# Fundamental Limits of Non-Adaptive Group Testing With Markovian Correlation

**Aditya Narayan Ravi** **Ilan Shomorony**
Electrical and Computer Engineering Department, University of Illinois Urbana-Champaign

## Abstract

We study a correlated group testing model where $n$ items are infected according to a Markov chain, which creates bursty infection patterns. In the sparse infections regime, where the expected number of infections scales as $O(n^\theta)$ with $\theta \in (0,1)$, we propose a non-adaptive testing strategy with an efficient decoding algorithm. Our approach outperforms an optimal yet computationally inefficient independent testing and decoding scheme (one that disregards correlation), under certain parameter regimes. At a high level, we use randomized block testing, where we first sample contiguous blocks of correlated items and then subsample items within selected blocks. Decoding then proceeds in two stages: a coarse elimination step to rule out items appearing in negative tests, followed by a fine thresholding step that declares an item infected if its test participation count exceeds a predefined threshold. Notably, when $\theta \to 0$, our method achieves asymptotically vanishing error while using a number of tests that is within a $1/\ln(2) \approx 1.44$ multiplicative factor of the fundamental entropy bound—a result that parallels the independent group testing setting. Further, we show that the number of tests reduces with an increase in the expected burst length of infected items, quantifying the advantage of exploiting correlation in test design.

## 1 INTRODUCTION

Consider a *correlated group testing* problem where the infection status of an item is correlated with the infection status of the item next to it. We are interested in non-adaptive group testing strategies (Balding et al., 1996) to determine the set of infected items.

Classical group testing, introduced by Dorfman (1943), often employs the combinatorial model, a widely used framework. In this model, a population of size $n$ contains exactly $k$ infected individuals, chosen uniformly at random from all possible subsets of size $k$. An alternative approach is the probabilistic i.i.d. model, which assumes that each individual is independently infected with probability $q_n$. Importantly, these two models are largely equivalent (Aldridge et al., 2019). This framework has been extensively explored in the literature (Coja-Oghlan et al., 2020a; Scarlett and Cevher, 2017, 2018; Coja-Oghlan et al., 2020b; Aldridge et al., 2019). In this setting, when infections are sparse, group testing can dramatically reduce the number of tests required.

However, the assumption of independence is rarely reflective of real-world infection dynamics. Recent studies have begun to acknowledge this limitation and have explored the potential of leveraging known community structures (Ahn et al., 2021; Nikolopoulos et al., 2021, 2023; Ahn et al., 2023; Bertolotti and Jadbabaie, 2020; Arasli and Ulukus, 2023) to enhance group testing efficiency. Moreover, in many practical scenarios, infections are not uniformly distributed but tend to occur in bursts (Colbourn, 1999; Li et al., 2023; Bui et al., 2022). For example, people living in the same living space or neighborhood are more likely to get infected and tested together. Such bursty infection patterns are naturally captured by correlated infection models.

Another important motivation for our work arises from imaging-based spatial transcriptomics technologies such as MERFISH (Moffitt and Zhuang, 2016) and Xenium (Janesick et al., 2023). These technologies use barcoded "probes" to simultaneously test the expression of a group of genes. If any of the genes tested are being expressed in a certain location on a tissue, that location lights up when an image is captured under a fluorescence microscope. The design of these probes can be cast as a group testing problem,

Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Mai Khao, Thailand. PMLR: Volume 258. 

where we can target multiple genes during each testing/imaging round. Furthermore, the genes that are tested often exhibit strong local dependencies, where genes involved in the same regulatory pathways or located near each other on a chromosome tend to activate together (Allocco et al., 2004), once again resulting in bursty patterns. This "co-regulation" has been exploited previously in Cleary et al. (2021), with correlated genes being measured together to reduce the number of tests required. Understanding correlated group testing can thus provide insights into how an optimal design of these probes could take advantage of gene co-regulation, reducing the number of testing rounds. Note that once the probes are designed and introduced in spatial transcriptomics experiments, modifying or adapting the experimental setup is infeasible. This further motivates the need for non-adaptive group testing strategies for this setting. By non-adaptive, we refer to test designs that are fully specified in advance and do not depend on the outcomes of earlier tests.

Delving into the problem more formally, consider a set of $n$ distinct items labeled $\{1, 2, \ldots, n\}$. Each item $i$ is assigned a binary random variable $U_i \in \{0, 1\}$, where $U_i = 1$ indicates that item $i$ is infected, and $U_i = 0$ otherwise. The collective infection status of all items is represented by the *infection vector* $U^n = (U_1, U_2, \ldots, U_n)$. The objective is to estimate $U^n$ using the outcomes of $T$ non-adaptive group tests. Each test simultaneously tests a subset of items. A test returns a positive result if at least one infected item is included in the group; otherwise, the result is negative.

To build intuition, consider the classical independent group testing problem, where each item in a population of size $n$ is independently infected with probability $q_n$. Consider a sparse regime, where $U_i$'s are independent and distributed as Bernoulli($q_n$), where $nq_n = kn^\theta$ for some constant $k > 0$. As $\theta \to 0$ (referred to as the *very sparse* regime (Aldridge et al., 2019)), it is well known (Aldridge et al., 2019; Coja-Oghlan et al., 2020a; Aldridge, 2017) that a vanishing probability of error can be achieved as $n \to \infty$, if the number of tests $T(n)$ satisfies (we define $\log(\cdot) := \log_2(\cdot)$)

$$T(n) > nq_n \log n := T_{\text{ind}}. \tag{1}$$

Importantly, the above bound is also optimal. It is established that no algorithm can further reduce the number of required tests while maintaining asymptotically vanishing error probability (Aldridge et al., 2019).

One can extend this approach to correlated group testing, where the infection vector $U^n$ exhibits known dependencies. A straightforward application of the previous scheme would disregard these correlations, effectively assuming an i.i.d. prior on $U^n$ and yielding the same number of tests. This raises a key question: can incorporating correlations into test design further reduce the number of required tests in (1)?

Prior work suggests that incorporating correlations into test design and decoding rule can reduce the number of required tests under some assumptions. Independent testing strategies become suboptimal in correlated settings, as they fail to leverage the information an item's infection status provides about others. Adaptive strategies (Ahn et al., 2021, 2023; Nikolopoulos et al., 2021, 2023) have successfully exploited these dependencies to reduce the number of tests. However, achieving similar provable gains with non-adaptive strategies remains a challenge.

For instance, Nikolopoulos et al. (2023) propose a non-adaptive approach that reduces tests, albeit under a non-vanishing false positive rate. Similarly, Portnoy and Cohen (2024) introduce a trellis-based test design that demonstrates test reductions empirically. These results strongly suggest that non-adaptive strategies should, in principle, benefit from correlation-aware designs. However, designing tests that provably reduce the number of tests under vanishing error remains challenging.

To make progress in the non-adaptive correlated setting, we consider the case where the infection vector $U^n$ is generated by a two-state Markov chain, initialized according to its stationary distribution, shown in Figure 1. For relevant parameter choices, this Markov chain induces bursty infection patterns. Specifically, define transition probabilities $\alpha_n = k' n^\theta / n$ and $\beta < 1$ (a constant). The parameter $\beta$ is the probability of the infection process leaving a burst of infected items and the expected burst length is $1/\beta$. The marginal infection probability remains $q_n = \frac{kn^\theta}{n}$, consistent with the sparse regime.

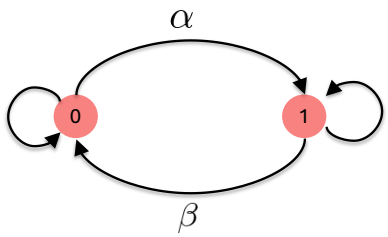


Figure 1: The 2-state Markov chain that generates the infection vector $U^n$

In this Markovian setting, we design a non-adaptive *per-item* randomized testing and polynomial-time decoding algorithm that explicitly leverages the correlation structure to minimize the number of tests. We

focus on the sparse regime where $nq_n = kn^\theta$ for some constant $k > 0$. Theorem 3 establishes a general achievability bound that holds for all $\theta \in (0, 1)$. As a concrete consequence, when $\theta \to 0$, one can make an explicit parameter choice leading to a vanishing error probability as $n \to \infty$, provided the number of tests satisfies

$$T(n) > \frac{\beta}{\ln 2} nq_n \log n := T_{\text{corr}}, \tag{2}$$

where $\beta$ denotes the transition probability from an infected item to a non-infected item.

While the above expression is a relaxed version of our more complicated full theorem which considers all $\theta \in (0, 1)$, it highlights an important aspect of correlation. Specificially, as $\theta \to 0$, whenever

$$\beta < \ln 2 \approx 0.69$$

this strictly improves upon the independent case (1).

Furthermore, in the Markovian setting, a fundamental general result in group testing (Wolf, 1985) implies that for any group testing strategy and any decoding strategy achieving a vanishing probability of error, as $\theta \to 0$ the number of tests $T$ must satisfy the following fundamental lower bound

$$T \geq H(U^n) = \beta nq_n \log n + o(nq_n \log n) := T_{\text{low}}, \tag{3}$$

where $H(\cdot)$ is the entropy. Hence, our achieved $T_{\text{corr}}$ is within a constant of the lower bound $T_{\text{low}}$ in the very sparse regime $\theta \approx 0$. The multiplicative gap $1/\ln 2$ as $\theta \to 0$ is expected, because we consider testing strategies that employ *per-item* decoding (which is sometimes referred to as separate decoding of items (Aldridge et al., 2019)). In this approach, the infection status of each item is determined solely based on the information of which tests contain that item and the test outcomes. In contrast, joint decoding (Aldridge et al., 2019; Scarlett and Cevher, 2017; Coja-Oghlan et al., 2020a) considers the infection status of all items simultaneously, leveraging the entire set of items tested in each test to jointly decode the infection status of the items. Our multiplicative gap of $1/\ln 2$ (as $\theta \to 0$) has an analogue in the classical independent group testing model. Specifically, when the infections are i.i.d. Bernoulli($\theta$), any nonadaptive design constrained to per-item (separate) decoding requires

$$(1/\ln 2)\, nq_n \log(n)$$

tests, whereas optimal joint decoding can achieve

$$nq_n \log(n)$$

tests. This $1/\ln 2$ penalty for per-item decoding was first identified by Malyutov and Mateev (1980) and later revisited by Scarlett and Cevher (2018).

