $H(\mathbf{U}|\mathbf{Y})$ in a standard way using

$$H(\mathbf{U}|\mathbf{Y}) \quad \leq \quad 1 + \mathbb{P}(E=1) H(\mathbf{U}|\mathbf{Y}, E=1) \tag{11}$$

where E is the indicator of the error event $\mathbf{U} \neq \mathbf{Z}$. In [6] this last term is bounded by $\log_2 \binom{N}{K}$, since a priori $\mathbf{U}$ could be any defective set. However, in practice, this is a significant overestimate. For example, in the noiseless case there is a relatively small collection of defective sets that a particular defective set $\mathbf{U}$ can mistakenly be estimated as (referred to as $A(\cdot)$ later in this paper). For example, any item appearing in a test pool $\mathcal{X}_t$ giving result $Y_t = 0$ cannot be defective. Essentially, Theorem I.1 exploits such facts.

Tan and Atia [32, Theorem 2] prove a strong converse for combinatorial group testing, however, they do not achieve exponential decay. Since the result of the test only depends on whether the items in the defective set $\mathcal{K}$ are present, we can restrict our attention to the submatrix $\mathsf{X}_{\mathcal{L}}$ indexed by subsets $\mathcal{L} \subseteq \mathcal{K}$.

Theorem II.3 ([32], Theorem 2)**.** *Define parameters $\zeta_T := T^{-1/4}\mathbb{P}(\mathrm{suc})^{-1/2}$ and $\eta_T := T^{-1} + h(T^{-1/4})$. If the components of X are independent and identically distributed then for*

each $\mathcal{L}$, then T (the number of tests required to achieve the given probability of success) satisfies:

$$T(I(\mathsf{X}_{\mathcal{K}\setminus\mathcal{L}};\mathsf{X}_{\mathcal{L}},\mathbf{Y})+\eta_T)\geq(1-\zeta_T)\log_2\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}.$$

Rearranging, and writing $I=I(\mathsf{X}_{\mathcal{K}\setminus\mathcal{L}};\mathsf{X}_{\mathcal{L}},\mathbf{Y})$ we obtain

$$\mathbb{P}(\text{suc})\leq\frac{1}{T^{1/2}\left(1-T(I+\eta_T)/\log_2\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}\right)^2}\sim\frac{1}{\delta^2T^{1/2}},$$

taking $T^*=\log_2\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}/I$ for $T=(1-\delta)T^*$. This gives a strong converse, though not the exponential decay achieved in (1) above. However Tan and Atia's results [32] are valid in a variety of settings and noise models.

Pedagogically, we note a parallel between these various approaches and treatments of the channel coding problem in the literature. That is (9), due to Chan et al. [6], is proved using Fano's inequality, parallelling the proof of Shannon's noisy coding theorem exemplified for example in [33, Section 8.9]. The argument of Tan and Atia [32] is based on Marton's blowing up lemma, mirroring the treatment of Shannon's theorem in the book of Csiszár and Körner [34, Section 6].

Our work in the non-adaptive case is based on the more recent work of Polyanskiy, Poor and Verdú [15], which has been adapted to the problem of data compression in [35]. We remark that Scarlett and Cevher use an approach parallel to ours, based on the Verdú–Han information spectrum method of [36]. This gives results such as [28, Corollary 9], for example, which can be adapted to deliver exponential convergence (on replacing Chebyshev's inequality in the proof by a result such as Chernoff's inequality). It is important to remark that the link between channel coding and hypothesis testing was first identified in a quantum information context by Hayashi and Nagaoka [31], with later work [37] by Hayashi clarifying this approach. Further, the work of Nagaoka [38] used similar ideas to derive strong converse results.

The paper [39] extends the approach of [15] to channels with feedback, corresponding to adaptive group testing. We remark that information spectrum methods were first introduced for channels with feedback by Wolfowitz [40] (see also Gallager [41]).

III. Hypothesis testing and non-adaptive group testing

We first state a result, Theorem III.2, which implies a generalization of Theorem I.1 in the non-adaptive case (Corollary III.3), and allows us to deduce strong converse results. The key observation comes from Polyanskiy, Poor and Verdú [15] (see also [31]) who used a relationship between channel coding and hypothesis testing. Since the Neyman–Pearson lemma gives the optimal hypothesis test, [15] deduces strong bounds on coding error probabilities.

Definition III.1. *Write $\beta_{1-\epsilon}(P,Q)$ for the smallest possible type II error for hypothesis tests (with type I error probability $\leq\epsilon$) deciding between P and Q.*

Note that as in [16], we do not assume that Q is a probability measure, and simply think of $\beta_{1-\epsilon}$ as the maximum value of $Q($ do not reject null $\mid$ alternative is true$)$.

We use the same analogy as [15] for the group testing problem, given a process generating random chosen defective sets $\mathbf{U}$ (a source). To some extent this is simply a question of adapting the notation of [15]. However, unlike [15] we do not require that $\mathbf{U}$ is uniform (allowing us to consider probabilistic as well as combinatorial group testing). In a pure channel coding scenario it seems less natural to consider non-uniform $\mathbf{U}$, however such $\mathbf{U}$ were considered in [16], where a list-decoding argument was used to analyse the joint source–channel coding problem. In fact, we combine the approaches of [16] and [15].

Since we consider non-adaptive group testing, we fix $\mathsf{X}=\mathcal{X}$ in advance. We identify the Only Defects Matter property, Definition I.4, as playing a key role. We write $P_{\mathbf{KY}}(\mathbf{k},\mathbf{y})$ for the joint probability distribution of $\mathbf{K}$ and $\mathbf{Y}$ and consider an algorithm which estimates (decodes) the defective set $\mathbf{Z}=\widehat{\mathbf{U}}$, using only outputs $\mathbf{Y}$ and test matrix $\mathsf{X}=\mathcal{X}$. Since $\mathsf{X}=\mathcal{X}$ is fixed, we can simplify (6) above and write $P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})$ for the probability that the estimator gives $\mathbf{Z}=\mathbf{z}$ when $\mathbf{Y}=\mathbf{y}$. We prove the following result in Appendix A:

Theorem III.2. *Suppose that the group testing model satisfies the Only Defects Matter property, Definition I.4. For any non-adaptive choice of test design, any estimation rule $P_{\mathbf{Z}|\mathbf{Y}}$ with $\mathbb{P}(\text{suc})\geq1-\epsilon$ and probability mass function $Q_{\mathbf{Y}}$ satisfies:*

$$\beta_{1-\epsilon}(P_{\mathbf{KY}},P_{\mathbf{K}}\times Q_{\mathbf{Y}})\leq\sum_{\mathbf{z}\in\{0,1\}^N}P_{\mathbf{U}}(\mathbf{z})\,Q^*(\mathbf{z}),\qquad(12)$$

where $Q^*(\mathbf{z})=\sum_{\mathbf{y}}Q_{\mathbf{Y}}(\mathbf{y})\,P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})$ is the probability that $\mathbf{Y}\sim Q_{\mathbf{Y}}$ is decoded to $\mathbf{z}$.

Corollary III.3. *In the noiseless non-adaptive case consider any defective set distribution $P_{\mathbf{U}}$ and write $\Pi_{\mathbf{U}}(m)$ for the sum of the largest m values of $P_{\mathbf{U}}(\mathbf{z})$. Then*

$$\mathbb{P}(\text{suc})\leq\Pi_{\mathbf{U}}(2^T).\qquad(13)$$

Proof. See Appendix A. □

In particular, Corollary III.3 extends Theorem I.1 under the additional assumption of non-adaptivity; we discuss how to remove this assumption in Theorem IV.3 below. That is, if for some set $\mathcal{M}$ of size $|\mathcal{M}|$, the $P_{\mathbf{U}}(\mathbf{z})=\mathbb{I}(\mathbf{z}\in\mathcal{M})/|\mathcal{M}|$ then

$$\mathbb{P}(\text{suc})\leq\frac{2^T}{|\mathcal{M}|}.\qquad(14)$$

In retrospect, perhaps this result is not surprising; we think of an optimal list decoding by simply choosing the defective set of highest probability compatible with each outcome $\mathbf{y}$.