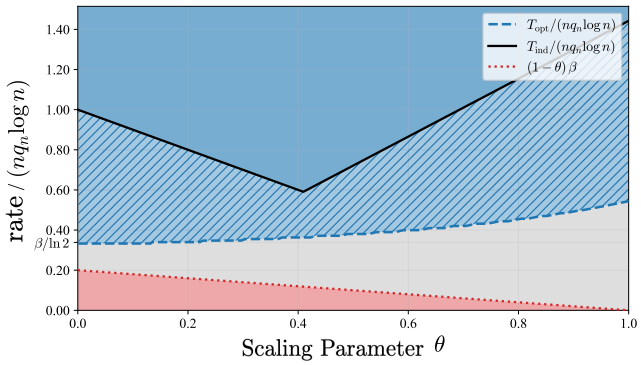


Figure 2: Comparison of the normalized number of tests for an expected burst length $1/\beta = 5$ as $\theta$ varies from 0 to 1. The solid black curve corresponds to the independent-testing baseline, while the dashed blue curve shows the achievability bound from our algorithm. The red region represents an infeasible regime. The gray band between the red line and the achievability curve denotes regimes for which achievability remains unknown. The light blue region between the achievability curve and the independent baseline highlights regimes where our method is feasible but independent testing is strictly suboptimal. Finally, in the darker blue region above the independent curve both strategies succeed.

The key difference from (1) is the influence of the factor $\beta$. Specifically, in the Markovian setting, the number of tests achieved by this algorithm is inversely proportional to the expected burst length ($1/\beta$). Our general result (Theorem 3) shows that for reasonable expected burst lengths (for example $1/\beta > 5$), our algorithm strictly improves over the independent case. Figure 2 shows that with expected burst length $1/\beta = 5$, our algorithm reduces the number of tests significantly.

Our approach utilizes a two-stage strategy to create the (non-adaptive) test matrix: a first stage that selects blocks to be included in each test, and a second stage that randomly samples items from the selected blocks to be included in the tests. Central to this strategy is our efficient, polynomial-time *per-item* decoder, which operates by first performing an initial screening that considers whether a given item is part of a non-infected tested group. If it is not, we follow up by applying a threshold on the number of times that item is tested to estimate its infection status. Surprisingly, this simple decoding rule leverages the testing information and provably reduces the number of tests required in certain parameter regimes. In contrast to earlier per-item decoders for the independent case, which often rely on log-likelihood or information-density statistics, our rule first eliminates any item appearing in a negative test and then applies a scalar threshold

to the number of tests containing that item. While simpler, this rule appears to use essentially the same underlying test information. Notably, the correlation structure is exploited entirely through the correlation-aware test design rather than the decoder itself. Related threshold-style ideas also appear in joint decoding methods such as SCOMP (Aldridge et al., 2019).

Our main contributions are the following: (1) We introduce a novel non-adaptive randomized block testing strategy tailored to the Markovian setting. (2) We introduce a simple *per-item* thresholding-based decoding strategy that is computationally efficient. (3) We prove that our algorithm improves over correlation-unaware testing across a broad range of parameters. (4) We prove that the testing and per-item decoding strategy achieves within a multiplicative factor of $1/\ln 2$ of the information-theoretic lower bound, as $\theta \to 0$.

**Related Work:** Group testing has been extensively studied under uniform or i.i.d. priors (Aldridge et al., 2019; Scarlett and Cevher, 2017, 2018; Colbourn, 1999; Berger and Levenshtein, 2002), as well as in heterogeneous populations with non-identical infection probabilities (Hwang, 1975; Li et al., 2014; Kealy et al., 2014). Several works have extended classical models to incorporate infection bursts (Colbourn, 1999; Li et al., 2023; Lin et al., 2012; Müller and Jimbo, 2004; Bui et al., 2021, 2022), typically by imposing deterministic constraints on factors like burst length and number of bursts, thereby ruling out rare long runs and enabling explicit worst-case test constructions. Our proposed setting can be seen as a probabilistic alternative to these burst models, in which both the burst lengths and the number of bursts are random under a Markov prior. This introduces tail events absent from deterministic formulations, such as unusually long bursts that can cross block boundaries and require separate probabilistic control in the analysis. Related deterministic multi-block models have also been studied in follow-up work (Bui et al., 2022).

Correlations between the infection status of items have also been introduced in the context of community-aware group testing. This setting leverages population structure by assuming the individuals can be partitioned into disjoint families or by employing stochastic block models with correlation-driven infection patterns (Nikolopoulos et al., 2021, 2023; Ahn et al., 2021, 2023). Related approaches, such as those in Bertolotti and Jadbabaie (2020); Arasli and Ulukus (2023); Lau et al. (2022), incorporate subgroup structures, while our approach utilizes a Markovian framework that operates without predefined partitions. Network structure is also present in graph-constrained group testing but, in this setting, the network imposes constraints on the test design (Harvey et al., 2007; Cheraghchi et al., 2012; Karbasi and Zadimoghaddam, 2012).

In Portnoy and Cohen (2024), the authors explore Markovian correlations similar to our setting, and propose a variation of the list Viterbi algorithm, demonstrating empirically a reduction in the number of tests. Our work takes a complementary approach, focusing on testing strategies that allow for rate analysis.

Adaptive strategies for group testing are explored in (Ahn et al., 2021, 2023; Nikolopoulos et al., 2021, 2023). Notably, non-adaptive methods such as the one proposed in Nikolopoulos et al. (2023) demonstrate reductions in the number of required tests, albeit with non-vanishing false positive rates.

## 2 PROBLEM SETTING

We consider a correlated group testing framework where $n$ distinct items, labeled $\{1, 2, \dots, n\}$, are each assigned a binary label $U_i \in \{0, 1\}$. We impose a Markovian structure on $P(U^n)$, specifically a 2-state Markov chain with state space $S = \{0, 1\}$ and transition probabilities $p_{0,1} = \alpha$ and $p_{1,0} = \beta$ as in Figure 1. We generate the infection random vector $U^n$ by initializing the Markov chain in its steady state

$$(q, 1-q) := \left(\frac{\alpha}{\alpha+\beta}, \frac{\beta}{\alpha+\beta}\right),$$

and then recording $n$ consecutive states.

Let $T := T(n)$ be the number of tests to be conducted. For fixed $n$ and $T$, a testing strategy is encoded by a matrix $\mathsf{X} \in \{0, 1\}^{T\times n}$. Specifically, $X_{t,i} = 1$ if item $i$ is included in test $t$, and $X_{t,i} = 0$ otherwise. We observe the test outcomes in a vector $Y = [Y_1, \dots, Y_T]$, where each entry $Y_t$ follows

$$Y_t = \begin{cases} 1 & \text{if } \exists\, i \in \{1, \dots, n\} : U_i = 1 \text{ and } X_{t,i} = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently,

$$Y_t = \mathbf{1}\left\{\sum_{i=1}^{n} U_i X_{t,i} > 0\right\},$$

with $\mathbf{1}\{\cdot\}$ denoting the indicator function. A decoding rule $g$ then maps $(\mathsf{X}, Y)$ to an estimate $\hat{U}^n \in \{0, 1\}^n$. Let $\mathcal{E} = \{U^n \neq \hat{U}^n\}$ denote the error event.

Our goal is to design a sequence of testing matrices and decoding rules for which $\Pr(\mathcal{E}) \to 0$ as $n \to \infty$. As suggested by the i.i.d. result in (1), $T(n)$ will need to scale as $nq_n \log n$. In the Markov setting that we consider

$$q_n = \alpha/(\alpha + \beta),$$

is the stationary probability of an infection. Based on this scaling, we define an achievable achievable number of normalized tests as follows:

**Definition 1** *A achievable number of normalized tests $\tau$ is achievable if there exists a sequence of test matrices $\mathsf{X}_n \in \{0,1\}^{T(n)\times n}$ and corresponding decoding rules $g_n$ such that*

$$\tau = \lim_{n\to\infty} \frac{T(n)}{nq_n \log n}$$

*and $\Pr(\mathcal{E}) \to 0$ as $n \to \infty$. Moreover, we let $\tau^*$ be the infimum of all achievable rates $\tau$ and refer to it as the minimal group achievable number of normalized tests.*

Following the terminology in the group testing literature, we focus on the *sparse regime* (Aldridge et al., 2019), which corresponds to

$$nq_n = O\big(n^\theta\big). \tag{4}$$

To be precise, we consider the case where $\alpha$ is defined as $\alpha = k'n^\theta/n$, for some constant $k'$. We can then compute $q_n$ as $q_n = \alpha/(\alpha+\beta) = kn^\theta/n + o\left(n^{\theta-1}\right)$, where $k := k'/\beta$. For simplicity, we omit lower-order terms of $o\left(n^\theta\right)$ that are added to $q_n$ and assume that $q_n = kn^\theta/n$, in the rest of this work. However, our results remain valid even when the lower-order terms are included.

**Notation:** Throughout the paper, $\log(\cdot)$ represents the logarithm in base 2, while ln represents the logarithm in base $e$. For functions $a(n)$ and $b(n)$, we say $a(n) = o(b(n))$ if $a(n)/b(n) \to 0$ as $n \to \infty$. Similarly, we say $a(n) = O(b(n))$ if there exists a constant $C > 0$ and an integer $n_0$ such that $|a(n)| \le C|b(n)|$ for all $n \ge n_0$.

## 3 MAIN RESULTS

Our main results include a converse bound (adapted from prior work) and an achievability result for our proposed strategy. The following result, adapted from Wolf (1985), establishes a fundamental lower bound on the number of tests required for any group testing strategy with a vanishing error probability.

**Theorem 2 (Converse)** *Under the Markovian correlation model, the minimal group achievable number of normalized tests $\tau^*$ satisfies*

$$\tau^* \ge \beta(1-\theta), \tag{5}$$

*where $\beta$ is the inverse of the expected infection burst length.*

The proof of Theorem 2 is based on an entropy bound and is considered in Appendix D. In the i.i.d. setting, where the components of $U^n$ are independent, the corresponding lower bound is

$$\tau^* \ge 1-\theta, \tag{6}$$

as shown in prior work (Wolf, 1985). In the i.i.d. setting, this lower bound characterizes the fundamental rate below which no non-adaptive strategy can succeed. In contrast, under the Markovian correlation model, the achievable rate scales directly with the factor $\beta$, reflecting the influence of correlations in the infection process. The following theorem formalizes this intuition by providing an explicit achievability bound under this model.

**Theorem 3 (Achievability)** *Under the Markovian correlation model, the minimal group achievable number of normalized tests $\tau^*$ satisfies*

$$\tau^* \le \min_{\varepsilon,c,C,\xi} \max\left\{ \frac{\theta}{\varepsilon\log(1/\varepsilon)}, -\frac{1}{\bar{\varepsilon}\log(1-e^{-m\ln 2})}, -\frac{\theta}{\bar{\varepsilon}\log(1-e^{-m\ln 2 c^C})} \right\} \frac{\beta}{\bar{c}\ln 2}, \tag{7}$$

*where $\bar{c} = 1-c \in [0,1]$, $\bar{\varepsilon} = 1-\varepsilon \in [0,1]$ and*

$$m := \left(1+\xi+\frac{(1-\beta)^C}{1-(1-\beta)^C} + (1-\xi)(1-\beta)^{\xi C}\right).$$

*In particular, when $\theta \to 0$, under an appropriate choice of $(\epsilon, c, C, \xi) \in [0,1]\times[0,1]\times\mathbb{N}\times[0,1]$,*

$$\tau^* \le \frac{\beta}{\ln 2} \approx 1.44\beta. \tag{8}$$

This bound is achievable using a non-adaptive randomized block testing design and a per-item decoding rule. A *per-item decoder* determines whether item $i$ is infected based solely on the $i$-th column of the test matrix $\mathsf{X}$ and the test outcome vector $Y$.

Comparing Theorems 2 and 3, we see as $\theta \to 0$ that our testing strategy achieves a number of tests $T(n)$ that is within a factor of $1/\ln 2$ of the fundamental limit. This parallels a similar result that exists for per-item decoders in the i.i.d. case (Scarlett and Cevher, 2018). This gap arises from the use of per-item decoding, which, while computationally efficient, sacrifices some optimality compared to joint decoding strategies. Nonetheless, the proposed strategy significantly improves upon naive designs that ignore the Markovian correlation structure.

## 4 ALGORITHM

To exploit the correlation among items, we adopt a two-stage testing procedure as shown in Figure 3. Partition the $n$ items into equally sized contiguous blocks

$\{B_1, B_2, \ldots, B_{n/C}\}$ with $C$ items each, where $C \in \mathbb{N}$ is a constant. Specifically,

$$B_i := \{(i-1)C + 1, (i-1)C + 2, \ldots, iC\}.$$

First each block $B_i$ is selected with probability $p_1$. We then select each item within the selected blocks independently with probability $p_2$. More formally, each row of the test matrix $\mathsf{X}$ is independently generated as follows. For each block $B_\ell$, $\ell = 1, \ldots, n/C$, we draw independent Bern($p_1$) random variables $W_\ell$, which determine whether the block is selected during the coarse selection phase. For each item $j = 1, \ldots, n$, we also draw independent Bern($p_2$) random variables $Z_j$, which determine whether an item within a selected block is tested during the fine selection phase. The entries on the $i$th row of the test matrix $X$ are then

$$X_{ij} = W_\ell Z_j, \quad \text{for } j \in B_\ell. \tag{9}$$

In this way, $X_{ij} = 1$ if and only if block $B_\ell$ containing item $j$ is selected in the first selection phase ($W_t = 1$) and item $j$ is independently selected in the fine selection phase ($Z_j = 1$).

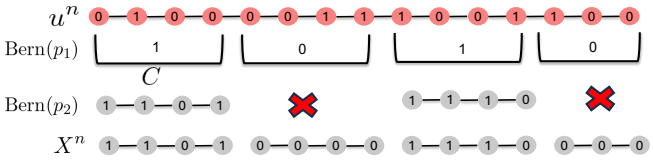


Figure 3: Grouped testing which selects groups with Bern($p_1$). Within selected groups, each item is selected with Bern($p_2$).

We employ a *per-item* decoding rule that determines the infection status of each item $i$ using two indicators: an intermediate estimate $\tilde{u}_i$ to flag potential infections and a final estimate $\hat{u}_i$. The decoding process consists of two steps: Observe that, for each item $i$, if there exists any test $t$ where the item was included ($X_{t,i} = 1$) and the test result was negative ($Y_t = 0$), the item is not infected. In this case, we set $\tilde{u}_i = 0$. If no such test exists, set $\tilde{u}_i = 1$, i.e.,

$$\tilde{u}_i = \begin{cases} 0, & \text{if } \exists\, t \text{ such that } (X_{t,i} = 1, Y_t = 0) \\ 1, & \text{otherwise.} \end{cases}$$

This rule corresponds exactly to the well-known Combinatorial Orthogonal Matching Pursuit (COMP) decoding algorithm (Aldridge et al., 2019), which only relies on the elimination of items appearing in negative tests. In our setting, however, we go beyond COMP by exploiting additional information available in the following way. If $\tilde{u}_i = 0$, the item is non-infected, so set $\hat{u}_i = 0$. If $\tilde{u}_i = 1$, compute the total number of tests in which item $i$ participates:

$$X_i := \sum_{t=1}^{T} X_{t,i}.$$

Compare $X_i$ to a threshold $\gamma = p(1-\varepsilon) \cdot T$, where $p := p_1 p_2$. This choice of $\gamma$ will be explained in the error analysis section below. If $X_i \geq \gamma$, set $\hat{u}_i = 1$, otherwise set $\hat{u}_i = 0$, i.e.

$$\hat{u}_i = \begin{cases} 1, & \text{if } X_i \geq \gamma \\ 0, & \text{otherwise.} \end{cases}$$

Note that this decoder is polynomial time, since it computes the sum of $T \sim O(n^\theta \log n)$ tests, over $n$ items. We now theoretically analyze the probability of error of the above testing and decoding rule.

## 5 PROOF OF THEOREM 3

Let $\mathcal{E}_i$ denote the event that item $i$ is misclassified; that is, a *false positive* or *false negative* event occurs. Then the total probability of error is bounded as

$$\Pr(\mathcal{E}) = \Pr\left(\bigcup_{i=1}^{n} \mathcal{E}_i\right) \overset{(a)}{\leq} \sum_{i=1}^{n} \Pr(\mathcal{E}_i) \overset{(b)}{\leq} n \Pr\left(\mathcal{E}_{\max}\right). \tag{10}$$

The step in (a) follows from the union bound, while in (b) we define $\Pr(\mathcal{E}_{\max}) = \max_i \Pr(\mathcal{E}_i)$.

To analyze the probability of error, we now focus on a single event $\mathcal{E}_i$. To analyze the probability of error, we now focus on a single event $\mathcal{E}_i$. The analysis below is essentially independent of the specific choice of $i$ and applies uniformly up to a mild edge effect in the block design. In particular, when $C$ is not an integer, the final block may have size $\lfloor C \rfloor$ while the remaining blocks have size $\lceil C \rceil$, which slightly breaks exact symmetry across items. To avoid a technically cumbersome treatment of this asymmetry, we work with the worst-case per-item error event $\mathcal{E}_{\max}$. Let $V_j \in \{0, 1\}$ denote the random variable that indicates if block $B_j$ is infected (i.e., contains at least one infected item). The probability $\tilde{q}$ that a given block is infected is

$$\begin{aligned} \tilde{q} &:= 1 - \Pr(U_1 = 0, U_2 = 0, \ldots, U_C = 0) \\ &= 1 - (1 - q_n)(1 - \alpha)^{C-1}. \end{aligned} \tag{11}$$

Let $\Pr_{\text{FN}} := \Pr(\hat{U}_1 = 0 \mid U_1 = 1)$ and $\Pr_{\text{FP}} := \Pr(\hat{U}_1 = 1 \mid U_1 = 0)$ denote the false negative and false positive probabilities for item 1, respectively. Then

$$\Pr\left(\mathcal{E}_1\right) = q_n \Pr_{\text{FN}} + (1 - q_n) \Pr_{\text{FP}}. \tag{12}$$

A false negative error occurs when item $i$ is infected ($U_i = 1$) but is incorrectly decoded as non-infected. This can only arise during the thresholding step, where $\hat{u}_i = 0$ is mistakenly assigned. This happens if the number of tests that contains the infected item falls below the threshold $\gamma$. The false negative probability of error, denoted by $\Pr_{\text{FN}}$, can be upper bounded, for any $\gamma < p \cdot T$, as

$$\Pr_{\text{FN}} = \Pr\left(\sum_{t=1}^{T} X_{t,1} \leq \gamma \middle| U_1 = 1\right) \leq 2^{-T \cdot D(\gamma/T \| p)}, \tag{13}$$

where $D(a\|b)$ denotes the Kullback–Leibler (KL) divergence between the distributions Bern($a$) and Bern($b$). This bound directly follows from the Chernoff bound (Hoeffding, 1963). To see this note that the block structure causes the variables $X_{t,i}$ to be dependent across $i$ for a fixed $t$, but they remain independent across $t$ for a fixed $i$. This independence ensures that the sequence of indicator random variables $\{X_{t,1}\}_{t=1}^{T}$ are i.i.d. with distribution Bern($p$), allowing us to apply the Chernoff bound.