Theorem III.2 also implies a converse for the non-adaptive binary symmetric channel case, which will hold more generally in the adaptive case. We discuss this in Section V-B below (since an upper bound on success probabilities for adaptive group testing implies an upper bound for non-adaptive group testing).

IV. Adaptive group testing

As discussed in Section I, a precise formulation of adaptive group testing (corresponding to channels with feedback) requires the use of directed probability distributions and information theory. For any t, we write $\mathbf{Y}_{1,t} = \{Y_1, \ldots, Y_t\}$ and $\mathcal{X}_{1,t} = \{\mathcal{X}_1, \ldots, \mathcal{X}_t\}$. We first prove the following representation of the joint probability distribution of $(\mathbf{U}, \mathsf{X}, \mathbf{Y})$ for adaptive group testing:

Lemma IV.1. *Assuming the Only Defects Matter property (Definition I.4) with transition matrix $\mathbb{P}(Y_t = y | K_t = k) = P(y|k)$ for all k, y, t, we can write*

$$P_{\mathbf{U},\mathsf{X},\mathbf{Y}}(\mathbf{u}, \mathcal{X}, \mathbf{y}) = P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}\left(\mathcal{X}\|\mathbf{y}^-\right) \prod_{t=1}^{T} P(y_t|k_t), \tag{15}$$

where $k_t = \mathbf{u}.\mathcal{X}_t$ (matrix product) is the number of defectives in the t-thtest and

$$P_{\mathsf{X}\|\mathbf{Y}^-}\left(\mathcal{X}\|\mathbf{y}^-\right) := \prod_{t=1}^{T} P_{\mathsf{X}_t|\mathbf{Y}_{1,t-1},\mathsf{X}_{1,t-1}}\left(\mathcal{X}_t|\mathbf{y}_{1,t-1}, \mathcal{X}_{1,t-1}\right) \tag{16}$$

is the causal conditional probability, with the key property that for any fixed $\mathbf{y}$:

$$\sum_{\mathcal{X}} P_{\mathsf{X}\|\mathbf{Y}^-}\left(\mathcal{X}\|\mathbf{y}^-\right) = 1. \tag{17}$$

Proof. We write (omitting the subscripts on $\mathbb{P}$ for brevity) $\mathbb{P}(\mathbf{u}, \mathcal{X}, \mathbf{y})/\mathbb{P}(\mathbf{u})$ as a collapsing product of the form:

$$\begin{aligned}
&\prod_{t=1}^{T} \frac{\mathbb{P}\left(\mathbf{u}, \mathcal{X}_{1,t}, \mathbf{y}_{1,t}\right)}{\mathbb{P}\left(\mathbf{u}, \mathcal{X}_{1,t-1}, \mathbf{y}_{1,t-1}\right)} \\
&= \prod_{t=1}^{T} \mathbb{P}\left(\mathcal{X}_t, y_t | \mathbf{u}, \mathcal{X}_{1,t-1}, \mathbf{y}_{1,t-1}\right) \\
&= \prod_{t=1}^{T} \mathbb{P}\left(y_t | \mathcal{X}_t, \mathbf{u}, \mathcal{X}_{1,t-1}, \mathbf{y}_{1,t-1}\right) \mathbb{P}\left(\mathcal{X}_t | \mathbf{u}, \mathcal{X}_{1,t-1}, \mathbf{y}_{1,t-1}\right) \\
&= \prod_{t=1}^{T} \mathbb{P}\left(y_t | k_t\right) \mathbb{P}\left(\mathcal{X}_t | \mathcal{X}_{1,t-1}, \mathbf{y}_{1,t-1}\right)
\end{aligned}$$

where we remove the conditioning in the final line since y_t is the result of sending $k_t = \mathbf{U}^T \mathcal{X}_t$ through a memoryless channel (the output of which is independent of previous test designs and their output) and since the choice of the t-thtest pool $\mathcal{X}_t$ is conditionally independent of $\mathbf{U}$, given the previous tests and their output. □

Recall from (6) that we write $P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{z}|\mathbf{y}, \mathcal{X})$ for the probability that some algorithm estimates the defective set as $\widehat{\mathbf{U}} = \mathbf{Z} = \mathbf{z} \in \{0,1\}^N$ when the group testing process with test matrix $\mathsf{X} = \mathcal{X}$ returns $\mathbf{Y} = \mathbf{y}$. We write $S^*_{\mathcal{X}}(\mathbf{y})$ for the set of values that $\mathbf{y}$ may be decoded to; that is $\mathbf{u} \in S^*_{\mathcal{X}}(\mathbf{y})$ if and only if $P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{y}, \mathcal{X}) > 0$.

Theorem IV.2. *Take any probability mass function $Q_{\mathbf{Y}}$ on $\{0,1\}^T$. For any model of group testing (adaptive or non-adaptive), satisfying the Only Defects Matter property Definition I.4, if there is a decoding rule $P_{\mathbf{Z}|\mathbf{Y},\mathcal{X}}$ with success probability $\mathbb{P}(\text{suc}) \geq 1 - \epsilon$ and a probability measure $Q_{\mathbf{U}}$ such that $Q_{\mathbf{U}}\left(S^*_{\mathcal{X}}(\mathbf{y})\right) \leq L$ for all $\mathbf{y}$, then*

$$\beta_{1-\epsilon}(P_{\mathbf{U}} P_{\mathsf{X}\|\mathbf{Y}^-} P_{\mathbf{Y}|\mathbf{K}}, Q_{\mathbf{U}} P_{\mathsf{X}\|\mathbf{Y}^-} Q_{\mathbf{Y}}) \leq L. \tag{18}$$

Proof. See Appendix B. □

Note that this result resembles the list decoding version of the meta-converse [16, Lemma 4]. We use arguments based on Theorem IV.2 to prove a result which extends Theorem I.1 for general defective set distributions $P_{\mathbf{U}}$ in the noiseless binary case. This result applies to both adaptive and non-adaptive group testing.

Theorem IV.3. *For noiseless adaptive binary group testing, if we write $\Pi_{\mathbf{U}}(m)$ for the sum of the largest m values of $P_{\mathbf{U}}(\mathbf{z})$ then*

$$\mathbb{P}(\text{suc}) \leq \Pi_{\mathbf{U}}(2^T).$$

Proof. See Appendix C. □

For combinatorial group testing, since $P_{\mathbf{U}}$ is uniform on a set of size $\binom{N}{K}$, Theorem IV.3 implies that $\Pi_{\mathbf{U}}(m) = m/\binom{N}{K}$ and we recover Theorem I.1. We show how sharp this result is in Figure 1, which is reproduced from [2, Figure 1].