We now bound the probability of a false positive. A false positive error occurs when item $i$ is not infected but is decoded as infected, which requires: (1) the item passes the initial screening, i.e., $\tilde{u}_i = 1$, and (2) it exceeds the threshold $\gamma$ of tests required to set $\hat{u}_i = 1$. Define the event

$$\mathcal{E}_{1,\text{FP}} := \{\tilde{u}_1 = 1, X_1 \geq \gamma\}.$$

Recall that $X_i = \sum_{t=1}^{T} X_{t,i}$. For $\gamma = p(1-\varepsilon)T$, we have

$$\begin{aligned}
\Pr_{\text{FP}} &= \Pr(\mathcal{E}_{1,\text{FP}} | U_1 = 0) \leq \frac{1}{1-q_n} \Pr(\mathcal{E}_{1,\text{FP}}) \\
&\overset{(a)}{=} \frac{1}{1-q_n} \sum_{t=p(1-\varepsilon)T}^{T} \Pr(\tilde{u}_1 = 1 |\, X_1 = t) \Pr(X_1 = t) \\
&\overset{(b)}{\leq} \frac{1}{1-q_n} \Pr(\tilde{u}_1 = 1 | X_1 = p\bar{\varepsilon}T) = \frac{1}{1-q_n} f(p\bar{\varepsilon}T),
\end{aligned} \tag{14}$$

where $\bar{\varepsilon} = 1-\varepsilon$ and $f(\gamma) := \Pr(\tilde{u}_1 = 1 | X_1 = \gamma)$. Here, $(a)$ uses the total law of probability and $(b)$ uses the fact that $\Pr(\tilde{u}_1 = 1 | X_1 = t)$ is a non-increasing function of $t$. To understand $(b)$, note that intuitively, a non-infected item that is tested more frequently is less likely to be mistakenly identified as infected in the first decoding step. This is because it has a higher chance of appearing in negative tests.

The following lemma formalizes this intuition by establishing that the probability of a false positive identification, $f(\gamma)$, is a non-increasing function of $\gamma$. Additionally, the lemma bounds this conditional probability as a function of $U^n$, the infection vector. Let

$$V := \sum_{j=2}^{n/C} V_j, \quad U := \sum_{j=2}^{C} U_j,$$

where each random variable $V_j \in \{0,1\}$ indicates whether at least one item in block $B_j$ is infected. Note that $V^n$ is a deterministic function of $U^n$.

**Lemma 1** *For fixed $p_1, p_2$, and $C$, the function $f(\gamma)$ is non-increasing with respect to $\gamma$. Furthermore, it is upper bounded by*

$$f(\gamma) \leq \mathbb{E}\left[\left(1 - (\bar{p}_1)^V (\bar{p}_2)^U\right)^{\gamma}\right],$$

*where $\bar{p}_i = 1 - p_i,\ i = 1, 2$.*

**Proof of Lemma 1.** We apply the law of total probability to $f(\gamma)$ with respect to the random vector $U^n$, which captures the infection status of all the blocks.

$$\begin{aligned}
f(\gamma) &= \sum_{u^n} \Pr(\tilde{u}_1 = 1 | X_1 = \gamma, u^n) \Pr(u^n) = \\
&\mathbb{E}_{U^n}\left[\Pr(\tilde{u}_1 = 1 | X_1 = \gamma, U^n)\right] \overset{(a)}{=} \mathbb{E}_{U^n}\left[g(U^n)^{\gamma}\right],
\end{aligned} \tag{15}$$

where we define the function

$$g(U^n) := \Pr(Y_1 = 1 | X_{1,1} = 1, U^n).$$

Here, for a fixed realization $U^n = u^n$, the randomness is only over the test design. Since different tests are drawn independently, the test outcomes are conditionally independent given $U^n$. Therefore, conditioned on $X_1 = \gamma$ and $U^n$, the event $\tilde{u}_1 = 1$ depends on $\gamma$ i.i.d. tests containing item 1, so

$$\Pr(\tilde{u}_1 = 1 | X_1 = \gamma, U^n) = g(U^n)^{\gamma},$$

where

$$g(U^n) \triangleq \Pr(Y = 1 | X_{1,1} = 1, U^n)$$

is the conditional probability that a single test containing item 1 is positive. Clearly $f(\gamma)$ is a non-increasing function of $\gamma$. To evaluate $f(\gamma)$, we look closely at $g(U^n) = 1 - \Pr(Y_1 = 0 | X_{1,1} = 1, U^n)$.

In particular, $Y_1 = 0$ occurs if no infected item in the pool contributes to a positive test outcome. For each block that item 1 does not belong to (namely $B_2, B_3, \ldots, B_C$), the event $\{Y_1 = 0\}$ requires one of the following scenarios:

1. The block is not infected. In this case, it cannot contribute to a positive test.

2. The block is infected but is not selected for testing, which happens with probability $(1-p_1)$.
3. The block is infected and is selected for testing (which occurs with probability $p_1$), but none of its infected items are ultimately tested. This last scenario happens with probability $(1-p_2)^{\sum_{j=1}^{C} U'_j}$, where $U'_j$ indicates whether a particular item in that block is infected, given that the block is infected.

Hence, if a block is infected, the probability it does not contribute to a positive test is

$$(1-p_1)+p_1\,(1-p_2)^{\sum_{j=1}^{C} U'_j} \geq (1-p_1).$$

Discarding the second term in the above equation helps simplify the analysis. There are $\sum_{j=2}^{n/C} V_j$ infected blocks among those that do not contain item 1. The probability that all those infected blocks fail to produce a positive test is (recall that $V_j$ is a deterministic function of $U_j$)

$$\left((1-p_1)+p_1\,(1-p_2)^{\sum_{j=1}^{C} U'_j}\right)^{\sum_{j=2}^{n/C} V_j} \geq (1-p_1)^{V}.$$

Within the block containing item 1, the probability that none of its other infected items are included in the test is $(1-p_2)^{\sum_{j=2}^{C} U_j}$. Putting these pieces together, $g(U^n)$ can be upper bounded by

$$g(U^n) \leq 1-(1-p_1)^{V}\,(1-p_2)^{U},$$

where $V := \sum_{j=2}^{n/C} V_i$ and $U := \sum_{j=2}^{C} U_i$. Collecting all steps, we arrive at the bound

$$f(\gamma) = \mathbb{E}[g(U^n)^\gamma] \leq \mathbb{E}\left[\left(1-(1-p_1)^{V}(1-p_2)^{U}\right)^{\gamma}\right],$$

as stated in the lemma. □

Consider the total number of infected blocks $V$. If the indicators $V_j$ were i.i.d., standard concentration inequalities could be directly applied to show that $V$ is close to its expectation with high probability. However, due to the Markov infection model, the $V_j$'s are not independent. To address this, we develop a new concentration inequality tailored for this sum, which may be of independent interest. Lemma 2 formalizes this result.

**Lemma 2** *For any $\delta > 0$ and $\xi \in (0,1)$, the number of infected blocks $V$ satisfies*

$$\Pr(V \geq (1+\delta)mn\alpha) \leq 2^{-n\alpha\xi(1-\beta)^{\xi C}\delta^2/(3(1+\delta))}, \quad (16)$$

*where $m := m_{\delta,\xi,C}$ is given by*

$$m = \left(1+\xi+(1+\delta)\frac{\bar{\beta}^C}{1-\bar{\beta}^C}+(1-\xi)(1-\beta)^{\xi C}\right),$$

*where $\bar{\beta} := 1-\beta$ and $\lim_{\delta,\xi\to(0,0)} \lim_{C\to\infty} m_{\delta,\xi,C} = 1$.*

The proof leverages the dependencies between infection bursts and their propagation through blocks, requiring a careful argument. The full proof of this theorem, is provided in Appendix A. Some works, such as Gillman (1998), analyze sums of functions over dependent mixing processes, which relate to the bound derived here, though they primarily focus on sums of functions applied on individual states and use significantly different proof techniques.

Define the event $\mathcal{B} := \{V \leq K\}$, where $K := (1+\delta)mn\alpha$. Using the bound from Lemma 1 and applying the law of total probability with event $\mathcal{B}$

$$\begin{aligned}
&\mathbb{E}\left[\left(1-\bar{p}_1^{V}\bar{p}_2^{U}\right)^{\gamma}\right] = \mathbb{E}\left[\left(1-\bar{p}_1^{V}\bar{p}_2^{U}\right)^{\gamma}\middle|\mathcal{B}\right]\Pr(\mathcal{B})+\\
&\mathbb{E}\left[\left(1-\bar{p}_1^{V}\bar{p}_2^{U}\right)^{\gamma}\middle|\overline{\mathcal{B}}\right]\Pr(\overline{\mathcal{B}}) \overset{(a)}{\leq} \mathbb{E}\left[\left(1-\bar{p}_1^{K}\bar{p}_2^{U}\right)^{\gamma}\middle|\mathcal{B}\right]\\
&+o(1/n), \qquad (17)
\end{aligned}$$

where $(a)$ is due to $\Pr(\mathcal{B}) \leq 1$ and Lemma 2. Consider the first term in the above equation. Recall that $V_1 = 1$ implies that block 1 is infected. Therefore, we obtain the upper bound

$$\begin{aligned}
&\mathbb{E}\left[\left(1-\bar{p}_1^{K}\bar{p}_2^{U}\right)^{\gamma}\middle|\mathcal{B}\right] \leq \mathbb{E}\left[\left(1-\bar{p}_1^{K}\bar{p}_2^{U}\right)^{\gamma}\middle|\mathcal{B},V_1=0\right]\\
&+\mathbb{E}\left[\left(1-\bar{p}_1^{K}\bar{p}_2^{U}\right)^{\gamma}\middle|\mathcal{B},V_1=1\right]\Pr(V_1=1|\mathcal{B})\\
&\overset{(a)}{\leq} \left(1-\bar{p}_1^{K}\right)^{\gamma}+\left(1-\bar{p}_1^{K}\bar{p}_2^{C}\right)^{\gamma}\frac{\tilde{q}}{1-o(1/n)}\\
&= \left(1-\bar{p}_1^{K}\right)^{\gamma}+(1+o(1))\left(1-\bar{p}_1^{K}\bar{p}_2^{C}\right)^{\gamma}\tilde{q}, \qquad (18)
\end{aligned}$$

where $(a)$ is justified as follows. Given $V_1 = 0$, we must have $U = 0$ since there are no infected items in block 1, which implies

$$\mathbb{E}\left[\left(1-\bar{p}_1^{K}\bar{p}_2^{U}\right)^{\gamma}\middle|\mathcal{B},V_1=0\right] = \left(1-\bar{p}_1^{K}\right)^{\gamma}.$$

Now, when $V_1 = 1$, a loose upper bound is obtained by assuming that all items in the block are infected, i.e., $U = C$, which leads to

$$\mathbb{E}\left[\left(1-\bar{p}_1^{K}\bar{p}_2^{U}\right)^{\gamma}\middle|\mathcal{B},V_1=1\right] \leq \left(1-\bar{p}_1^{K}\bar{p}_2^{C}\right)^{\gamma}.$$

Finally, we obtain

$$\Pr(V_1=1|\mathcal{B}) \leq \frac{\Pr(V_1=1)}{\Pr(\mathcal{B})} = \frac{\tilde{q}}{\Pr(\mathcal{B})}.$$

Using Lemma 2, we substitute $\Pr(\mathcal{B}) \geq 1-o(1/n)$. Combining (14) (17), (18) and Lemma 1 we can conclude that

$$\begin{aligned}
\Pr_{\mathrm{FP}} \leq \frac{1}{1-q_n}&\left(\left(1-\bar{p}_1^{K}\right)^{\gamma}+(1+o(1))\left(1-\bar{p}_1^{K}\bar{p}_2^{C}\right)^{\gamma}\tilde{q}\right)\\
&+o\left(\frac{1}{n}\right). \qquad (19)
\end{aligned}$$

Now from (10), (12), (13) and (19), the total error $\Pr(\mathcal{E})$ is upper bounded by

$$\Pr(\mathcal{E}) \leq nq2^{-T\cdot D(\gamma/T\|p)} + n\left(1-p_1^K\right)^\gamma + (1+o(1))n\tilde{q}\cdot\left(1-\bar{p}_1^K\bar{p}_2^C\right)^\gamma + o(1). \quad (20)$$

The following lemma then shows that under appropriate choices of $p_1, p_2$, and $C$, and $\tau$ as defined in Definition 1 , the total probability of error vanishes as $n\to\infty$.

**Lemma 3** *Let the parameters $p_1, p_2$ and $T$ be set as*

$$p_1 = \frac{\ln 2}{n\alpha}, \quad p_2 = 1-c, \quad T = \tau\cdot(nq_n\log n)$$

*where $\tau$ is the achievable number of normalized tests. Then for any $\tau$ such that*

$$\tau > \min_{\varepsilon,c,C,\xi}\max\left\{\frac{\theta}{\varepsilon\log(1/\varepsilon)}, -\frac{1}{\bar{\varepsilon}\log(1-e^{-m\ln 2})}, -\frac{\theta}{\bar{\varepsilon}\log(1-e^{-m\ln 2 c^C})}\right\}\frac{\beta}{\bar{c}\ln 2},$$

*where $\bar{c} := 1-c$, $\bar{\varepsilon} := 1-\varepsilon$, the overall probability of error satisfies*

$$\Pr(\mathcal{E}) \leq nq\,2^{-T\cdot D(\gamma/T\|p)} + n(1-\bar{p}_1^K)^\gamma + (1+o(1))\,n\tilde{q}(1-\bar{p}_1^K\bar{p}_2^C)^\gamma \to 0, \quad (21)$$

*as $n\to\infty$.*

Lemma 3 is proved in Appendix B. The definition of $\tau^*$ (Definition 1) implies that Lemma 3 is equivalent to part I of Theorem 3. Part II of Theorem 3 is proved in Appendix C.

# 6 CONCLUSION

In this work, we introduced a novel Markovian correlation model for correlated group testing. We proposed a novel *per-item* testing and decoding strategy for non-adaptive group testing in the presence of correlated infection patterns.

Future work could extend this framework to more general correlation structures, such as higher-order Markov chains or graphical models, to capture long-range dependencies and complex item relationships, as well as higher order-dependent bursty patterns. Exploring joint decoding strategies, which consider the infection status of all items simultaneously, could further reduce the number of required tests, though this requires correlated non-adaptive test designs that remain analytically tractable. Another natural direction is to explore alternative test constructions that incorporate correlations within the design itself, and to develop converse bounds for correlated models to better quantify the gap between simple per-item decoding and more powerful joint methods.

# 7 ACKNOWLEDGEMENTS

We would like to thank Prof. Jonathan Scarlett for noticing an error in an earlier version of this work and for the subsequent discussions that helped us improve both the paper and the results.

The work of A.N.R. and I.S. was supported in part by the National Science Foundation (NSF) under grants 2046991 and 2508507.

## References

David J Balding, William J Bruno, David C Torney, and Emanuel Knill. A comparative survey of non-adaptive pooling designs. In *Genetic mapping and DNA sequencing*, pages 133–154. Springer, 1996.

Robert Dorfman. The detection of defective members of large populations. *The Annals of mathematical statistics*, 14(4):436–440, 1943.

Matthew Aldridge, Oliver Johnson, Jonathan Scarlett, et al. Group testing: an information theory perspective. *Foundations and Trends® in Communications and Information Theory*, 15(3-4):196–392, 2019.

Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, and Philipp Loick. Information-theoretic and algorithmic thresholds for group testing. *IEEE Transactions on Information Theory*, 66(12):7911–7928, 2020a.

Jonathan Scarlett and Volkan Cevher. Phase transitions in the pooled data problem. *Advances in Neural Information Processing Systems*, 30, 2017.

Jonathan Scarlett and Volkan Cevher. Near-optimal noisy group testing via separate decoding of items. *IEEE Journal of Selected Topics in Signal Processing*, 12(5):902–915, 2018.

Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, and Philipp Loick. Optimal group testing. In *Conference on Learning Theory*, pages 1374–1388. PMLR, 2020b.

Surin Ahn, Wei-Nina Chen, and Ayfer Özgür. Adaptive group testing on networks with community structure. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1242–1247. IEEE, 2021.

Pavlos Nikolopoulos, Sundara Rajan Srinivasavaradhan, Tao Guo, Christina Fragouli, and Suhas Diggavi. Group testing for connected communities. In *International Conference on Artificial Intelligence and Statistics*, pages 2341–2349. PMLR, 2021.

Pavlos Nikolopoulos, Sundara Rajan Srinivasavaradhan, Tao Guo, Christina Fragouli, and Suhas N Diggavi. Community-aware group testing. *IEEE*

*Transactions on Information Theory*, 69(7):4361–4383, 2023.

Surin Ahn, Wei-Ning Chen, and Ayfer Özgür. Noisy adaptive group testing for community-oriented models. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1621–1626. IEEE, 2023.

Paolo Bertolotti and Ali Jadbabaie. Network group testing. *arXiv preprint arXiv:2012.02847*, 2020.

Batuhan Arasli and Sennur Ulukus. Group testing with a graph infection spread model. *Information*, 14(1):48, 2023.

Charles J Colbourn. Group testing for consecutive positives. *Annals of Combinatorics*, 3(1):37–41, 1999.

Yun-Han Li, Ryan Gabrys, Jin Sima, Ilan Shomorony, and Olgica Milenkovic. Finding a burst of positives via nonadaptive semiquantitative group testing. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1848–1853. IEEE, 2023.

Thach V. Bui, Yeow Meng Chee, Jonathan Scarlett, and Van Khu Vu. Group testing with blocks of positives. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 1082–1087, 2022. doi: 10.1109/ISIT50566.2022.9834611.

Jeffrey R Moffitt and Xiaowei Zhuang. Rna imaging with multiplexed error-robust fluorescence in situ hybridization (merfish). In *Methods in enzymology*, volume 572, pages 1–49. Elsevier, 2016.

Amanda Janesick, Robert Shelansky, Andrew D Gottscho, Florian Wagner, Stephen R Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A Morrison, Michelli F Oliveira, Jordan T Sicherman, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nature Communications*, 14(1): 8353, 2023.

Dominic J Allocco, Isaac S Kohane, and Atul J Butte. Quantifying the relationship between co-expression, co-regulation and gene function. *BMC bioinformatics*, 5:1–10, 2004.

Brian Cleary, Brooke Simonton, Jon Bezney, Evan Murray, Shahul Alam, Anubhav Sinha, Ehsan Habibi, Jamie Marshall, Eric S Lander, Fei Chen, et al. Compressed sensing for highly efficient imaging transcriptomics. *Nature biotechnology*, 39(8): 936–942, 2021.

Matthew Aldridge. The capacity of bernoulli nonadaptive group testing. *IEEE Transactions on Information Theory*, 63(11):7142–7148, 2017. doi: 10.1109/TIT.2017.2748564.

Ayelet C Portnoy and Alejandro Cohen. Non-adaptive multi-stage algorithm for group testing with prior statistics. In *2024 60th Annual Allerton Conference on Communication, Control, and Computing*, pages 1–8. IEEE, 2024.

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

Mikhail Borisovich Malyutov and Plamen Stoyanov Mateev. Planning of screening experiments for a nonsymmetric response function. *Mathematical notes of the Academy of Sciences of the USSR*, 27: 57–68, 1980.

Toby Berger and Vladimir I Levenshtein. Asymptotic efficiency of two-stage disjunctive testing. *IEEE Transactions on Information Theory*, 48(7):1741–1749, 2002.

FK Hwang. A generalized binomial group testing problem. *Journal of the American Statistical Association*, 70(352):923–926, 1975.