Corollary IV.4. *Consider a sequence $\mathbf{U}^{(i)}$ of defectivity vectors of length i, generated as independent realisations of a stationary ergodic stochastic process of entropy rate H. Given $T^{(i)} = (H - \epsilon)i$ tests to solve the ith noiseless adaptive group testing problem, the success probability tends to zero. (Hence the strong capacity cannot be more than 1).*

Proof. We define the typical set

$$\mathcal{T}^{(i)}_{\epsilon} = \left\{ \left| \frac{-\log P_{\mathbf{U}^{(i)}}(\mathbf{u})}{i} - H \right| \leq \frac{\epsilon}{2} \right\} \tag{19}$$

By the Shannon-McMillan-Breiman theorem (AEP) (see for example [33, Theorem 15.7.1]), the probability $\mathbb{P}(\mathcal{T}^{(i)}_{\epsilon}) \to 1$. Then, in Theorem IV.3, the $2^{T(i)}$ strings of largest probability will certainly be contained in a list containing the elements of $(\mathcal{T}^{(i)}_{\epsilon})^c$ and the $2^{T(i)}$ strings of largest probability in $\mathcal{T}^{(i)}_{\epsilon}$. Since, by definition, any string in $\mathcal{T}^{(i)}_{\epsilon}$ has probability less than $2^{-iH + i\epsilon/2}$, we deduce that

$$\begin{aligned}
\mathbb{P}(\text{suc}) \leq \Pi_{\mathbf{U}^{(i)}}(2^{T(i)}) &\leq \mathbb{P}\left((\mathcal{T}^{(i)}_{\epsilon})^c\right) + 2^{T(i)} 2^{-iH + i\epsilon/2} \\
&= \mathbb{P}\left((\mathcal{T}^{(i)}_{\epsilon})^c\right) + 2^{-i\epsilon/2}
\end{aligned}$$

Given a quantitative form of the Shannon-McMillan-Breiman theorem (proved for example using the concentration inequalities described in [42]), we can deduce an explicit (exponential) rate of convergence to zero of $\mathbb{P}(\text{suc})$. □

Note that (since it is proved using concentration inequalities only), although this result gives a strong converse, it may do so with a sub-optimal exponent (rate of convergence). It remains of interest to categorise the optimal strong converse exponent in these problems.

We give more explicit bounds which show how Theorem IV.3 can be applied in the noiseless probabilistic case in Section V-A below. Section V-B contains an illustrative example of results that can be proved using Theorem IV.2, in the noisy

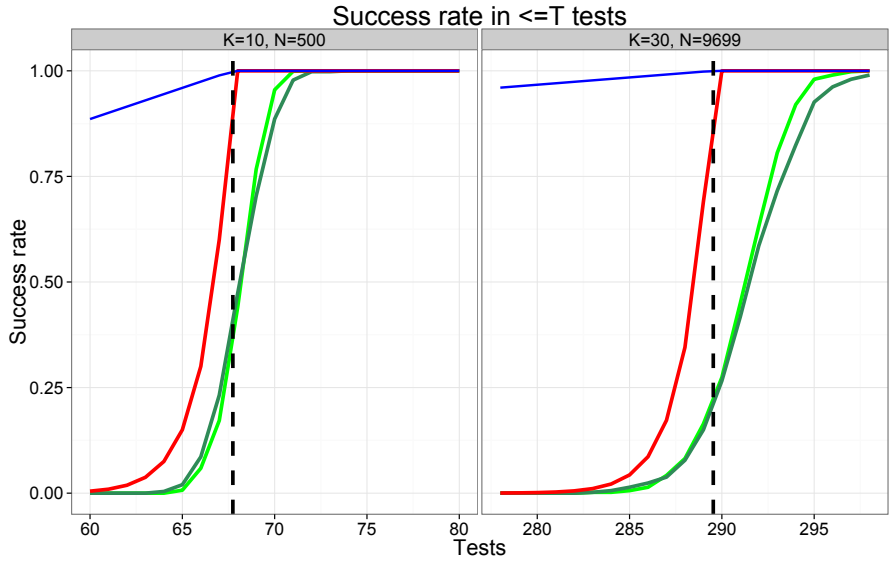


Fig. 1. (Reproduced from [2, Figure 1]). Success probability for noiseless adaptive combinatorial group testing with $(K, N) = (10, 500)$ and $(30, 9699)$ (these numbers are chosen to match the regime $K = N^{1-\beta}$, as in Figure 3). The upper bound on success probability of Theorem I.1 is plotted in red, and the upper bound (9) (from [6, Equation (6)]) in blue. The dotted vertical line is at $\log_2 \binom{N}{K}$ (the magic number). The converse bound of Theorem I.1 closely matches simulated achievability results of practical algorithms. The empirical success probability of the HGBSA of Hwang [3] is plotted as a bright green line, and the related algorithm analysed in [2, Section IV] is plotted in dark green.

adaptive case Baldassini's thesis [43] developed and analysed algorithms in the noisy adaptive case. However, it remains an open problem to find capacity-achieving algorithms, even for binary symmetric noise.

V. Adaptive group testing examples

A. Noiseless adaptive probabilistic group testing

In this section, we give an example of bounds which can be proved using Theorem IV.3 for noiseless adaptive probabilistic group testing. Note that the control of the source strings with highest probabilities is an operation that lies at the analysis of the finite blocklength data compression problem in [35].

Example V.1. *We consider the identical Probabilistic case, where $p_i \equiv p < 1/2$, so $P_{\mathbf{U}}(\mathbf{z}) = p^w(1-p)^{N-w}$, where $w = w(\mathbf{z})$ is the Hamming weight of $\mathbf{z}$. Write*

$$L^*_{N,T} := \min\left\{L : \sum_{i=0}^{L}\binom{N}{i} \geq 2^T\right\} \tag{20}$$

and define $s \geq 0$ via

$$2^T = \sum_{i=0}^{L^*_{N,T}-1}\binom{N}{i} + s, \tag{21}$$

meaning the 2^T highest probability defective sets are all of those of weight $\leq L^*_{N,T} - 1$, plus s of weight $L^*_{N,T}$. We evaluate $\Pi_{\mathbf{U}}(2^T)$ to obtain a bound on which we plot in Figure 2:

$$\Pi_{\mathbf{U}}(2^T) = \sum_{i=0}^{L^*_{N,T}-1}\binom{N}{i}p^i(1-p)^{N-i} + sp^{L^*_{N,T}}(1-p)^{N-L^*_{N,T}}. \tag{22}$$

Remark V.2. *We give a Gaussian approximation to the bound* (22)*, in the spirit of [15]. Since we need to control tail probabilities we use the approximation given by Chernoff bounds (see Theorem D.1). If $L = L(y) := Np + y\sqrt{Np(1-p)}$ and $T(y) = Nh(L(y)/N)$ then* (39) *gives*

$$\mathbb{P}(\mathrm{Bin}(N, 1/2) \leq L(y)) \simeq 2^{-N+T(y)}, \tag{23}$$

giving an approximate solution to (20) *(as discussed in Appendix D, here $\simeq$ denotes equality on an exponential scale). Substituting in* (22) *and using a second normal approximation, we obtain an approximation in parametric form, that with $T(y)$ tests the*

$$\mathbb{P}(\mathrm{suc}) \leq \Pi_{\mathbf{U}}(2^{T(y)}) \simeq \mathbb{P}(\mathrm{Bin}(N, p) \leq L(y)) = \Phi(y) + o(1). \tag{24}$$

For example, if $y = 0$ then $T = T(0) = Nh(p)$ (the magic number) and $L = Np$, and $|\Pi_{\mathbf{U}}(2^T) - 1/2| = o(1)$.

Indeed using the Chernoff bound, we use (24) to deduce a strong capacity result:

Corollary V.3. *Noiseless binary probabilistic group testing has strong capacity $C = 1$ in any regime where $p \to 0$ and $Np \to \infty$.*

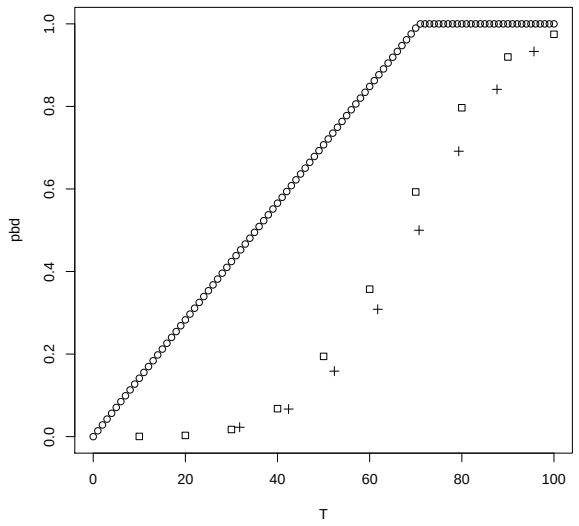


Fig. 2. Noiseless probabilistic adaptive group testing in the case of $p_i \equiv 1/50$, $N = 500$. We vary the number of tests T between 0 and 100, and plot the success probability on the y axis. We plot the upper bound on $\mathbb{P}(\text{suc})$ given by (22) using $\square$. For comparison, we plot the (weaker) Fano bound (10) of [8] as $\circ$. The approximation (24) is plotted as $+$.