Tongxin Li, Chun Lam Chan, Wenhao Huang, Tarik Kaced, and Sidharth Jaggi. Group testing with prior statistics. In *2014 IEEE International Symposium on Information Theory*, pages 2346–2350. IEEE, 2014.

Tom Kealy, Oliver Johnson, and Robert Piechocki. The capacity of non-identical adaptive group testing. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 101–108. IEEE, 2014.

Yaw-Ling Lin, Charles Ward, and Steven Skiena. Synthetic sequence design for signal location search. In *Research in Computational Molecular Biology: 16th Annual International Conference, RECOMB 2012, Barcelona, Spain, April 21-24, 2012. Proceedings 16*, pages 165–179. Springer, 2012.

Meinard Müller and Masakazu Jimbo. Consecutive positive detectable matrices and group testing for consecutive positives. *Discrete mathematics*, 279(1-3):369–381, 2004.

Thach V Bui, Mahdi Cheraghchi, and Thuc D Nguyen. Improved algorithms for non-adaptive group testing with consecutive positives. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1961–1966. IEEE, 2021.

Ivan Lau, Jonathan Scarlett, and Yang Sun. Model-based and graph-based priors for group testing. *IEEE Transactions on Signal Processing*, 70:6035–6050, 2022.

Nicholas JA Harvey, Mihai Patrascu, Yonggang Wen, Sergey Yekhanin, and Vincent WS Chan. Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs. In *IEEE*

*INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pages 697–705. IEEE, 2007.

Mahdi Cheraghchi, Amin Karbasi, Soheil Mohajer, and Venkatesh Saligrama. Graph-constrained group testing. *IEEE Transactions on Information Theory*, 58(1):248–262, 2012.

Amin Karbasi and Morteza Zadimoghaddam. Sequential group testing with graph constraints. In *2012 IEEE information theory workshop*, pages 292–296. Ieee, 2012.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

Aditya N Ravi and Ilan Shomorony. Fundamental limits of non-adaptive group testing with markovian correlation. *arXiv preprint*, https://arxiv.org/abs/2501.12588, 2025. Preprint.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

Daniel G Brown. How i wasted too long finding a concentration inequality for sums of geometric variables. *online*, 2011.

## Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

# A Proof of Lemma 2

**Lemma 2** *For any $\delta > 0$ and $\xi \in (0,1)$, the number of infected blocks $V$ satisfies*

$$\Pr(V \geq (1+\delta)mn\alpha) \leq 2^{-n\alpha\xi(1-\beta)^{\xi C}\delta^2/(3(1+\delta))}, \tag{16}$$

*where $m := m_{\delta,\xi,C}$ is given by*

$$m = \left(1 + \xi + (1+\delta)\frac{\bar{\beta}^C}{1-\bar{\beta}^C} + (1-\xi)(1-\beta)^{\xi C}\right),$$

*where $\bar{\beta} := 1-\beta$ and $\lim_{\delta,\xi\to(0,0)} \lim_{C\to\infty} m_{\delta,\xi,C} = 1$.*

Define the number of 0 to 1 transitions in the infection vector $U^n$ as $\tilde{Z}$, given by

$$\tilde{Z} := \sum_{i=1}^{n-1} \mathbf{1}\{U_i = 0, U_{i+1} = 1\}.$$

A crucial observation is that whenever $U_i = 0$, the transition $(U_i = 0) \to (U_{i+1} = 1)$ occurs with probability $\alpha$, independent of the rest of the process. Introducing an independent sequence of Bern($\alpha$) random variables $\{Z_i\}_{i=1}^n$, we can rewrite the transition indicator as

$$\mathbf{1}\{u_i = 0, u_{i+1} = 1\} = \mathbf{1}\{u_i = 0\} \cdot Z_i.$$

This establishes the bound

$$\tilde{Z} = \sum_{i=1}^{n-1} \mathbf{1}\{u_i = 0\} \cdot Z_i \leq \sum_{i=1}^{n} Z_i := Z.$$

Thus, $Z$ is an upper bound on the total number of initial infection points. Intuitively, for large values of $C$, we expect most bursts of infections to remain confined within single blocks. However, if a burst of infections starts near the end of a block, a single burst can cause multiple blocks to become infected.

To analyze this, we split the summation $Z$ into two separate sums. Define the set of indices corresponding to the first $(1-\xi)$ fraction of each block as

$$S := \bigcup_{j=0}^{\frac{n}{C}-1} \{i : i \in \{jC+1, jC+2, \ldots, jC+(1-\xi)C\}\}.$$

Note that $|S| = n(1-\xi)$. We can then decompose $Z$ as

$$Z = \sum_{i\in S} Z_i + \sum_{i\in S^c} Z_i = Z_S + Z_{S^c}.$$

This decomposition allows us to separately analyze the contributions from infections occurring within the main body of each block and those occurring near block boundaries. Let $M_j$ be the number of blocks infected by the $j^{\text{th}}$ burst.

If $j \in S$, then for the infection to spread beyond the current block, it must first accumulate at least $\xi C$ infections to reach the boundary. Once the boundary is reached, each subsequent block requires an additional $C$ infections for further spread. Therefore, we can bound $M_j$ using a sum of two random variables. To get an upper bound, we assume the worst case, i.e. that all infections start from $(1-\xi)C$. Now we can define $N_j \sim \text{Bern}((1-\beta)^{\xi C})$, which indicates whether the infection successfully crosses the first boundary. If $N_j = 0$, the burst of infections remains within its initial block. If $N_j = 1$, the burst of infections spills into the next block. After crossing the first boundary, the infection continues spreading according to a Geometric distribution, where each subsequent boundary crossing occurs with probability $(1-\beta)^C$. Let $O_j \sim \text{Geom}(1-(1-\beta)^C)$ represent the number of additional blocks infected beyond the first successful crossing. Here we assume that the support of the Geometric RV is $\{0, 1, 2, \ldots\}$, thus accounting for the infection of the first block. Thus, we obtain a deterministic upper bound to $M_j$ as a sum of two random variables, given by

$$M_j \leq N_j + O_j + 1.$$

If $j \notin S$, to obtain an upper bound, we assume all infections starts near the boundary of a block. Unlike the previous case, the infection does not require a full $\xi C$ accumulation to reach the boundary since it is already positioned close to the next block. This makes it more likely that the infection will at least spill over into one additional block. To account for this, we employ a looser upper bound by assuming that at least two blocks are infected. As before we have each subsequent boundary crossing occurs with probability $(1-\beta)^C$. Thus, we have the bound

$$M_j \leq O_j + 2.$$

This decomposition provides an upper bound on the number of blocks an infection burst can reach. We now assemble all the components to derive a high-probability tail bound on $V$, the total number of infected blocks.

$$\begin{aligned} V \leq \sum_{j=1}^{n} M_j Z_j &= \sum_{j:Z_j=1} M_j \leq \sum_{j:Z_j=1, j\in S} (N_j + O_j + 1) + \sum_{j:Z_j=1, j\notin S} (O_j + 2) \\ &= \sum_{j\in S, Z_j=1} N_j + \sum_{j\in S, Z_j=1} O_j + \sum_{j\in S, Z_j=1} 1 + \sum_{j\notin S, Z_j=1} O_j + \sum_{j\notin S, Z_j=1} 2 \\ &= \sum_{j=1}^{Z_S} N_j + \sum_{j=1}^{Z} O_j + Z_S + 2Z_{S^c} - Z_{S^c} = \underbrace{\sum_{j=1}^{Z} O_j}_{:=O} + \underbrace{\sum_{j=1}^{Z_S} N_j}_{:=N} + Z + Z_{S^c}. \end{aligned}$$

The change in indices is justified by the i.i.d. assumption, which allows us to rewrite the sums of i.i.d. random variables in terms of counts rather than specific indices. Since each $Z_i \sim$ Bernoulli$(\alpha)$ independently, their expectations are:

$$\mathbb{E}[Z_S] \;=\; n(1-\xi)\alpha, \quad \mathbb{E}[Z_{S^c}] \;=\; n\,\xi\,\alpha.$$

**Chernoff Bounds for $Z_S$ and $Z_{S^c}$:** First we develop concentration inequalities on $Z_S, Z_{S^c}$. Since they are just sums of i.i.d. Bern$(\alpha)$ RVs, we can employ a standard Chernoff bound, to obtain

$$\begin{aligned} \Pr\left(Z_S \geq (1+\delta)n(1-\xi)\alpha \text{ or } Z_{S^c} \geq (1+\delta)n\xi\alpha\right) &\overset{(a)}{\leq} \Pr(Z_S \geq (1+\delta)n(1-\xi)\alpha) + \Pr(Z_{S^c} \geq (1+\delta)n\xi\alpha) \\ &\overset{(b)}{\leq} 2^{-n(1-\xi)\frac{\alpha}{1+\delta}\delta^2} + 2^{-n\xi\frac{\alpha}{1+\delta}\delta^2}, \end{aligned} \tag{22}$$

where $(a)$ is due to the union bound and $(b)$ is due to the following version of the Chernoff bound from Hoeffding (1994).

**Lemma 4** *For i.i.d. Bern$(p)$ random variables $X_1, X_2 \ldots X_n$,*

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq (1+\delta)p\right) \leq 2^{-np\delta^2/(1+\delta)}. \tag{23}$$

Now define

$$\mathcal{B}_z = \{Z_S \leq (1+\delta)n(1-\xi)\alpha, Z_{S^c} \leq (1+\delta)n\xi\alpha)\}.$$

**Bounding $O$ and $N$:** Conditioned on $\mathcal{B}_z$, we have $Z \leq (1+\delta)n\alpha$, so

$$O \leq \sum_{j=1}^{(1+\delta)n\alpha} O_j. \tag{24}$$

This is a finite sum of Geometric random variables, which has the Negative Binomial distribution NB$((1+\delta)n\alpha, 1-(1-\beta)^C)$. The following lemma is adapted from Brown (2011) and provides a concentration bound for this quantity when the support of $O_j$ is $\{0,1,\ldots\}$.