Sketch proof. For any $p \leq 1/2$ and $\epsilon > 0$, we consider the asymptotic regime where $T = Nh(p-\epsilon)$ as $N \to \infty$. Choosing $L = N(p - \epsilon/2)$, we know that using standard bounds (see for example [33, Equation (12.40)])

$$\sum_{i=0}^{L} \binom{N}{i} \geq \binom{N}{L} \geq \frac{2^{Nh(L/N)}}{N+1} \geq \frac{2^{Nh(p-\epsilon/2)}}{N+1},$$

which is larger than 2^T in the asymptotic regime. Hence, summing over the strings of weight $\leq L$ will give at least the 2^T strings of highest probability, and we deduce by Theorem D.1 that

$$\mathbb{P}(\text{suc}) \leq \mathbb{P}(\text{Bin}(N,p) \leq N(p-\epsilon/2)) \leq 2^{-ND(p-\epsilon/2\|p)},$$

which tends to zero exponentially fast. This complements the performance guarantee proved in [7], strengthening the result of [7, Corollary 1.5] where the corresponding weak capacity result was stated using (10). $\square$

B. Noisy adaptive group testing example

We now use Theorem IV.2 to prove a bound on $\mathbb{P}(\text{suc})$ in a noisy example. For simplicity we state the following example in the case of uniform $\mathbf{U}$. Further generalizations (in the spirit of Theorem IV.3) are possible by adapting the proofs along the lines of Section C.

Example V.4. *Suppose $\mathbf{U}$ is uniformly distributed on a set $\mathcal{M}$ of size M and suppose the output of standard combinatorial noiseless non-adaptive group testing $\mathbf{X}$ is fed through a memoryless binary symmetric channel with error probability $p < 1/2$ to produce $\mathbf{Y}$. We write $x_i = \mathbb{I}(k_i \geq 1)$, and observe that $P(y_i|k_i) = (1-p)^{T-d(x_i,y_i)}p^{d(x_i,y_i)}$, where d represents the Hamming distance.*

Consider the optimal rule for deciding between null hypothesis $P_{\mathbf{U}}P_{\mathbf{X}\|\mathbf{Y}^-}P_{\mathbf{Y}|\mathbf{K}}$ and alternative $Q_{\mathbf{U}}P_{\mathbf{X}\|\mathbf{Y}^-}Q_{\mathbf{Y}}$. If $Q_{\mathbf{U}} = P_{\mathbf{U}}$ and $Q_{\mathbf{Y}}(\mathbf{y}) \equiv 1/2^T$, the likelihood ratio is

$$\begin{aligned}\frac{P_{\mathbf{U}}(\mathbf{u})\,P_{\mathbf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-)\,P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k})}{Q_{\mathbf{U}}(\mathbf{u})\,P_{\mathbf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-)\,Q_{\mathbf{Y}}(\mathbf{y})} &= \frac{P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k})}{Q_{\mathbf{Y}}(\mathbf{y})} \\ &\propto \left(\frac{p}{1-p}\right)^{d(\mathbf{x},\mathbf{y})}.\end{aligned}$$

By the Neyman–Pearson lemma, the optimal rule is to accept the null if $d(\mathbf{x},\mathbf{y}) < d^*$, to accept the null with probability λ if $d(\mathbf{x},\mathbf{y}) = d^*$ and to reject the null otherwise, where we calculate d^* and λ as follows:

$$\begin{aligned}\frac{1}{M} &\geq \beta_{1-\epsilon}(P_{\mathbf{U}}P_{\mathbf{X}\|\mathbf{Y}^-}P_{\mathbf{Y}|\mathbf{K}}, Q_{\mathbf{U}}P_{\mathbf{X}\|\mathbf{Y}^-}Q_{\mathbf{Y}}) \\ &= \mathbb{P}(\text{Bin}(T,1/2) \leq d^*-1) + \lambda\mathbb{P}(\text{Bin}(T,1/2) = d^*), \quad (25)\end{aligned}$$

where the first inequality follows from Theorem IV.2. Then, for this value of d^* we write that

$$\begin{aligned}\mathbb{P}(\text{suc}) &= 1 - \mathbb{P}(\textit{ type I error }) \\ &= \sum_{\mathbf{x},\mathbf{y}} P_{\mathbf{KY}}(\mathbf{k},\mathbf{y})\,\mathbb{P}(\textit{ accept } P_{\mathbf{KY}}) \\ &= \mathbb{P}(\text{Bin}(T,p) \leq d^*-1) + \lambda\mathbb{P}(\text{Bin}(T,p) = d^*). \quad (26)\end{aligned}$$

In Figure 3(a), we plot this in the case $N = 500$, $K = 10$, $p = 0.11$, and for comparison plot the Fano bound taken from [6, Theorem 2]:

$$\mathbb{P}(\text{suc}) \leq \frac{T(1-h(p))}{\log_2 \binom{N}{K}}. \quad (27)$$

In Figure 3(b) we give the group testing analogue of [15, Figure 1]. We use the regime of [4]; that is, we vary N and take $K = \lceil N^{1-\beta} \rceil$, where $\beta = 0.37$ (this gives the value $K = 10$ for $N = 500$). Again taking $p = 0.11$, we fix $\mathbb{P}(\text{suc}) = 0.999$, and use the lower bound on T corresponding to the analysis above. This gives an upper bound on the rate $\log_2 \binom{N}{K}/T$, which we plot in Figure 3(b). Note that in this finite size regime, exactly as in [15, Figure 1], the resulting rate bound is significantly smaller than the capacity $C = 1 - h(p) = 0.500$, which we only approach asymptotically.

Remark V.5. *As in Remark V.2, we can give a Gaussian approximation in the setting of Example V.4 and deduce a capacity result. Using* (38) *we deduce from 25 that d^* satisfies*

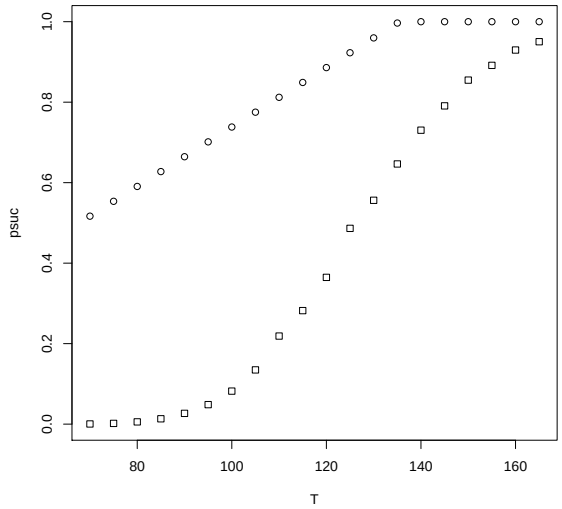
$$\frac{1}{M} \simeq \mathbb{P}(\text{Bin}(T,1/2) \leq d^*) \simeq 2^{-T+Th(d^*/T)},$$

or $h(d^*/T) = 1 - \log_2 M/T + o(1)$. Hence for a sequence of problems where the ith problem has $\mathbf{U}^{(i)}$ uniformly distributed on a set of size $M(i)$ we know that

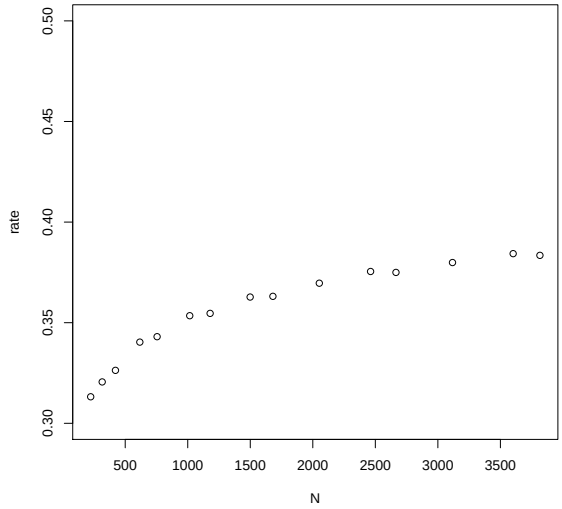
$$\liminf_{i\to\infty} \frac{\log_2 M(i)}{T(i)} \geq C + \epsilon := 1 - h(p) + \epsilon,$$

so $d^*/T < p - \delta$ and we can deduce by Theorem D.1 that

$$\mathbb{P}(\text{suc}) \simeq \mathbb{P}(\text{Bin}(T,p) \leq d^*) \leq 2^{-TD(p-\epsilon\|p)},$$



(a)



(b)

Fig. 3. Combinatorial group testing with $N = 500$ and $K = 10$, where the output $\mathbf{X}$ of standard noiseless group testing is fed into a memoryless binary symmetric channel with $p = 0.11$. (a) We vary the number of tests T between 70 and 165, and plot the success probability on the y axis. We plot the upper bound on $\mathbb{P}(\mathrm{suc})$ given by Example V.4 using $\square$. For comparison, we plot the (weaker) Fano bound (27) taken from [6] as $\circ$. (b) In each case we choose T large enough such that the success probability $\mathbb{P}(\mathrm{suc}) = 0.999$. We plot the upper bound on the rate given by Example V.4, and observe that this is significantly lower than the value of the capacity $C = 0.500$ in this finite blocklength regime.

so tends to zero exponentially fast, meaning that the strong capacity must be less than $1 - h(p)$.

Note the similarity between the calculations in Examples V.1 and V.4; in the former case we control source probabilities (see also [15, Theorem 35]) in the latter case we control concentration of channel probabilities, . This fits with the idea that we analyse group testing as a joint source-channel coding problem.

APPENDIX A
PROOF OF NON-ADAPTIVE RESULTS

We use an argument based on [15], adapted to the scenario where $\mathbf{U}$ need not be uniform. Note that while this case is considered in [16], that paper uses a different (list-decoding) rule; in effect we combine the hypothesis testing rule of [15] and the model of [16]. We use the list-decoding rule of [35] in Appendix B.

Consider a hypothesis testing problem where we are given a pair $(\mathbf{k}, \mathbf{y})$ and asked to test the null hypothesis that it comes from joint distribution $P_{\mathbf{KY}}$ against an alternative of some other specific $Q_{\mathbf{KY}}$. This is a counterfactual exercise; in group testing we do not know $\mathbf{K}$, however, it is helpful to imagine a separate user who is asked to make inference using this information, and uses the following hypothesis testing rule:

> given pair $(\mathbf{k}, \mathbf{y})$ send $\mathbf{y}$ to the decoder to produce $\mathbf{z}$, and then accept $P_{\mathbf{KY}}$ with probability
>
> $$P_{\mathbf{U}|\mathbf{K}}(\mathbf{z}|\mathbf{k}) = \frac{P_{\mathbf{U}}(\mathbf{z})\, P_{\mathbf{K}|\mathbf{U}}(\mathbf{k}|\mathbf{z})}{P_{\mathbf{K}}(\mathbf{k})}. \qquad (28)$$

Proof of Theorem III.2. The key is to notice that $\mathbf{U} \to \mathbf{K} \to \mathbf{Y} \to \mathbf{Z}$ form a Markov chain, so for estimation algorithm $P_{\mathbf{Z}|\mathbf{Y}}$ we obtain

$$P_{\mathbf{Z}|\mathbf{U}}(\mathbf{w}|\mathbf{z}) \;=\; \sum_{\mathbf{k},\mathbf{y}} P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{w}|\mathbf{y})\, P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k})\, P_{\mathbf{K}|\mathbf{U}}(\mathbf{k}|\mathbf{z}).$$

Using this, there is an equivalence betwen group testing error probability and $\mathbb{P}(\text{Type I error})$ since

$$\begin{aligned}
&\mathbb{P}(\mathrm{suc}) \\
&= \sum_{\mathbf{z},\mathbf{w}} P_{\mathbf{U}}(\mathbf{z})\, \mathbb{I}(\mathbf{w} = \mathbf{z}) P_{\mathbf{Z}|\mathbf{U}}(\mathbf{w}|\mathbf{z}) \\
&= \sum_{\mathbf{z},\mathbf{w}} P_{\mathbf{U}}(\mathbf{z})\, \mathbb{I}(\mathbf{w} = \mathbf{z}) \\
&\qquad \times \left[\sum_{\mathbf{k},\mathbf{y}} P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{w}|\mathbf{y})\, P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k})\, P_{\mathbf{K}|\mathbf{U}}(\mathbf{k}|\mathbf{z}) \right] \\
&= \sum_{\mathbf{k},\mathbf{y},\mathbf{z}} P_{\mathbf{KY}}(\mathbf{k},\mathbf{y})\, P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})\, \frac{P_{\mathbf{U}}(\mathbf{z})\, P_{\mathbf{K}|\mathbf{U}}(\mathbf{k}|\mathbf{z})}{P_{\mathbf{K}}(\mathbf{k})} \qquad (29) \\
&= \sum_{\mathbf{k},\mathbf{y}} P_{\mathbf{KY}}(\mathbf{k},\mathbf{y}) \sum_{\mathbf{z}} P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})\, P_{\mathbf{U}|\mathbf{K}}(\mathbf{z}|\mathbf{k}) \\
&= \sum_{\mathbf{k},\mathbf{y}} P_{\mathbf{KY}}(\mathbf{k},\mathbf{y})\, \mathbb{P}(\text{accept } P_{\mathbf{KY}} \text{ given pair } (\mathbf{k},\mathbf{y})) \\
&= 1 - \mathbb{P}(\text{ Type I error })
\end{aligned}$$

where we use the expression (28) to deal with (29). We find the probability of a Type II error in the same way. We focus on the case where $Q_{\mathbf{KY}} = P_{\mathbf{K}} \times Q_{\mathbf{Y}}$ (so $\mathbf{K}$ and $\mathbf{Y}$ are independent under $Q_{\mathbf{KY}}$), where

$$\begin{aligned}
&\mathbb{P}(\text{ Type II error }) \\
&= \sum_{\mathbf{k},\mathbf{y}} Q_{\mathbf{KY}}(\mathbf{k},\mathbf{y})\, \mathbb{P}(\text{accept } P_{\mathbf{KY}} \text{ given pair } (\mathbf{k},\mathbf{y})) \\
&= \sum_{\mathbf{k},\mathbf{y}} P_{\mathbf{K}}(\mathbf{k})\, Q_{\mathbf{Y}}(\mathbf{y}) \sum_{\mathbf{z}} P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})\, \frac{P_{\mathbf{U}}(\mathbf{z})\, P_{\mathbf{K}|\mathbf{U}}(\mathbf{k}|\mathbf{z})}{P_{\mathbf{K}}(\mathbf{k})}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\mathbf{k},\mathbf{z}} P_{\mathbf{U}}(\mathbf{z}) P_{\mathbf{K}|\mathbf{U}}(\mathbf{k}|\mathbf{z}) \sum_{\mathbf{y}} Q_{\mathbf{Y}}(\mathbf{y}) P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y}) \\
&= \sum_{\mathbf{k},\mathbf{z}} P_{\mathbf{U}}(\mathbf{z}) P_{\mathbf{K}|\mathbf{U}}(\mathbf{k}|\mathbf{z}) Q^*(\mathbf{z}) \\
&= \sum_{\mathbf{z}} P_{\mathbf{U}}(\mathbf{z}) Q^*(\mathbf{z}) \qquad (30)
\end{aligned}
$$