**Lemma 5** *Let* NB$(n,p)$ *be a negative binomially distributed random variable that arises as the sum of $n$ i.i.d. geometrically distributed random variables with expectation $\frac{1-p}{p}$. Then,*

$$\mathbb{E}[\text{NB}(n,p)] = \frac{n(1-p)}{p},$$

*and for any* $\delta > 0$,

$$\Pr\left(\mathrm{NB}(n,p) > (1+\delta)\left(\frac{1-p}{p}\right)n\right) \leq \exp\left(-\frac{n\delta^2}{2(1+\delta)}\right).$$

Since $O_j \sim \mathrm{Geom}(1-(1-\beta)^C)$, its expectation is given by

$$\mathbb{E}[O_j] = \frac{(1-\beta)^C}{1-(1-\beta)^C}.$$

Applying Lemma 5 and using the bound in (24) we obtain

$$\Pr\left(O \geq (1+\delta')(1+\delta)n\alpha\frac{(1-\beta)^C}{1-(1-\beta)^C}\middle|\mathcal{B}_z\right) \leq \exp\left(-\frac{n\alpha(1+\delta)(\delta')^2}{2(1+\delta')}\right).$$

Now consider the random variable $N$. Again note that conditioning on $\mathcal{B}_z$ we have

$$N \leq \sum_{j=1}^{(1+\delta)n\alpha(1-\xi)} N_j. \tag{25}$$

This is just a sum of i.i.d. $\mathrm{Bern}((1-\beta)^{\xi C})$ random variables. We can employ the Chernoff bound in Lemma 4 to upper bound this. Therefore using (25), we can state that

$$\Pr\left(N \geq (1+\delta')(1+\delta)n\alpha(1-\xi)(1-\beta)^{\xi C}\middle|\mathcal{B}_z\right) \leq 2^{-n(1-\xi)(1+\delta)\alpha r(\delta')^2/(1+\delta')},$$

where $r = (1-\beta)^{\xi C}$.

**Establishing an upper bound on $V$:** We establish an upper bound on $V$ by explicitly applying the union bound. We define the following events:

$$\begin{aligned}
A_1 &:= \{Z_{S^c} \leq (1+\delta)n\xi\alpha\},\\
A_2 &:= \{Z \leq (1+\delta)n\alpha\},\\
A_3 &:= \left\{O \leq (1+\delta')(1+\delta)n\alpha\frac{(1-(1-\beta)^C)}{(1-\beta)^C}\right\},\\
A_4 &:= \left\{N \leq (1+\delta')(1+\delta)n\alpha(1-\xi)(1-\beta)^{\xi C}\right\}.
\end{aligned}$$

Each satisfies

$$\begin{aligned}
\Pr(A_1^c) &\leq 2^{-n\xi\frac{\alpha}{1+\delta}\delta^2},\\
\Pr(A_2^c) &\leq 2^{-n(1-\xi)\frac{\alpha}{1+\delta}\delta^2} + 2^{-n\xi\frac{\alpha}{1+\delta}\delta^2},\\
\Pr(A_3^c) &\leq \exp\left(-\frac{n\alpha(1+\delta)(\delta')^2}{2(1+\delta')}\right),\\
\Pr(A_4^c) &\leq 2^{-n(1-\xi)\alpha r(1+\delta)(\delta')^2/(1+\delta')}.
\end{aligned}$$

By the union bound, we have:

$$\begin{aligned}
&\Pr(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c) \leq \Pr(A_1^c) + \Pr(A_2^c) + \Pr(A_3^c) + \Pr(A_4^c)\\
&\leq 2^{-\frac{n\xi\alpha}{1+\delta}\delta^2} + 2^{-\frac{n(1-\xi)\alpha}{1+\delta}\delta^2} + 2^{-\frac{n\xi\alpha}{1+\delta}\delta^2} + 2^{-\frac{n\alpha(1+\delta)\ln 2(\delta')^2}{2(1+\delta')}} + 2^{-\frac{n(1-\xi)\alpha r(1+\delta)}{(\delta')^2(1+\delta')}}.
\end{aligned}$$

Setting $\delta' = \delta$ we obtain assuming $1-\xi > \xi$, or equivalently $\xi < 0.5$, we conclude that:

$$2^{-n\alpha(1-\xi)\delta^2/(1+\delta)} \leq 2^{-n\alpha\xi\delta^2/(1+\delta)}.$$

For the fourth term, using the fact that $\frac{\ln 2}{2} > \frac{\xi}{3}$, we derive:

$$2^{-\frac{n\alpha \ln 2\, \delta^2}{2(1+\delta)}} \leq 2^{-n\alpha\xi\delta^2/(3(1+\delta))}.$$

For the last term, for sufficiently large $n$, since $r \leq 1$, we can show:

$$2^{-n\alpha\xi\delta^2/(3(1+\delta))} \leq 2^{-nr\alpha\xi\delta^2/(3(1+\delta))}.$$

Summing all terms, we obtain the final bound:

$$\Pr(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c) \leq 5 \cdot 2^{-nr\alpha\xi\delta^2/(3(1+\delta))}.$$

Therefore setting $\delta = \delta'$, we note that

$$V \leq O + N + Z_{S^c} + Z \leq (1+\delta)n\alpha \left(1 + \xi + (1+\delta)\frac{(1-\beta)^C}{1-(1-\beta)^C} + (1-\xi)(1-\beta)^{\xi C}\right) = n\alpha(1+\delta)m,$$

where

$$m = m_{\delta,\xi,C} := \left(1 + \xi + (1+\delta)\frac{((1-\beta)^C)}{1-(1-\beta)^C} + (1-\xi)(1-\beta)^{\xi C}\right).$$

Therefore,

$$\Pr(V \geq (1+\delta)kn\alpha) \leq 2^{-n\alpha\xi(1-\beta)^{\xi C}\delta^2/(3(1+\delta))}.$$

## B Proof of Lemma 3

**Lemma 3** *Let the parameters $p_1, p_2$ and $T$ be set as*

$$p_1 = \frac{\ln 2}{n\alpha}, \quad p_2 = 1 - c, \quad T = \tau \cdot (nq_n \log n)$$

*where $\tau$ is the achievable number of normalized tests. Then for any $\tau$ such that*

$$\tau > \min_{\varepsilon,c,C,\xi} \max\left\{\frac{\theta}{\varepsilon\log(1/\varepsilon)}, -\frac{1}{\bar{\varepsilon}\log(1-e^{-m\ln 2})}, -\frac{\theta}{\bar{\varepsilon}\log(1-e^{-m\ln 2}c^C)}\right\}\frac{\beta}{\bar{c}\ln 2},$$

*where $\bar{c} := 1-c$, $\bar{\varepsilon} := 1-\varepsilon$, the overall probability of error satisfies*

$$\begin{aligned}\Pr(\mathcal{E}) \leq\ & nq\, 2^{-T\cdot D(\gamma/T\|p)} + n(1-\bar{p}_1^K)^\gamma \\ & + (1+o(1))\, n\tilde{q}(1-\bar{p}_1^K\bar{p}_2^C)^\gamma \to 0,\end{aligned} \tag{21}$$

*as $n \to \infty$.*

**Proof of Lemma 3** We consider each term separately. Consider the first term. We can rewrite it as

$$nq2^{-T\cdot D(\gamma/T\|p)} = kn^\theta 2^{-T\cdot D(\gamma/T\|p)} \overset{(a)}{=} k2^{\theta\log n - T\cdot D(\gamma/T||p)} \leq k2^{\theta\log n - T\cdot p\varepsilon\log(1/\varepsilon)}.$$

Here $(a)$ follows from

$$D(p(1-\epsilon)\|p) \geq p\varepsilon\log(1/\varepsilon), \tag{26}$$

where we retain just the first term in the definition of $D(x\|y)$. Setting

$$T = (1+\delta_1)\frac{\theta\log n}{p\varepsilon\log(1/\varepsilon)} := T_1(n),$$

allows the first term to vanish as $n \to \infty$. Now consider the second term. Similar to the previous steps

$$n(1-\bar{p}_1^K)^{T\cdot p(1-\varepsilon)} = 2^{\log n + Tp(1-\varepsilon)\log(1-\bar{p}_1^K)}.$$

Setting

$$T = -(1+\delta_2)\frac{\log n}{p(1-\epsilon)\log\left(1-\bar{p}_1^K\right)} := T_2(n),$$

allows the second term to vanish as $n \to \infty$. Finally consider the third term. Similar to previous steps

$$(1+o(1))n\tilde{q}\,\left(1-\bar{p}_1^K\,\bar{p}_2^C\right)^{T\cdot p(1-\varepsilon)} = 2^{\log(n\tilde{q})+Tp(1-\varepsilon)\log\left(1-\bar{p}_1^K\bar{p}_2^C\right)}.$$

Again, setting

$$T = -(1+\delta_3)\frac{\log(n\tilde{q})}{p(1-\epsilon)\log\left(1-\bar{p}_1^K\bar{p}_2^C\right)} := T_3(n),$$

allows the third term to vanish to as $n \to \infty$. Therefore recall the definition of

$$\tau = \lim_{n\to\infty}\frac{T(n)}{nq_n\log n},$$

and combining the above bounds we can claim that

$$\tau = \lim_{n\to\infty}\left(\frac{1}{nq_n\log n}\right)\max\left(T_1,T_2,T_3\right) = \max\left(\lim_{n\to\infty}\frac{T_1(n)}{nq_n\log n},\lim_{n\to\infty}\frac{T_2(n)}{nq_n\log n},\lim_{n\to\infty}\frac{T_3(n)}{nq_n\log n}\right)$$