Hence, since $\beta_{1-\epsilon}$ gives the minimum type II error, we deduce the Proposition. □

Proof of Corollary III.3. Taking $Q_{\mathbf{Y}} \equiv 1/2^T$, the optimal rule is to accept $P_{\mathbf{KY}}$ with probability $1-\epsilon$ if $\mathbf{x} = \mathbf{y}$, and to reject $P_{\mathbf{KY}}$ otherwise (this corresponds to taking $\lambda = 1-\epsilon$ and $d^* = 0$ in Example V.4 above). We obtain by Theorem III.2 that

$$
\frac{(1-\epsilon)}{2^T} = \beta_{1-\epsilon}(P_{\mathbf{KY}}, P_{\mathbf{K}} \times Q_{\mathbf{Y}}) \leq \sum_{\mathbf{z}\in\{0,1\}^N} P_{\mathbf{U}}(\mathbf{z}) Q^*(\mathbf{z}). \qquad (31)
$$

For each defective set $\mathbf{U}$, we write $\mathbf{X} = \mathbf{Y} = \boldsymbol{\theta}(\mathbf{U})$. For a particular $\mathbf{y}$, we write $A(\mathbf{y}) = \boldsymbol{\theta}^{-1}(\mathbf{y}) = \{\mathbf{z} : \boldsymbol{\theta}(\mathbf{z}) = \mathbf{y}\}$ for the defective sets that get mapped to $\mathbf{y}$ by the testing procedure. We write $p_{\max}(\mathbf{y}) = \max_{\mathbf{z}\in A(\mathbf{y})} P_{\mathbf{U}}(\mathbf{z})$ for the maximum probability in $A(\mathbf{y})$ and $\mathcal{U}^*(\mathbf{y}) = \{\mathbf{u} : P_{\mathbf{U}}(\mathbf{u}) = p_{\max}(\mathbf{y})\}$ for the collection of defective sets achieving this probability. For each $\mathbf{y}$, pick a string $\mathbf{u}^*(\mathbf{y}) \in \mathcal{U}^*(\mathbf{y})$ in any arbitrary fashion; and note that there are up to 2^T strings $\mathbf{u}^*(\mathbf{y})$, which are distinct, since they each map to a different value under θ. These various definitions are illustrated in Figure 4. Using (31) we deduce:

$$
\begin{aligned}
\mathbb{P}(\text{suc}) &= (1-\epsilon) \\
&\leq 2^T \sum_{\mathbf{z}\in\{0,1\}^N} P_{\mathbf{U}}(\mathbf{z}) Q^*(\mathbf{z}) \\
&\leq \sum_{\mathbf{y}\in\{0,1\}^T} \sum_{\mathbf{z}\in\{0,1\}^N} P_{\mathbf{U}}(\mathbf{z}) P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y}) \\
&\leq \sum_{\mathbf{y}\in\{0,1\}^T} \sum_{\mathbf{z}\in A(\mathbf{y})} p_{\max}(\mathbf{y}) P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{u}|\mathbf{y}) \qquad (32) \\
&\leq \sum_{\mathbf{y}\in\{0,1\}^T} p_{\max}(\mathbf{y}) \\
&= \sum_{\mathbf{y}\in\{0,1\}^T} P_{\mathbf{U}}(\mathbf{u}^*(\mathbf{y})) \\
&\leq \Pi_{\mathbf{U}}(2^T).
\end{aligned}
$$

Here (32) follows since for given $\mathbf{y}$ the success probability is maximised by restricting to $P_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y})$ supported on the set $\mathbf{z} \in A(\mathbf{y})$, so we know that $P_{\mathbf{U}}(\mathbf{z}) \leq p_{\max}(\mathbf{y})$. The result follows since there are at most 2^T separate messages $\mathbf{X} = \mathbf{x}$, so at most 2^T distinct values $\mathbf{u}^*(\mathbf{x})$. This result generalizes (14). □

Note that (as expected) the success probability is maximised by the maximum likelihood decoder $P_{\mathbf{Z}|\mathbf{Y}}$ which places all its support on members of $\mathbf{U}^*(\mathbf{y})$.

APPENDIX B
PROOF OF ADAPTIVE RESULT, THEOREM IV.2

Proof of Theorem IV.2. We use the machinery of Kostina and Verdú [16]. That is, given $(\mathbf{U}, \mathsf{X}, \mathbf{Y})$ we perform a hypothesis test between

$$
\begin{aligned}
H_0: &\ P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k}) \\
H_1: &\ Q_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) Q_{\mathbf{Y}}(\mathbf{y}),
\end{aligned}
$$

for some probability mass functions $Q_{\mathbf{U}}$, $Q_{\mathbf{Y}}$. For estimation algorithm $P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}$, recall that we write $S^*_{\mathcal{X}}(\mathbf{y})$ for the set of values that $\mathbf{Z}$ may be decoded to; that is $\mathbf{u} \in S^*_{\mathcal{X}}(\mathbf{y})$ if and only if $P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{y},\mathcal{X}) > 0$. and use the (sub-optimal) decision rule that we choose H_0 if $\mathbf{U} \in S^*_{\mathcal{X}}(\mathbf{Y})$.

We write $\mathbf{u}$, $\mathcal{X}$ and $\mathbf{y}$ as indices of summation for brevity, to refer to sums over $\mathbf{u} \in \{0,1\}^N$, $\mathcal{X} \in \{0,1\}^{N\times T}$ and $\mathbf{y} \in \{0,1\}^T$. Using the fact that for the estimation algorithm $P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}$

$$
\begin{aligned}
&\mathbb{P}(\text{suc}|\mathbf{U}=\mathbf{u}, \mathsf{X}=\mathcal{X}, \mathbf{Y}=\mathbf{y}) \\
&\quad = \sum_{\mathbf{z}} P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{z}|\mathbf{y},\mathcal{X}) \mathbb{I}(\mathbf{z}=\mathbf{u}) = P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{y},\mathcal{X}),
\end{aligned}
$$

as in the proof of Theorem III.2 there is a relationship between between group testing error probability and $\mathbb{P}(\text{Type I error})$ since:

$$
\begin{aligned}
&1 - \mathbb{P}(\text{Type I error}) \\
&= \sum_{\mathbf{u},\mathcal{X},\mathbf{y}} P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k}) \mathbb{I}(\mathbf{u} \in S^*_{\mathcal{X}}(\mathbf{y})) \\
&= \sum_{\mathbf{u},\mathcal{X},\mathbf{y}} P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k}) \\
&\qquad \times \sum_{\mathbf{z}} \mathbb{I}(\mathbf{u} \in S^*_{\mathcal{X}}(\mathbf{y})) P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{z}|\mathbf{y},\mathcal{X}) \qquad (33) \\
&\geq \sum_{\mathbf{u},\mathcal{X},\mathbf{y}} P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k}) \\
&\qquad \times P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{y},\mathcal{X}) \qquad (34) \\
&= \sum_{\mathbf{u},\mathcal{X},\mathbf{y}} P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k}) \\
&\qquad \times \mathbb{P}(\text{suc}|\mathbf{U}=\mathbf{u}, \mathsf{X}=\mathcal{X}, \mathbf{Y}=\mathbf{y}),
\end{aligned}
$$

which we recognise as $\mathbb{P}(\text{suc})$, where the result follows since the inner sum in (33) includes the term $P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{y},\mathcal{X})$. Hence, if $\mathbb{P}(\text{suc}) \geq 1-\epsilon$ then $\mathbb{P}(\text{Type I error}) \leq \epsilon$, so the type II error probability of this decision rule satisfies

$$
\beta_{1-\epsilon}(P_{\mathbf{U}} P_{\mathsf{X}\|\mathbf{Y}^-} P_{\mathbf{Y}|\mathbf{K}}, Q_{\mathbf{U}} P_{\mathsf{X}\|\mathbf{Y}^-} Q_{\mathbf{Y}}) \leq \mathbb{P}(\text{Type II error}).
$$

We can write this type II error probability using a similar argument (based on [16, Eq. (60-63)]) as

$$
\begin{aligned}
&\mathbb{P}(\text{Type II error}) \\
&= \sum_{\mathbf{u},\mathcal{X},\mathbf{y}} Q_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) Q_{\mathbf{Y}}(\mathbf{y}) \mathbb{I}(\mathbf{u} \in S^*_{\mathcal{X}}(\mathbf{y})) \\
&= \sum_{\mathcal{X},\mathbf{y}} P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) Q_{\mathbf{Y}}(\mathbf{y}) \sum_{\mathbf{u}} Q_{\mathbf{U}}(\mathbf{u}) \mathbb{I}(\mathbf{u} \in S^*_{\mathcal{X}}(\mathbf{y})) \\
&\leq \sum_{\mathcal{X},\mathbf{y}} P_{\mathsf{X}\|\mathbf{Y}^-}(\mathcal{X}\|\mathbf{y}^-) Q_{\mathbf{Y}}(\mathbf{y}) L \\
&= L,
\end{aligned}
$$

where the final line follows from (17). □

APPENDIX C
PROOF OF THEOREM IV.3

Proof of Theorem IV.3. In general, in the noiseless case, for each defective set $\mathbf{U}$, we write $\mathbf{X} = \mathbf{Y} = \boldsymbol{\theta}(\mathbf{U}, \mathcal{X})$. For a particular $\mathbf{Y} = \mathbf{y}$ and $\mathsf{X} = \mathcal{X}$, we write $A(\mathbf{y}, \mathcal{X}) = \boldsymbol{\theta}^{-1}(\mathbf{y}, \mathcal{X}) = \{\mathbf{z} : \boldsymbol{\theta}(\mathbf{z}, \mathcal{X}) = \mathbf{y}\}$ for the defective sets that get mapped to $\mathbf{y}$ by the testing procedure defined by $\mathcal{X}$. We write $p_{\max}(\mathbf{y}, \mathcal{X}) = \max_{\mathbf{z} \in A(\mathbf{y}, \mathcal{X})} P_{\mathbf{U}}(\mathbf{z})$ for the maximum probability in $A(\mathbf{y}, \mathcal{X})$ and $\mathcal{U}^*(\mathbf{y}, \mathcal{X}) = \{\mathbf{u} : P_{\mathbf{U}}(\mathbf{u}) = p_{\max}(\mathbf{y}, \mathcal{X})\}$ for the collection of defective sets achieving this probability. For each $\mathbf{y}$, pick a string $\mathbf{u}^*(\mathbf{y}, \mathcal{X}) \in \mathcal{U}^*(\mathbf{y}, \mathcal{X})$ in any arbitrary fashion; and note that there are up to 2^T strings $\mathbf{u}^*(\mathbf{y}, \mathcal{X})$, which are distinct, since they each map to a different value under $\boldsymbol{\theta}(\cdot, \mathcal{X})$. These definitions are illustrated in Figure 4.

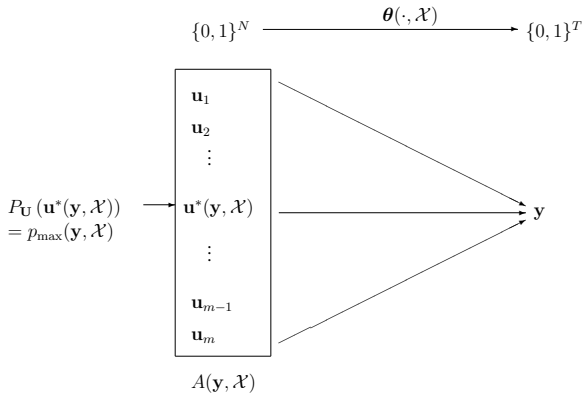


Fig. 4. Schematic illustration of the sets used to prove Theorem IV.3.

Since the channel is noiseless, $\mathbf{u} \in A(\mathbf{y}, \mathcal{X})$, so using the fact that for any $\mathbf{u}$ the $1 = \sum_{\mathbf{w}} \mathbb{I}(\mathbf{u} \in A(\mathbf{w}, \mathcal{X}))$, we deduce from (34) that $\mathbb{P}(\mathrm{suc})$ is

$$\begin{aligned}
&\sum_{\mathbf{u}, \mathcal{X}, \mathbf{y}} P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X} \| \mathbf{Y}^-}\left(\mathcal{X} \| \mathbf{y}^-\right) P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k}) P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{y}, \mathcal{X}) \\
&= \sum_{\mathbf{u}, \mathcal{X}, \mathbf{y}} P_{\mathbf{U}}(\mathbf{u}) P_{\mathsf{X} \| \mathbf{Y}^-}\left(\mathcal{X} \| \mathbf{y}^-\right) P_{\mathbf{Y}|\mathbf{K}}(\mathbf{y}|\mathbf{k}) \\
&\qquad \times P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{y}, \mathcal{X}) \sum_{\mathbf{w}} \mathbb{I}(\mathbf{u} \in A(\mathbf{w}, \mathcal{X})) \\
&= \sum_{\mathbf{w}, \mathcal{X}} P_{\mathsf{X} \| \mathbf{Y}^-}\left(\mathcal{X} \| \mathbf{w}^-\right) \sum_{\mathbf{u}} P_{\mathbf{U}}(\mathbf{u}) \\
&\qquad \times P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{w}, \mathcal{X}) \mathbb{I}(\mathbf{u} \in A(\mathbf{w}, \mathcal{X})) \\
&\leq \sum_{\mathbf{w}, \mathcal{X}} P_{\mathsf{X} \| \mathbf{Y}^-}\left(\mathcal{X} \| \mathbf{w}^-\right) p_{\max}(\mathbf{w}) \qquad (35) \\
&\leq \sum_{\mathbf{w}} p_{\max}(\mathbf{w}) \sum_{\mathcal{X}} P_{\mathsf{X} \| \mathbf{Y}^-}\left(\mathcal{X} \| \mathbf{w}^-\right) \\
&= \sum_{\mathbf{w}} p_{\max}(\mathbf{w}) \qquad (36) \\
&\leq \Pi_{\mathbf{U}}(2^T), \qquad (37)
\end{aligned}$$

where (35) follows because $P_{\mathbf{U}}(\mathbf{u}) \leq p_{\max}(\mathbf{w})$ on this set and since $\sum_{\mathbf{u}} P_{\mathbf{Z}|\mathbf{Y},\mathsf{X}}(\mathbf{u}|\mathbf{w}, \mathcal{X}) = 1$, (36) follows by (17). Finally (37) follows since there are at most 2^T separate messages $\mathbf{w}$, so at most 2^T distinct values $\mathbf{u}^*(\mathbf{w})$. □

APPENDIX D
CONCENTRATION INEQUALITY

We require an exponential bound in terms of relative entropy. There is a wide literature on this subject, and we take a one-sided form of the Chernoff bound stated as [42, Theorem 5] (for $p \leq 1/2$, we take $d = (1-p)$ and $\sigma^2 = p(1-p)$ in the result stated there):

Theorem D.1. *For $q < p \leq 1/2$, we bound the probability*

$$\mathbb{P}\left(\mathrm{Bin}(n, p) \leq nq\right) \leq 2^{-nD(q\|p)},$$

where we write $D(q\|p)$ for the relative entropy from a Bernoulli(q) random variable to a Bernoulli(p), calculated using logarithms to base 2.