Now we evaluate each limit. Recall that $nq_n = kn^\theta$ and $n\alpha = k'n^\theta$ The first limit is evaluated as follows

$$\lim_{n\to\infty}\frac{T_1(n)}{nq_n\log n} = \lim_{n\to\infty}(1+\delta_1)\frac{\theta}{nq_np\varepsilon\log(1/\varepsilon)} \tag{27}$$

$$\overset{(a)}{=}(1+\delta_1)\frac{k'\theta}{(1-c)k(\ln 2)\varepsilon\log(1/\varepsilon)} \overset{(b)}{=}(1+\delta_1)\frac{\beta}{(\ln 2)(1-c)\varepsilon\log(1/\varepsilon)}, \tag{28}$$

where $(a)$ follows from the choice of $p_1 = \frac{\ln 2}{n\alpha}$ and $p_2 = 1-c$. The second limit is evaluated as follows (recall $K := (1+\delta)mn\alpha$)

$$\lim_{n\to\infty}\frac{T_2(n)}{nq_n\log n} = -\lim_{n\to\infty}\frac{-(1+\delta_2)n\alpha}{nq_n(\ln 2)(1-\varepsilon)\log\left(1-(1-\ln(2)/n\alpha)^K\right)}$$

$$\overset{(a)}{=}-(1+\delta_2)\frac{\beta}{(1-\epsilon)(\ln 2)(1-c)\log(1-e^{-m\ln 2})}, \tag{29}$$

where recall that from Lemma 2 that

$$m := \left(1+\xi+(1+\delta)\frac{((1-\beta)^C)}{1-(1-\beta)^C}+(1-\xi)(1-\beta)^{\xi C}\right).$$

Here $(a)$ follows from the fact choice of $p_1 = \frac{\ln 2}{n\alpha}$ and noting that

$$\log\left(1-(1-\ln(2)/n\alpha)^K\right) \to \log(1-\exp(-\ln(2)K/n\alpha)) = \log(1-\exp(-m\ln(2)),$$

since $K = nm\alpha$. The third limit is evaluated as follows

$$\lim_{n\to\infty}\frac{T_3(n)}{nq_n\log n} = -\lim_{n\to\infty}\frac{(1+\delta_3)k'\log(n\tilde{q})}{k(\log n)(\ln 2)p_2(1-\epsilon)\log\left(1-\bar{p}_1^K\bar{p}_2^C\right)}$$

$$= -\lim_{n\to\infty}\frac{(1+\delta_3)k'\theta(\log n+\log C)}{k(\log n)p_2(\ln 2)(1-\epsilon)\log\left(1-\bar{p}_1^K\bar{p}_2^C\right)} \overset{(a)}{=} -\lim_{n\to\infty}\frac{(1+\delta_3)k'\theta(1+o(1))}{k(\ln 2)p_2(1-\epsilon)\log\left(1-\bar{p}_1^K\bar{p}_2^C\right)}$$

$$\overset{(b)}{=} -\frac{(1+\delta_3)k'\theta}{kp_2(\ln 2)(1-\epsilon)}\lim_{n\to\infty}\frac{1}{\log\left(1-\bar{p}_1^K\bar{p}_2^C\right)} = -\frac{\beta}{(1-c)\ln 2}\frac{(1+\delta_3)\theta}{(1-\varepsilon)\log(1-e^{-m\ln 2}c^C)},$$

where we note that $\bar{p}_2 = 1-p_2 = 1-(1-c) = c$ Therefore setting $\delta_i \to 0$ for $i \in \{1,2,3\}$, we obtain the first part of Theorem 3, which is

$$\tau^* \le \min_{\varepsilon,c,C,\xi}\max\left\{\frac{\theta}{\varepsilon\log_2(1/\varepsilon)},-\frac{1}{(1-\varepsilon)\log\left(1-e^{-m\ln 2}\right)},\frac{\theta}{(1-\varepsilon)\log\left(1-e^{-m\ln 2}c^C\right)}\right\}\frac{\beta}{\bar{c}\ln 2}. \tag{30}$$

□

## C Proof of part II of Theorem 3

We choose

$$\xi = \frac{1}{\sqrt{C}}, \quad C = \frac{1}{2}\sqrt{\log\left(\frac{1}{\theta}\right)}, \quad c = 2^{-\sqrt{\log\left(\frac{1}{\theta}\right)}}, \quad \varepsilon = \theta,$$

so that

$$c^C = 2^{-C\sqrt{\log\left(\frac{1}{\theta}\right)}} = 2^{-\frac{1}{2}\log\left(\frac{1}{\theta}\right)} = \sqrt{\theta},$$

and note that $\bar{c} = 1 - c \to 1$ as $\theta \to 0$. Moreover, since $\xi = 1/\sqrt{C}$ with $C = \frac{1}{2}\sqrt{\log(1/\theta)}$, we have $\xi \to 0$; also, because $0 < 1 - \beta < 1$ we obtain $(1-\beta)^C \to 0$ and similarly $(1-\beta)^{\xi C} \to 0$. Hence,

$$m = \left(1 + \xi + (1+\delta)\frac{(1-\beta)^C}{1-(1-\beta)^C} + (1-\xi)(1-\beta)^{\xi C}\right) \to 1.$$

Substituting $\varepsilon = \theta$ into the bound

$$\tau^* \leq \min_{\varepsilon,\, c,\, C,\, \xi} \max\left\{\frac{\theta}{\varepsilon \log_2(1/\varepsilon)},\ -\frac{1}{(1-\varepsilon)\log\left(1 - e^{-m\ln 2}\right)},\ -\frac{\theta}{(1-\varepsilon)\log\left(1 - e^{-m\ln 2}\, c^C\right)}\right\} \frac{\beta}{\bar{c}\ln 2},$$

the first term becomes $\frac{1}{\log_2(1/\theta)} \to 0$; with $m \to 1$ we have $e^{-m\ln 2} \to 1/2$ so that the second term is

$$-\frac{1}{(1-\theta)\log\left(1-\frac{1}{2}\right)} = \frac{1}{(1-\theta)} \to 1,$$

while the third term, using $c^C = \sqrt{\theta}$ and approximating $\log\left(1 - \frac{1}{2}\sqrt{\theta}\right) \sim -\frac{1}{2}\sqrt{\theta}$, becomes $-\frac{2\sqrt{\theta}}{1-\theta} \to 0$. Thus the maximum is 1 and, accounting for $\bar{c} \to 1$, we conclude that

$$\tau^* \leq \frac{\beta}{\ln 2}.$$

## D Proof of Theorem 2

**Proof of Converse 2.** We begin by considering a lower bound on the minimal group testing rate, as derived in Wolf (1985). The result is given by:

$$\tau^* > \lim_{n\to\infty} \frac{H(U^n)}{n q_n \log n}, \tag{31}$$

where $q_n = \frac{k n^\theta}{n}$ represents the infection rate, and $H(U^n)$ is the entropy of the infection pattern $U^n = (U_1, U_2, \ldots, U_n)$.

Under the Markovian correlation model parameterized by $\alpha$ and $\beta$, the entropy $H(U^n)$ can be expanded using the chain rule for entropy:

$$H(U^n) = H(U_1) + \sum_{i=2}^{n} H(U_i|U_{i-1}). \tag{32}$$

Since the Markov chain is stationary, the conditional entropy $H(U_i|U_{i-1})$ is the same for all $i \geq 2$, and equals $H(U_2|U_1)$. Therefore, the total entropy simplifies to:

$$H(U^n) = H(U_1) + (n-1)H(U_2|U_1). \tag{33}$$

Next, we expand $H(U_2|U_1)$ using the Markov transition probabilities. By definition:

$$H(U_2|U_1) = qH(U_2|U_1 = 1) + (1-q)H(U_2|U_1 = 0), \tag{34}$$

where $q$ is the stationary probability of $U_1 = 1$. For the conditional entropy $H(U_2|U_1 = 1)$, note that $U_2|U_1 = 1 \sim \text{Bern}(1-\beta)$. Hence,

$$H(U_2|U_1 = 1) = -\beta \log \beta - (1-\beta)\log(1-\beta). \tag{35}$$

Similarly, for $H(U_2|U_1 = 0)$, we have $U_2|U_1 = 0 \sim \text{Bern}(\alpha)$. Thus,

$$H(U_2|U_1 = 0) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha). \tag{36}$$

Substituting these results into the expression for $H(U_2|U_1)$, we obtain

$$\begin{aligned} H(U_2|U_1) = q\left(-\beta \log \beta - (1-\beta)\log(1-\beta)\right) \\ + (1-q)\left(-\alpha \log \alpha - (1-\alpha)\log(1-\alpha)\right). \end{aligned} \tag{37}$$

For large $n$, the infection rate $\alpha = \frac{k'n^\theta}{n}$ dominates due to its asymptotic behavior. Specifically, the leading term of $H(U_2|U_1)$ is given by

$$H(U_2|U_1) = \alpha \log \frac{1}{\alpha} + o\left(\alpha \log \frac{1}{\alpha}\right). \tag{38}$$

Substituting $\alpha = \frac{k'n^\theta}{n}$, we obtain

$$\begin{aligned} H(U_2|U_1) &= \frac{k'n^\theta}{n} \log\left(\frac{n}{k'n^\theta}\right) + o\left(\frac{n^\theta \log n}{n}\right) \\ &= \frac{k'(1-\theta)n^\theta \log n}{n} + o\left(\frac{n^\theta \log n}{n}\right). \end{aligned} \tag{39}$$

Thus, for large $n$, the entropy $H(U^n)$ is approximated as

$$\begin{aligned} H(U^n) &= n \cdot \frac{k'(1-\theta)n^\theta \log n}{n} + o\left(n^\theta \log n\right) \\ &= k'(1-\theta)n^\theta \log n + o\left(n^\theta \log n\right). \end{aligned} \tag{40}$$

Returning to the bound in (31), we substitute $H(U^n)$ and $nq_n = k \log n$, where $k = \frac{k'}{\beta}$, to obtain:

$$\begin{aligned} \tau^* &> \lim_{n\to\infty} \frac{k'(1-\theta)n^\theta \log n + o\left(n^\theta \log n\right)}{kn^\theta \log n} \\ &= \beta(1-\theta). \end{aligned} \tag{41}$$

This completes the proof. □