Since this is generally a tight bound, we use it to motivate the following approximation, which comes from writing $D(q\|1/2) = \log 2 - h(q)$. For any L we deduce that

$$\mathbb{P}(\mathrm{Bin}(N, 1/2) \leq L) \simeq 2^{-ND(L/N\|1/2)} = 2^{-N} 2^{Nh(L/N)}. \qquad (38)$$

Here and throughout the paper, $\simeq$ refers to equality on an exponential scale (the logarithms of both sides are approximately equal). If we take $L = L(y) := Np + y\sqrt{Np(1-p)}$ and $T(y) = Nh(L(y)/N)$ we deduce that

$$\mathbb{P}(\mathrm{Bin}(N, 1/2) \leq L(y)) \simeq 2^{-N+T(y)}. \qquad (39)$$

ACKNOWLEDGMENT

The author thanks Matthew Aldridge, Leonardo Baldassini and Thomas Kealy for useful discussions regarding the group testing problem, and Vanessa Didelez for help in understanding causal conditional probability. The author would like to thank two anonymous referees and the Associate Editor for their thorough and careful reading of the paper and very helpful suggestions.

REFERENCES

[1] R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, pp. 436–440, 1943.

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

[3] 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.

[4] M. P. Aldridge, L. Baldassini, and O. T. Johnson, "Group testing algorithms: bounds and simulations," *IEEE Trans. Inform. Theory*, vol. 60, no. 6, pp. 3671–3687, 2014.

[5] G. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1880 –1901, March 2012.

[6] 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 *Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2011, pp. 1832 –1839.

[7] T. Kealy, O. T. Johnson, and R. Piechocki, "The capacity of non–identical adaptive group testing," in *Proceedings of the 52nd Annual Allerton Conference on Communication, Control and Computing*, 2014, pp. 101–108.

[8] T. Li, C. L. Chan, W. Huang, T. Kaced, and S. Jaggi, "Group testing with prior statistics," in *Proceedings of the 2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 2346–2350.

[9] M. P. Aldridge, "Adaptive group testing as channel coding with feedback," in *Proceedings of the 2012 IEEE International Symposium on Information Theory*, July 2012, pp. 1832 –1836.

[10] ——, "Interference mitigation in large random wireless networks," Ph.D. dissertation, Science Faculty, University of Bristol, 2011, arxiv:1109.1255.

[11] H. Marko, "The bidirectional communication theory–a generalization of information theory," *IEEE Transactions on Communications*, vol. 21, no. 12, pp. 1345–1351, 1973.

[12] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*. Citeseer, 1990, pp. 303–305.

[13] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology, Zürich, 1998.

[14] P.-O. Amblard and O. J. J. Michel, "The relation between Granger Causality and directed information theory: A review," *Entropy*, vol. 15, no. 1, p. 113, 2012. [Online]. Available: http://www.mdpi.com/1099-4300/15/1/113

[15] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[16] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, 2013.

[17] C. Aksoylar, G. Atia, and V. Saligrama, "Sparse signal processing with linear and non-linear observations: A unified Shannon theoretic approach," in *Proceedings of the 2013 IEEE Information Theory Workshop*, Sept 2013, pp. 1–5.

[18] G. Atia, S. Aeron, E. Ermis, and V. Saligrama, "On throughput maximization and interference avoidance in cognitive radios," in *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*. IEEE, 2008, pp. 963–967.

[19] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing," in *Proceedings of the 2010 IEEE International Symposium on Information Theory*. IEEE, 2010, pp. 1913–1917.

[20] Y. Erlich, A. Gordon, M. Brand, G. Hannon, and P. Mitra, "Compressed genotyping," *IEEE Trans. Inform. Theory*, vol. 56, no. 2, pp. 706–723, 2010.

[21] N. Shental, A. Amir, and O. Zuk, "Identification of rare alleles and their carriers using compressed se (que) nsing," *Nucleic acids research*, vol. 38, no. 19, pp. e179–e179, 2010.

[22] D. Balding, W. Bruno, D. Torney, and E. Knill, "A comparative survey of non-adaptive pooling designs," in *Genetic mapping and DNA sequencing*. Springer, 1996, pp. 133–154.

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

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

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

[26] M. Malyutov, "Search for sparse active inputs: a review," in *Information Theory, Combinatorics and Search Theory*, ser. Lecture notes in Computer Science. London: Springer, 2013, vol. 7777, pp. 609–647.

[27] ——, "Recovery of sparse active inputs in general systems: a review," in *Computational Technologies in Electrical and Electronics Engineering (SIBIRCON), 2010 IEEE Region 8 International Conference on*. IEEE, 2010, pp. 15–22.

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

[29] T. Wadayama, "An analysis on non-adaptive group testing based on sparse pooling graphs," in *Proceedings of the 2013 IEEE International Symposium on Information Theory*, 2013, pp. 2681–2685.

[30] T. Wadayama, T. Izumi, and K. Mimura, "Bitwise MAP estimation for group testing based on holographic transformation," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 2787–2791.

[31] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1753–1768, 2003.

[32] V. Tan and G. Atia, "Strong impossibility results for sparse signal processing," *IEEE Signal Processing Letters*, vol. 21, no. 3, pp. 260–264, March 2014.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley, 1991.

[34] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge: Cambridge University Press, 2011, 2nd Edition.

[35] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4316–4330, 2015.

[36] S. Verdú and T. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[37] M. Hayashi, "Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding," *Physical Review A*, vol. 76, no. 6, p. 062301, 2007.

[38] H. Nagaoka, "Strong converse theorems in quantum information theory," in *Proceedings of ERATO Workshop on Quantum Information Science 2001, Univ. Tokyo, Tokyo, Japan, September 68, 2001, Asymptotic Theory in Quantum Statistical Inference*, M. Hayashi, Ed. World Scientific, 2005.

[39] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 4903–4925, Aug 2011.

[40] J. Wolfowitz, *Coding theorems of information theory*, ser. Ergebnisse der Mathematik und Ihrer Grenzgebiete. Springer, 1961, no. 31.

[41] R. G. Gallager, *Information theory and reliable communication*. Springer, 1968, vol. 2.

[42] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications and coding," *Foundations and Trends in Communications and Information Theory*, vol. 10, no. 1–2, pp. 1–246, 2013.

[43] L. Baldassini, "Rates and algorithms for group testing," Ph.D. dissertation, Science Faculty, University of Bristol, 2015.

Oliver Johnson received the B.A. degree in 1995, Part III Mathematics in 1996, and the Ph.D. degree in 2000, all from the University of Cambridge, Cambridge, U.K. He was Clayton Research Fellow at Christ's College and Max Newman Research Fellow at Cambridge University until 2006, during which time he published the book *Information Theory and the Central Limit Theorem (Singapore: World Scientific, 2004)*. Since 2006, he has been at University of Bristol , Bristol, U.K, and is now Reader in Information Theory. Much of his research has recently focused on the entropy of discrete random variables; including using transportation of measure to prove the Shepp-Olkin conjecture, log-Sobolev inequalities, maximum entropy, monotonicity and other problems. He has also recently published papers concerning wireless communication schemes, ecological modelling and change-point estimation.